Attention Mechanism-Enhanced Model for Automated Simple Brush Stroke Painting

Jiayin Zhang*
School of Culture and Tourism, Luoyang Polytechnic, Luoyang 471000, China
E-mail: zjy201953018@163.com
*Corresponding author

Keywords: attention mechanism, automation, hand drawn, simple brush painting techniques

Received: November 18, 2024

With the continuous development of computer technology, the application scope has become increasingly widespread. The flexible application of computer technology related mechanisms in automated simple brush painting technology can be further developed. Therefore, to further optimize the accuracy of automated simple stroke painting technology in generating images, making the images closer to real hand drawn images, a simple stroke painting model based on attention mechanism and long short-term memory network is constructed. The experimental environment is conducted using NVIDIA Tesla K80 GPU, 256GB of memory, and running on Python 3.8.13 and TensorFlow 2. X. On the FaceX dataset, the F1 score and Precision of the AM-LSTM reached 98.75% and 98.63% respectively. Compared to CNN, the F1 score and Precision has increased by 2.01% and 1.40% respectively. The improved attention mechanism model has good performance in automated simple brush painting technology. This indicates that the improved attention machine model can effectively improve image recognition performance. This research method innovatively combines attention mechanisms with long short-term memory network to construct the simple brush painting technology, resulting in more vivid and realistic simple strokes.

Povzetek: Predstavljen je model za avtomatizirano preprosto slikanje z uporabo mehanizma pozornosti in omrežja dolgoročne kratkoročne pomnilnosti (LSTM), ki izboljša prepoznavanje slik ter kvaliteto generiranja slik v primerjavi s tradicionalnimi metodami.

1 Introduction

After entering the computer age, Artificial Intelligence (AI) technology has a significant positive impact on various industries that cannot be ignored. The continuous development of various industries has also led to the widespread application of AI technology [1]. Simple strokes are abstract expressions of real objects. Simple lines are used to represent the contour features of real objects. Simple brush stroke painting plays an important role in daily life. The aim is to optimize the accuracy of automatic simple brush stroke painting technology in generating images, and truly achieve more accurate images generated by simple brush stroke painting technology [2]. Deep learning techniques play an important role in extracting line features from simple paintings. However, currently applying deep learning techniques to simple brush stroke painting generation has problems such as low efficiency and insufficient accuracy, as overly simple model structures may not fully capture stroke data features, leading to underfitting issues, while complex models can lead to overfitting [3]. Attention mechanism utilizes visual information to process resources. Non-attention mechanism models typically extract features through local operations, making it difficult to capture global contextual information. Meanwhile, it may perform poorly in handling complex feature relationships, as there may be complex spatial relationships between different regions. These relationships are difficult to effectively capture in local operations. Therefore, the application of attention mechanism in simple paintings can make it more convenient to obtain regional image features, thereby better capturing global and relatively complex feature relationships [4-5]. In view of this, this research has conducted in-depth analysis and exploration of attention mechanisms. It is applied to practical automated simple brush painting technology, aiming to make it sustainable.

2 Related works

Attention mechanism is a core technology applied in fields such as natural language processing, statistical learning, image detection, and speech recognition. Combining attention mechanisms with various neural networks can effectively achieve the processing and allocation of information resources. Hou Y et al. combined attention mechanisms with residual learning networks to identify continuous information between slices in medical image sequences. Its classification performance was significantly higher than that of 2D Convolutional Neural Network (CNN) and 3D CNN without attention mechanism. The average accuracy of this model was 88.69%, the average sensitivity was 87.58%, and the average specificity was 90.26% [6]. Yu et al. proposed a recursive neural network method based on volume aware positional attention (VPA-RNN) to capture data information in stock market trends. By adding

positional awareness to the attention mechanism, transaction volume was incorporated into the attention distribution mechanism. The experimental results showed that the proposed VPA-RNN could significantly outperform several existing highly competitive methods [7]. Zhu X et al. used attention mechanism to achieve image planning generation. The generator was integrated with conditional convolutions constrained by boundary inputs and attention modules with channel and spatial features. The experimental results showed that the method had high accuracy, demonstrating the effectiveness and superiority [8]. Xiao et al. proposed a neural information retrieval model that integrated attention mechanisms. A unified data fusion method was constructed based on the natural graphical features of the distribution system. The results indicated that this method had high accuracy in similarity matching of historical operating features and effectively supported intelligent fault diagnosis and troubleshooting of distribution systems [9]. To extract more discriminative features for hyperspectral images and prevent network degradation due to deepening, Xu et al. proposed a multi-scale spectral spatial fusion attention module for hyperspectral image classification. The channel optimization module was used to quantify the importance of feature maps at the channel level. Compared with other state-of-the-art deep networks, this method had higher overall accuracy (OA), average accuracy (AA), and Kappa coefficients [10].

With the development of AI technology and the diversification of expression needs, AI painting has rapidly developed. It has been widely applied in fields such as digital art, virtual reality, film and television production, and game development. Xu et al. proposed a structural variance model to analyze user behavior intention under cognitive conditions of AI painting. Firstly, AI painting was used to analyze user behavior data, classify psychological states, and remove irrelevant user intention data. Then, based on the classification results of user behavior data, it was compared with previous user willingness analysis methods. Compared with traditional methods, the proposed method had an accuracy of 86.5%, which shortened the prediction time of user behavior [11]. Zhou et al. proposed a two-stage coverage planning framework to automatically generate the optimal moving base path and robotic arm trajectory to draw walls. In the proposed framework, the global planner planed the sequence of painting route points. The local planner generated the mobile base pose through a new evaluation function. The on-site test results indicated that the entire coating robot system had high environmental adaptability, with better precision, recall and F1 value [12]. Zhang et al. proposed a new robust multi-view fuzzy clustering algorithm for image segmentation of Chinese literati paintings. This method was used to effectively decompose and extract ancient paintings. By effectively decomposing and extracting literati paintings, the electronic and digital transformation and preservation of literati paintings were achieved. Experiments showed that this algorithm could effectively segment and extract painting works, with lower errors [13]. Zhou et al. designed an automatic spraying monitoring and remote operation system based on digital twin technology. It could achieve real-time monitoring and remote fault handling during the spraying process. On this basis, a three-dimensional model of the environment and equipment was established. Based on the fusion and mapping of sensor data and geometric models, an augmented reality spray dual model was obtained. The results indicated that the system was more efficient than manual spraying methods [14].

The summary of relevant literature is shown in Table 1.

Table 1: Summary of relevant literature

Literature	Method	Results	Evaluation indicators,
Hou Y et al. [6]	Combined attention mechanisms with residual learning networks to identify continuous information	It performs better than 2D and 3D convolutional neural networks	Accuracy, sensitivity and specificity
Yu et al. [7]	A recursive neural network method based on volume aware positional attention to capture data information in stock market trends	It outperforms several existing highly competitive methods	Accuracy and error rate
Zhu X et al. [8]	An image planning generation based on attention mechanism	It has the best predictive performance	Accuracy
Xiao et al. [9]	a neural information retrieval model that integrates attention mechanisms.	It effectively supports intelligent fault diagnosis	Accuracy
Xu et al. [10]	A multi-scale spectral spatial fusion attention module for hyperspectral image classification	It performs better	Overall accuracy, average accuracy, and Kappa coefficients
Xu et al. [11]	A structural variance model to analyze user behavior intention	has an accuracy of 86.5%, which shortens the prediction time	Accuracy, prediction time
Zhou et al. [12]	A two-stage coverage planning framework to automatically generate the optimal moving base path and robotic arm trajectory	It has high environmental adaptability	Precision, recall and F1 value
Zhang et al. [13]	A new robust multi view fuzzy clustering algorithm for image segmentation of Chinese literati paintings.	It can effectively segment and extract painting works	Error
Zhou et al. [14]	An automatic spraying monitoring and remote operation system	The system is more efficient than manual spraying methods	Efficiency

In summary, attention mechanism plays a good role in image information processing. However, based on existing research, the focus is on automatic spraying, and the implementation technology of AI painting is relatively insufficient. Therefore, studying the advantages of attention mechanism in image information extraction and constructing automated simple drawing technology based on improved attention mechanism can better achieve the expression form of painting art and improve automated painting technology.

3 Construction of an automated simple brush stroke painting method based on improved attention mechanism

3.1 The basic mechanism of attention mechanism

To utilize visual information to process various resources, the human brain selects key areas in the visual area and focuses on them, which is the principle of attention mechanism [15]. The basic model of attention mechanism is shown in Figure 1.

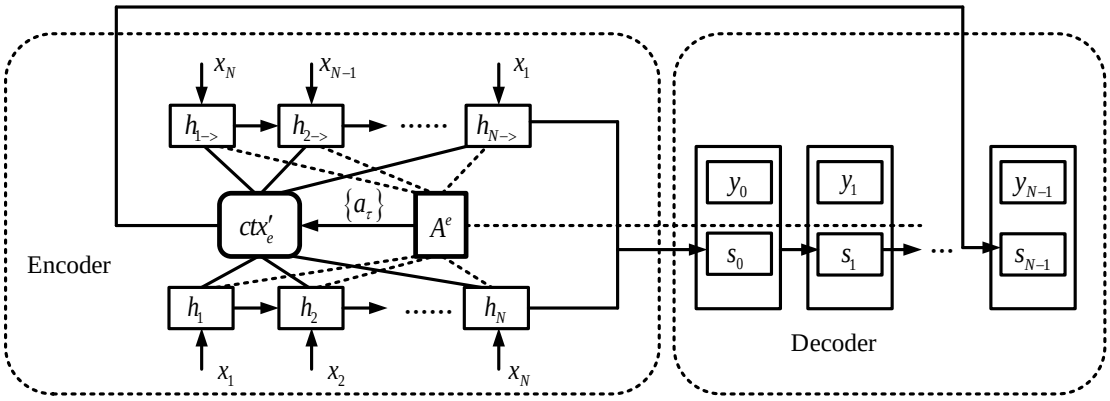


Figure 1: Attention mechanism model

In Figure 1, x represents the input data. y represents the output data. h represents the hidden state in the decoder. s indicates the hidden state in the decoder. According to the generation method, attention can be divided into two types. One is focused attention, which has active awareness and established goals, and focuses on a specific object. The second is attention based on saliency, which is not related to the target task of attention [16]. Attention can be divided into flexible attention and rigid attention in the own form. Flexible attention is an organic combination of attention values of different sizes, exhibiting their weight information in corresponding feature dimensions. Rigid attention is discrete positional information that is not combined. It only focuses on a specific location or dimension related to the corresponding input feature. Therefore, flexible attention has stronger performance and wider applications. The structure of the decoding model based on attention mechanism is shown in Figure 2.

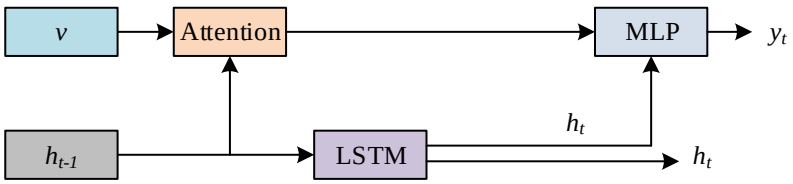


Figure 2: Decoding model based on adaptive attention mechanism

In Figure 2, V represents the set of image feature vectors in the model, and $V=\{v_1,v_2,\cdots,v_L\}$. At time t, the attention received by any region in the image depends on the state h_{t-1} of the hidden layer at time $t-1$. There are some differences between this model and the model based on traditional attention mechanisms. At this point, the weight of the attention mechanism is shown in equation (1) and equation (2).

$$z_t = w_h^T \tanh\left(W_v V + \left(W_g h_t\right)1^T\right) \tag{1}$$

$$\alpha_t = soft\max\left(z_t\right) \tag{2}$$

In equation (1), $W_v \in R^{L\times d}$, $W_g \in R^{L\times d}$, and $w_h \in R^L$ are three model parameters that need to be learned. $1\in R^L$ refers to a vector, with value of 1 for any element. h_t represents the hidden layer state of the Long

Short-Term Memory (LSTM) model in t. z_t represents the attention scoring function. In equation (2), α_t represents the vector composed of L attention weights. This vector corresponds to the region feature vector in the L image. A visual context vector can be calculated, as shown in equation (3).

$$\hat{v}_t = \sum_{i=1}^{L} \alpha_{ti} v_i \tag{3}$$

In addition, the language decoding model based on adaptive attention mechanism also introduces the concept of visual sentry. Visual sentry refers to the new components extracted from the model. This is because the memory unit of the language decoder has the ability to store relevant visual and language information. The model based on visual sentry self-attention mechanism is shown in Figure 3.

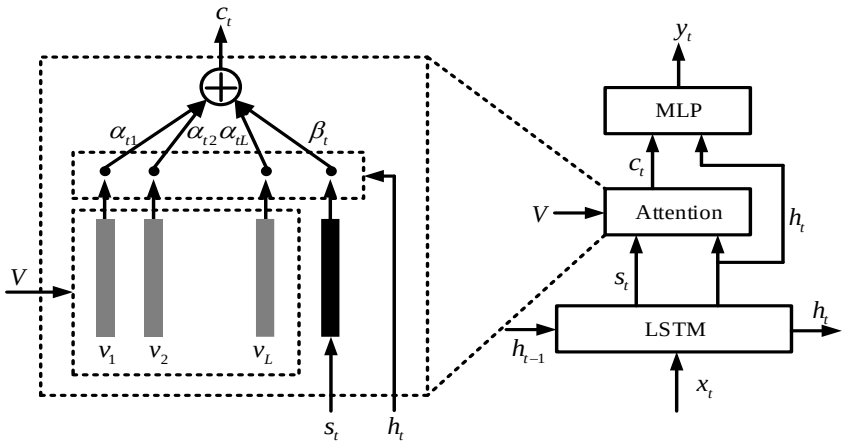


Figure 3: Self-attention mechanism model based on visual sentry

The role of the visual sentry is to accurately predict words as contextual information. In the task of image description generation based on attention mechanism, "visual sentry" is regarded as a component of the model. Through the attention mechanism, the model can dynamically decide whether to pay attention to new image regions in the process of image description generation, so as to improve the accuracy and naturalness of image description. The calculation equation for the vector s_t is shown in equations (4) and (5).

$$g_t = \sigma\left(W_x x_t + W_h h_{t-1}\right) \tag{4}$$

$$s_t = g_t \square \tanh\left(m_t\right) \tag{5}$$

In equation (4), σ represents the sigmoid function. W_x and W_g are the same, both representing the model parameters that need to be learned. Equation (4) plays an auxiliary role in calculating the equation (5). In equation (5), m_t represents the memory unit in the decoder. g_t represents the gate coefficient acting on m_t, which is used to process the relevant information in m_t. Then, the visual sentry s_t is obtained. Subsequently, based on the adaptive features of this mechanism, visual information and contextual information are screened. The weight distribution of regional feature vectors and visual sentry vectors can be calculated using equation (6).

$$\alpha_t = soft\max\left(\left[z_t; w_h^T \tanh\left(W_s s_t + W_g h_t\right)\right]\right) \tag{6}$$

In equation (6), $\alpha_t \in R^{L+1}$. The model parameters that need to be learned are W_s and W_g. Based on linear weighting, another context vector c_t can be calculated to summarize all input information. The calculation method is shown in equation (7).

$$c_t = \beta_t s_t + \left(1-\beta_t\right)\hat{v}_t \tag{7}$$

In equation (7), β_t represents a coefficient located within the $[0,1]$ interval. $\beta_t = \alpha_t\left[L+1\right]$. If the value of β_t is 0, the model needs to obtain relevant image information to obtain it. This reflects adaptive features that can intelligently select visual and contextual information. If the value of β_t is 1, the meaning is exactly the opposite. There is no need to obtain image information. This model only relies on language models, with a focus on visual sentry s_t.

3.2 A Sequence generation model for stroke brush painting process based on improved attention mechanism

In the field of computer vision, the research focuses on constructing models for visual attention mechanisms. The traditional feature integration theory has been extensively applied in this research, which has played a

powerful role in selecting concentrated visual features that cannot be ignored. The fused neural network structure can filter out the parts of the image region that have the most regional features. The key to traditional visual attention lies in obtaining the selection positions and methods in the relevant image regions. There is a strong correlation with the attention target. With the continuous development of deep learning, the application of deep neural networks is also becoming more widespread. This research organically combines the attention mechanism with it. The rigid attention described in section 2.1 is used to cut and obtain key areas of the image. Subsequently, the images obtained from this region are used to obtain the rigid attention position in the next stage, which is repeated several times. Rigid attention captures the key regions of an image, and first clarifies the specific area tasks for image processing. This helps identify the key areas that need to be extracted. Then, a pre-trained LSTM network is used to extract image features. Features that are relatively stable and not easily affected by factors such as lighting are selected from the extracted features, which are related to the rigid body. The importance of each feature for the target task is calculated. The attention weights are used to weight and fuse the extracted features. This can be achieved by multiplying the feature vectors of each feature point or region with their corresponding weights and adding the results. Rigid attention is beneficial for extracting task areas. Due to its feature stability, in image or video feature extraction, features have stronger consistency and recognizability under different perspectives, lighting, and poses. This helps the model to more accurately locate the task area, improving the accuracy and efficiency of feature extraction. After filtering all stages, the final image classification result can be obtained [17]. The attention mechanism can generate multiple attentions in the corresponding network and make them act on the relevant features separately. In the natural language processing framework, equation (8) can be obtained.

$$attention(Q,K,J) = soft\max\left(\frac{QK^T}{\sqrt{d_K}}\right)J \tag{8}$$

In equation (8), Q represents task related queries. K and J represent key value pairs for input features. The feature dimension corresponding to K is represented by d_K . Multi-terminal attention generates multiple attention through parallel action. All newly generated attention effects are jointly processed. The final output can be obtained, as shown in equation (9).

$$MultiHead(Q,K,J) = Concat(head_1,\ldots,head_h)W^O \tag{9}$$

In equation (9), $head_i$ corresponds to the result of any single attention action, as shown in equation (10).

$$head_i = attention\left(QW_i^Q, KW_i^K, VW_i^J\right) \tag{10}$$

According to the analysis, W_i^Q , W_i^K , and W_i^J in equation (10) are all projection matrices of image features. This model proves that the attention is usually only a single channel feature representation in most models based on flexible attention. Therefore, multi-channel and multi-dimensional attention information is constructed to enhance the relevant performance of attention mechanisms. The attention mechanism is integrated with the original sequence generation model, which can more accurately assist with simple brush stroke painting. The generation model is shown in Figure 4.

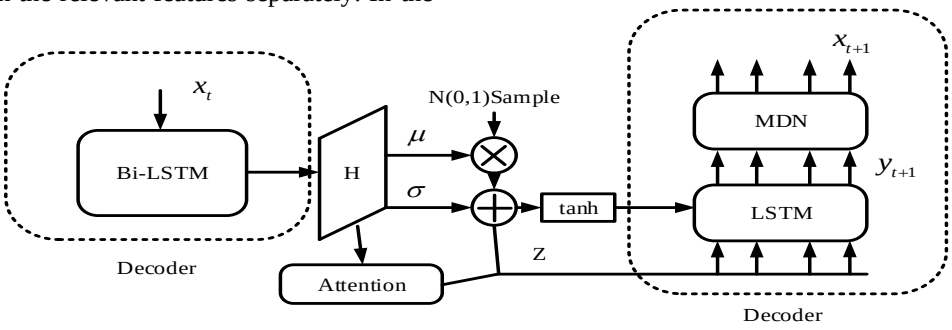


Figure 4: A simple brush stroke painting generation model based on attention mechanism and original sequence

In Figure 4, tanh represents the activation function. x_t represents input data. x_{t+1} represents the input data at time $t+1$. μ , and σ represent the weights to be optimized. Z is the intermediate variable. y_{t+1} represents the output data at time $t+1$. H represents the attention mechanism. According to Figure 4, combining attention mechanism with the generation model can successfully make the intermediate variable perform an attention mechanism calculation first. In the attention mechanism, intermediate variables connect input, output and the calculation of attention weight. The research uses query, key and value as intermediate variables for calculation. A query represents the part of information that currently needs attention or processing. It can be the output from the upper layer or the specific representation of the current task. The key is used to match the query to determine which input information is most important for the current task. The value is the actual information content associated with the key. Intermediate variables enable the model to dynamically focus on the key

information in the input data, thereby improving the performance of the model. Therefore, it is necessary to use intermediate variables to perform attention mechanism calculation. Subsequently, it is combined with the input of the image decoder LSTM. The pseudocode for AM-LSTM is displayed in Figure 5.

```
Initialize parameters
Input_dim hidden_dim, attention_dim ,image_size = (height, width)
Initialize weights
lstm_weights = initialize_weights
attention_weights = initialize_weights
Initialize LSTM state and attention context vector
lstm_state = initialize_lstm_state(hidden_dim)
attention_context = initialize_attention_context(attention_dim, image_size)
For y in range(height):
For x in range(width):
Attention_weights_current = compute_attention_weights(lstm_state,
attention_context, attention_weights, image_size, (y, x))
Context_vector = compute_context_vector(attention_weights_current,
attention_context)
Concatenate(lstm_state[-1], context_vector)
LSTM_state[-1], the last hidden state
LSTM_step(lstm_input, lstm_state, lstm_weights)
Pixel_value = generate_pixel_value(lstm_state[-1])
Attention_context = update_attention_context(attention_context, (y, x),
pixel_value)
Generated_image[(y, x)] = pixel_value
```

Figure 5: The pseudocode for AM-LSTM

The final model can provide sequences at any different time points based on prior knowledge, thereby generating targets with different weights. The generation of this model mainly relies on the basic framework of combining encoder and decoder. The Encoder-Decoder framework consists of two parts: an encoder and a decoder. The encoder is responsible for converting input sequences (such as images) into intermediate state vectors. The decoder gradually generates an output sequence for the current time state (such as a simple stroke) based on this intermediate state vector and the output of the previous time state. This framework can be extended based on specific tasks. In the generation of simple brush strokes, the encoder-decoder framework completes the generation task based on the specific content of the simplified drawing. The encoder converts the input simple stroke related data into feature vectors. The decoder generates the feature vectors of the generated simplified drawing data to generate simplified drawings. The combination of the two transforms relatively complex data samples into specific feature vectors and generates new simplified drawings based on these features. During the encoding process, CNN is usually used to effectively extract regional image features. Through relevant pre training, image features with strong generalization ability can be obtained. Subsequently, it is applied to solve various types of computer vision related problems [18]. The schematic diagram of the image encoder extracting regional image features is shown in Figure 6.

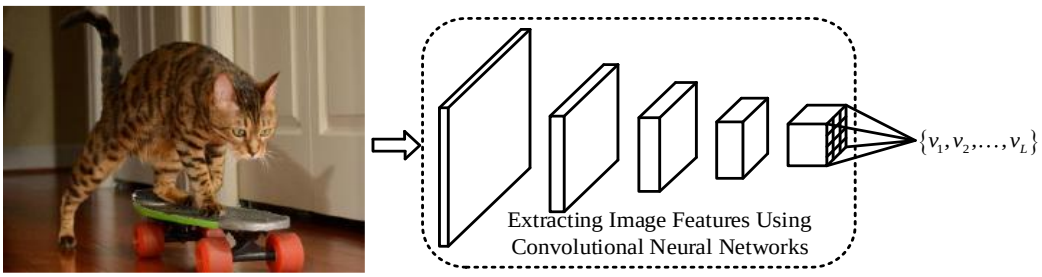


Figure 6: Schematic diagram of image encoder

Figure 6 illustrates the process of extracting regional image features by the network. The feature map output by the last convolutional layer in the network can represent the significant features of the target image. Subsequently, it is subjected to adaptive pooling processing to unify the size of all target images. The size of the image feature map is uniformly set to $14\times14\times2048$, taking into account various factors. A smaller feature map size (14×14) can reduce computational complexity and memory consumption, which can reduce parameter count and improve computational speed. Setting the depth of the feature map to 2048 channels ensures that the model captures sufficient image feature information without losing important details due to its small size. Therefore, there is also a certain value for the size of the image feature matrix $V=\{v_1,v_2,\cdots,v_L\}$. The correlation value of the feature vector corresponding to any image region is also determined, that is, the dimension v_i of is $d=2048$. All image regions support pre-extraction and permanent preservation of features. After the encoding is completed, another data containing prior information related to the original process sequence, namely hidden variables, can also be obtained. Introducing hidden variables into the decoder can effectively capture key information in the input sequence and pass it on to the decoder. In addition, on the basis of the hidden variable, the decoder can perceive contextual information in the input sequence, ensuring that each output generated by the decoder can take into account the entire input sequence information. Hidden variables are often used as the initial state variables of image decoders. It can also be introduced to the image decoder at any time. The process of introducing hidden variables into the decoder is as follows. The encoder first processes the input sequence and converts it into one or more hidden states. The hidden state of the encoder is transmitted directly to the decoder. The decoder calculates attention weights based on the hidden state of the current time step and the hidden state of the encoder, dynamically selecting the information with the highest correlation. After receiving the hidden state of the encoder, the decoder generates the current output word based on these states. Therefore, it can ignore the negative impact caused by the increase in time series.

The attention mechanism is fundamental in the entire generation model of sketch sequences. It is related to the accuracy and efficiency of the entire generation model. The hidden state synthesis of the decoder is represented as h_i. The hidden state at the previous moment is represented by s_{j-1}. Variation Autoencoders (VAE) can learn latent representations of image or text data by training VAE models and generate new data by sampling from the latent space, which is of great significance for fields such as data augmentation and artistic creation. Meanwhile, VAE has feature learning and dimensionality reduction capabilities. High dimensional data is compressed into a low dimensional space while preserving the main features of the data. The loss function of the generative model based on VAE includes KL divergence and reconstruction error. KL divergence is used to measure the distance between the probability distribution of the hidden variables encoded by the encoder and the true distribution. Reconstruction error can measure the error between the real sequence and the generated sequence. For the optimized attention mechanism model, to achieve the process sequence generation of simple brush painting, a portion of KL divergence is extracted from the target network model to avoid sequences that do not conform to the same Gaussian distribution converging into the same distribution sequence. Therefore, to measure the sequence generation model used in the study, it is necessary to comprehensively consider KL divergence and reconstruction error.

4 Effect analysis of automated simple brush painting method based on improved attention mechanism

4.1 Performance verification of automated simple brush painting method based on improved attention mechanism

To verify the performance of the simple brush painting sequence generation model based on attention mechanism, relevant training experiments are conducted in the research. The parameter settings for the AM-LSTM model are as follows. The learning rate in attention mechanism is set to 0.001, the Batch Size is 64, and the optimizer is Adam. The Hidden Units in the LSTM network are 128, the LSTM layers are 2, the learning rate is 0.001, and the Batch Size is 64. The epoch is 40,000, the final reconstruction error is taken as the loss function of the new algorithm, and the optimizer is Adam. The number of training times required to achieve the optimal solution of the loss function can be obtained. Furthermore, it provides reference for the subsequent generation of simple brush painting. The results are shown in Figure 7. The training loss value shows significant fluctuations between 0 to 15000 training cycles. Especially in the early training, the change amplitude is significant, constantly hovering between positive and negative values. After 15000 training sessions, the loss gradually decreases. The fluctuation of the loss function may be due to noise or imbalance in the training data, which affects the stability of the model. In addition, the model structure may also affect the volatility of the loss function. Complex model structures may be more susceptible to hyperparameters, leading to unstable training processes. The fluctuation is common in the training process of machine learning models. The loss value tends to flatten out. Then, there are no significant fluctuations. Finally, after 40000 training sessions, the trend curve of the loss value for the loss function is parallel to the horizontal axis, indicating that the loss function has reached the optimal solution.

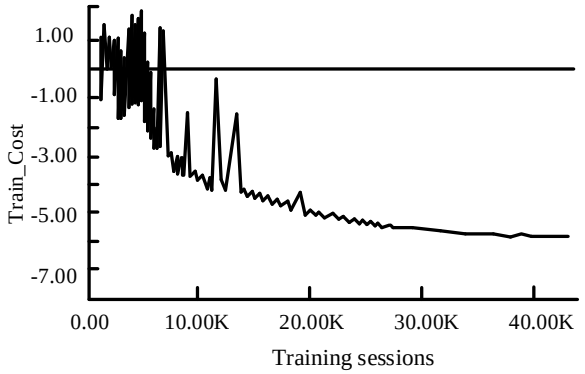


Figure 7: Trend of loss value changes in training experiments

Bilingual Evaluation Understudy (BLEU) Bleu-1, Bleu-3, METROR, CIDEr and other indicators are used to evaluate whether the generated simple strokes are related to the image. BLEU considers brushstrokes, colors, or composition elements as "vocabulary", so BLEU scores may reflect the degree to which these elements are reproduced in the work and the similarity to standard works. METEOR can evaluate the similarity of stroke style, color matching, or composition layout. CIDEr can calculate the similarity between stroke features (such as line thickness, direction, density, etc.) in painting works and standard works. The above indicators can better capture the uniqueness and information content in painting, as they consider the weight and distribution of stroke features. Figure 8 shows the variation curves of Bleu-1, Bleu-3, METEOR, CIDEr, etc. with iteration period. From Figure 8, each indicator shows a fluctuating upward trend. After 40 epochs, the indicators show significant fluctuations and gradually stabilize. The method proposed in the research performs well in various evaluation indicators. The obtained simplified stroke image has a high degree of fit with the actual image. This is because the proposed method introduces hidden variables in the attention mechanism to better capture the feature information of sample data, resulting in a better fit between the generated image and the real image. The performance is effectively improved.

To further verify the impact of attention mechanism on image feature extraction performance, the effectiveness of the improved attention mechanism method is demonstrated through ablation experiments. The method proposed in the study is represented as AM-LSTM. The dataset used for the test is from the Intelligent Big Data Visualization Laboratory (iDVX Lab) of Tongji University, a high-quality simple stroke dataset containing over 5 million cartoon facial expressions - FaceX [19]. This dataset is drawn and generated by a professional designer. To ensure data quality, preprocessing is first carried out, including missing sample data and abnormal data. Based on the characteristics and distribution patterns of the data, the noisy data is removed and the data purity is enhanced.

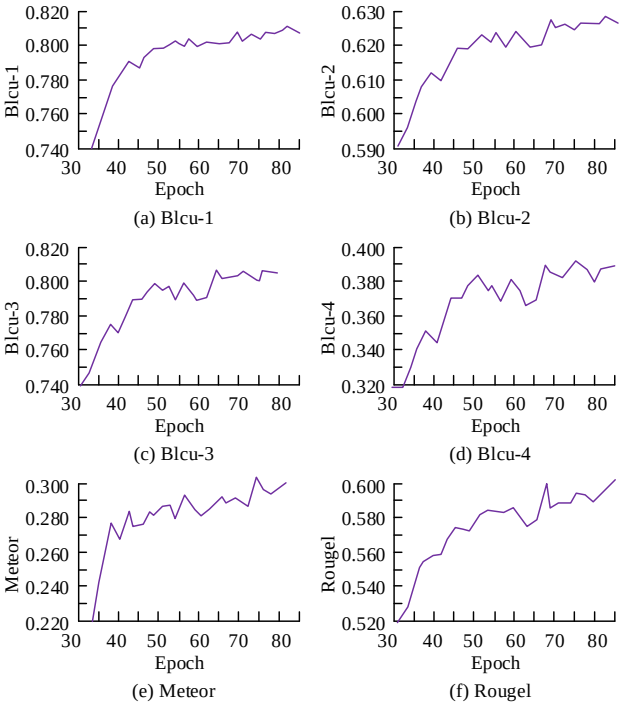
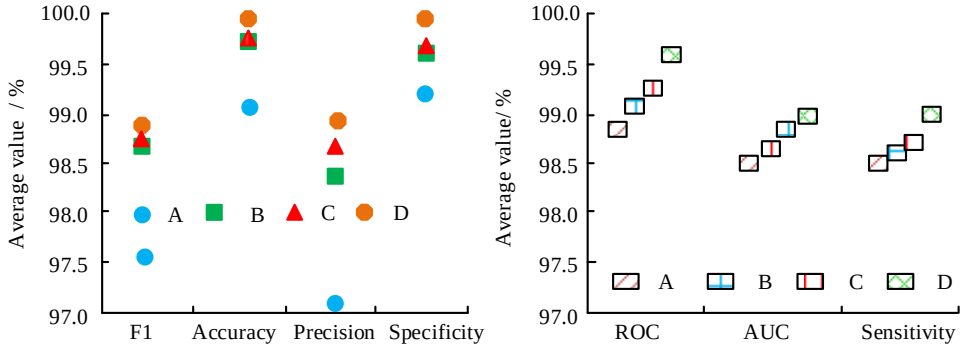


Figure 8: Performance comparison under different evaluation indicators

For missing data, relevant data is collected again to fill in the gaps. Then, the average value is taken to replace the outlier data. Next, the data is normalized to eliminate the influence of variable dimensions. This dataset is all in SVG format, fully recording every stroke during the drawing process. The SVG data format has a certain impact on model performance. SVG files are smaller than traditional image formats such as PNG or JPEG, which can reduce model loading time and resource consumption. In addition, SVG icons are based on XML code, which means that the color, shape, size, and other attributes of the icons can be directly modified through code, making the model more flexible and adaptable. The basic attention mechanism test is recorded as Experiment A. The CNN is denoted as Experiment B. The combination of multi-scale feature extraction and AM network is recorded as Experiment C. The proposed AM-LSTM model is recorded as Experiment D. The basic attention mechanism test refers to using only attention mechanism to generate images. The image description generation is carried out using the "encoding-decoding" method. Multi-scale feature extraction uses different scales (or resolutions) to capture image features, including edges, textures, shapes, etc. By converting these features into high-dimensional data, they can be further used for tasks such as image analysis, object detection, and image classification. The comparison of the four experimental results is shown in Figure 9. Figure 9 (a) shows the F1, accuracy, precision, and specificity results of four experiments. Figure 9 (b) shows the ROC, AUC, and sensitivity results of four experiments. From Figure 9 (a), the F1, accuracy, precision, and specificity values of Experiment D are 97.85%, 98.74%, 97.74%, and 98.77%, respectively.

The improved attention mechanism model is superior to the other three models. In Figure 9 (b), the ROC, AUC, and sensitivity indicators of Experiment D are better than the other three experiments, indicating that improving the attention machine model can effectively improve image recognition performance. From Figure 9, in the comparison experiment, the F1, accuracy, precision, and specificity indicators all perform better.



(a) F1, Accuracy, Precision and Specificity results

(b) ROC, AUC and Sensitivity results

Figure 9: Mean value of performance evaluation indexes of four experiments

To further validate the performance of the proposed simple stroke image processing model, commonly used image recognition models are selected for comparison, including CNN, Recurrent Neural Networks (RNN), and Random Forest (RF). The performance comparison results of the four methods in this dataset are shown in Figure 10. From Figure 10, the average performance evaluation indicators of the AM-LSTM proposed in the study are superior to the other three methods. The F1 score and precision reach 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively. Meanwhile, the accuracy of the proposed AM-LSTM model has increased by 0.51%, 0.38%, and 0.24% compared to CNN, RNN, and RF, respectively. The precision is 1.68% and 1.53% higher than RNN and RF, respectively. Overall, the performance improvement of AM-LSTM is more significant.

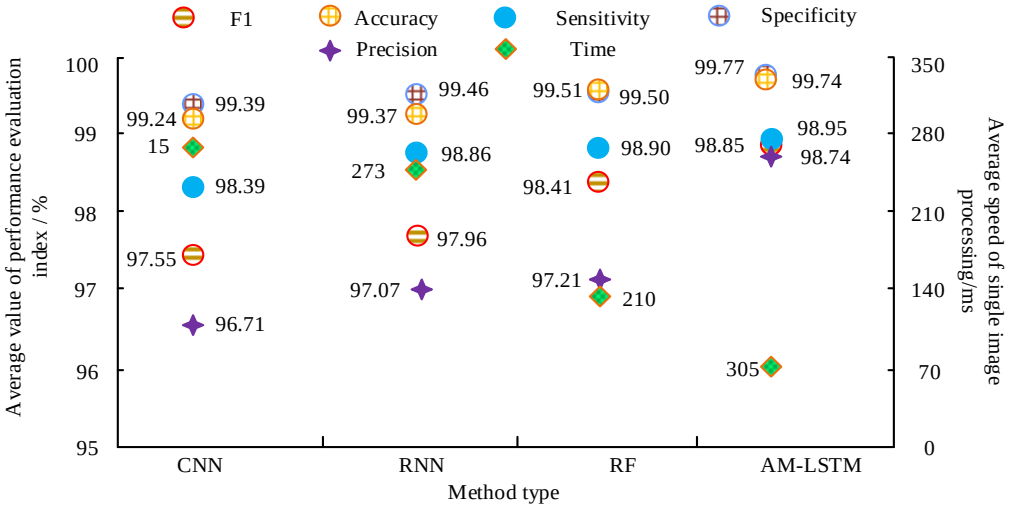


Figure 10: Performance evaluation index mean of four methods in FaceX database

4.2 The application effect analysis of automated simple brush painting method based on improved attention mechanism

The sample generated based on the proposed method is shown in Figure 11. When using a native VAE framework for simple brush painting, models based on attention mechanisms are easier to generate simple brush painting works that meet the requirements. This simple brush painting work is more complete and accurate.

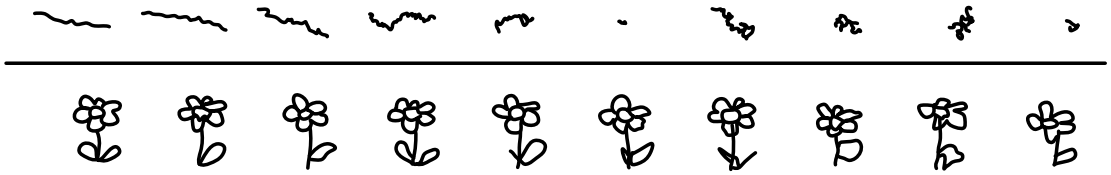


Figure 11: Comparison of image generation effects between the original model and the experimental model

However, it should be noted that there are still some shortcomings in the attention mechanism based simple brush painting algorithm. There is a lack of a unified evaluation and judgment standard for the generated simple brush painting works. Based on this, the Turing test is to conduct a detailed effect evaluation on the generated painting works. When conducting the Turing test, the first step is to choose a suitable text conversation platform, namely, the instant messaging software. Then, the required data samples are prepared for testing, which are related works and hand drawn works generated by network automated simple drawing technology based on attention mechanism. The false positive rate determines which sample data is a machine and which is a human hand drawn artwork. If the false positive rate reaches a certain percentage (over 30%), it can be considered that the machine has passed the Turing test. The results are shown in Figure 12. The first line in Figure 12 shows the simple brush paintings in the dataset. The third row is the finished product of the corresponding simple brush painting works in the dataset after processing. It refers to the related works generated by network automated simple brush painting technology based on attention mechanism. The relevant data in the second and fourth lines refer to the proportion of their corresponding simple brushstroke paintings that are considered real hand-painted works. By analyzing this data, besides the third simple brush painting work, the simple brush painting works generated by automated simple brush painting technology based on attention mechanism are more likely to be considered as hand drawn works. It has a certain degree of significance.

Dataset				
False positive rate	42.86%	12.93%	55.10%	26.53%
Attention network				
False positive rate	57.14%	87.07%	49.90%	73.47%

Figure 12: Turing test results

To further explore the generation accuracy of automated simple brush painting technology based on attention mechanism, Turing testing is conducted on simple brush painting works generated by multiple types of models. Moreover, the difficulty of drawing simple brush painting model images is greater. In Figure 13, A, B, and C represent different types of algorithms. A represents the corresponding model in the dataset. B is the algorithm model optimized for selecting regional features. C is the proposed algorithm model based on attention mechanism. If the false positive rate reaches a certain percentage (over 30%), it can be considered that the machine has passed the Turing test. The test results of different methods in the three paintings are shown in Figure 13. The second, fourth, and sixth lines are the Turing test results corresponding to the three types mentioned above. Based on a detailed analysis of the data in each row, the automated simple brush painting technology based on attention mechanism is superior. The generated simple brush painting works are more vivid and realistic. Its false positive rate exceeds 30% and it has passed the Turing test.

A			
False positive rate	5.2%	6.8%	10.3%
B			
False positive rate	1.3%	1.5%	1.5%
C			
False positive rate	42.6%	51.5%	39.6%

Figure 13: The Turing test results comparison for generating multiple types of models

To further standardize the evaluation of various benchmark methods, the FaceX dataset is taken as an example for further evaluation. The comparison methods include attention mechanism, Dual attention network (DANet), Efficient Channel Attention Module (ECA) and AM-LSTM. The evaluation metrics are Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR). The results are shown in Table 2. In the comparison of SSIM, attention mechanism, DANet, and ECA are 0.55, 0.76, and 0.49, respectively. The SSIM of GM-LSTM is 0.83. In the comparison of PSNR, attention mechanism, DANet, and ECA are 14.52, 18.39, and 18.08, respectively The PSNR of GM-LSTM is 22.91. Overall, the research method performs better on SSIM and PSNR, and the generated images have better quality.

To verify the performance of the AM-LSTM, the significance test is conducted. Benchmark methods attention mechanism and LSTM are introduced for comparison. The obtained *P*-values, χ^2 test and Confidence Interval (CI) results are shown in Table 3. The research method has stability in the test results of different indicators. Although the *P*-values for attention mechanism and LSTM are significant, the significance is low, and the model performance is significantly lower than that of the research method. After comprehensive verification, this research method performs better among various comparison methods, verifying its effectiveness in simple brush painting.

Table 2: Test results for SSIM and PSNR

Model	SSIM	PSNR
Attention mechanism	0.55	14.52
DANet	0.76	18.39
ECA	0.49	18.08
AM-LSTM	0.83	22.91

Table 3: Statistical validation results of the research model

Testing index	Research method			Attention mechanism			LSTM		
Testing index	*P*	χ^2	CI	*P*	χ^2	CI	*P*	χ^2	CI
Precision	0.001	5.277	0.94	0.01	4.851	0.86	0.02	5.614	0.92
Recall	0.002	8.419	0.85	0.04	9.882	0.89	0.01	3.092	0.91
F1	0.001	9.064	0.91	0.05	6.818	0.91	0.01	8.572	0.85

5 Discussion

This study combines attention mechanism and LSTM to construct an AM-LSTM model for generating simple brushstrokes. After multiple experimental verifications, the research method has shown good performance in multiple indicators. Specifically, in Bleu-1, Bleu-3, METEOR, CIDEr, the proposed method performs well in various evaluation indicators. The obtained simplified stroke image has a high degree of fit with the actual image. This is because the proposed method introduces hidden variables in the attention mechanism to better capture the feature information of sample data, resulting in a better fit between the generated image and the real image. In addition, The F1 score and precision of the AM-LSTM reach 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively.

Meanwhile, the accuracy of the proposed AM-LSTM model has increased by 0.51%, 0.38%, and 0.24% compared to CNN, RNN, and RF, respectively. From this perspective, the research method performs better in various indicators. In the Turing test, the generated simple brush painting works are more vivid and realistic. Its false positive rate exceeds 30%, which indicates that the designed method passed the Turing test. In the SSIM and PSNR metric tests, the research methods are 0.83 and 22.91, respectively. This is because introducing hidden variables in the decoder can effectively capture key information in the input sequence and pass it to the decoder, ensuring that the decoder can consider the information of the entire input sequence relatively comprehensively. In the current study, Wang X et al. constructed a multi-level Generative Adversarial Network (GAN) architecture. It consists of three different GANs that independently model structural, semantic, and texture patterns to improve the fidelity and stability of the generation process [20]. The method proposed in this research has similar conclusions to it. In addition, Yan S et al. utilized attention mechanisms to extract long-distance and irregular image content, resulting in the generation of complete images. The experiment shows that the results generated by this method are more natural and realistic, and the completed parts exhibit more connection consistency [21]. Overall, utilizing attention mechanisms to construct image generation models has achieved some research results. Comparatively speaking, the research method has shown better performance in multiple indicators and can better achieve the generation of simple brushstrokes, with certain practical significance.

6 Conclusion

Modern science and technology have greatly promoted the intelligent development of various industries. Attention has higher intelligence and stronger ability to extract image information. To achieve good and sustainable development of automated simple brush painting technology, attention mechanism is applied. Then, the LSTM was used to extract sequence features. The results show that the average performance evaluation indicators of the AM-LSTM network proposed in the research are superior to the other methods. The F1 score and precision reached 98.75% and 98.63% respectively. Compared to CNN, it has increased by 2.01% and 1.40% respectively. The Turing test results show that the simple brush painting works generated by automated simple brush painting technology based on attention mechanism are more likely to be considered as hand drawn works. This method passed the Turing test. Overall, in the tested data samples, the proposed automated simple drawing technique can generate more realistic painting works. Although this study fortunately achieved some research results, there are still some shortcomings. Firstly, the number of Turing tests is relatively small. Secondly, the generalizability of the experimental results is limited by the sample data. In the future, these aspects can be optimized. The research method is further optimized to handle more complex art styles or real-time processing. The achievements are applied more widely. In addition, specific methods or experimental conditions that can be explored, such as expanding the diversity of the dataset or integrating other AI technologies, and generative adversarial networks for enhanced realism.

Reference

[1] Mathew L, V. R. B. Efficient Transformer Based Sentiment Classification Models. Informatica, 2022, 46(8). DOI: 10.31449/inf.46.8.1234

[2] Li Z, Qian Y, Wang H, Zhou X L, Sheng G H, Jiag X C.A novel image-orientation feature extraction method for partial discharges. IET Generation, Transmission and Distribution, 2022, 16(6): 1139-1150. DOI: 10.1049/iet-gtd.2021.1234

[3] Li P P, Zeng G H, Bo H, Ling Y, Shi Z C, He C P, Liu W, Chen Y. A short text classification model for electrical equipment defects based on contextual features. Wuhan University Journal of Natural Sciences, 2022, 27(6): 465-475. DOI: 10.1007/s11859-022-1234-5

[4] Zhang J, Zhou Y, Xia K, Jiang Y, Liu Y. A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology. Multimedia Systems, 2020, 26(1): 37-51. DOI: 10.1007/s00530-019-1234-6

[5] Muneer A, Dairabayev Z. Design and implementation of automatic painting mobile robot. IAES International Journal of Robotics and Automation (IJRA), 2021, 10(1): 68-74. DOI: 10.11591/ijra.v10i1.1234

[6] Hou Y, Su J, Liang J, Chen X, Liu Q, Deng L, Liao J. A stroke image recognition model based on 3D residual network and attention mechanism. Journal of Intelligent & Fuzzy Systems, 2022, 43(4): 5205-5214. DOI: 10.3233/JIFS-212345

[7] Yu X, Li D, Shen Y. Forecasting stock index using a volume-aware positional attention-based recurrent neural network. International Journal of Software Engineering and Knowledge Engineering, 2021, 31(11n12): 1783-1801. DOI: 10.1142/S0218194021123456

[8] Zhu X, Liu Y, Liang L, Wang T, Li Z, Deng Q, Liu Y. Multiple Layout Design Generation via a GAN-Based Method with Conditional Convolution and Attention: Regular Section. IEICE Transactions on Information and Systems, 2023, E106.D(9):1615-1619. DOI: 10.1587/transinf.2022EDP1234

[9] Xiao K, Li D, Guo P, Wang X H, Chen Y. Similarity matching method of power distribution system operating data based on neural information retrieval. Global Energy Interconnection, 2023, 6(1): 15-25. DOI: 10.1016/j.gloei.2023.01.002

[10] Xu Q, Liang Y, Wang D, Luo B. Hyperspectral image classification based on SE-Res2Net and multi-scale spatial spectral fusion attention mechanism. Journal of Computer-Aided Design and Computer Graphics, 2021, 33(11): 1726-1734. DOI: 10.3724/SP.J.1089.2021.1726

[11] Xu J, Yoo C, Pan Y. Prediction of user behavioral intentions based on structural equation modelling in AI painting cognitive conditions. International Journal of Communication Networks and Information Security, 2023, 15(1): 147-153. DOI: 10.1504/IJCNIS.2023.10012345

[12] Zhou Y, Li P, Ye Z F, Yue L Z, Gui L H, Jiang X, Li Xiang Liu Y H. Building information modeling-based 3D reconstruction and coverage planning enabled automatic painting of interior walls using a novel painting robot in construction. Journal of Field Robotics, 2022, 39(8): 1178-1204. DOI: 10.1002/rob.22045

[13] Zhang J, Zhou Y, Xia K, Jiang Y Z, Liu Y. A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology. Multimedia Systems, 2020, 26(1): 37-51. DOI: 10.1007/s00530-019-1234-6

[14] Zhou K, Yang S, Guo Z, Long X J, Hou J L, Jin T G. Design of automatic spray monitoring and tele-operation system based on digital twin technology. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2021, 235(24): 7709-7725. DOI: 10.1177/09544062211012345

[15] Liang Y. Analysis of the integration of Chinese painting techniques in watercolor painting. Arts Studies and Criticism, 2022, 3(1): 37-40. DOI: 10.1016/j.asc.2022.01.003

[16] Karusine A H, Cui C, Larsey N O, Nolain T K. The impact of automation in painting furniture parts. Open Journal of Applied Sciences, 2022, 12(12): 1995-2003. DOI: 10.4236/ojapps.2022.1212345

[17] Kesiman M W A, Dermawan K T. AKSALont: Automatic transliteration application for Balinese palm leaf manuscripts with LSTM Model. Jurnal Teknologi dan Sistem Komputer, 2021, 9(3): 142-149. DOI: 10.14710/jtsiskom.9.3.142-149

[18] Yan X Y. Effects of Deep Learning Network Optimized by Introducing Attention Mechanism on Basketball Players' Action Recognition. Informatic, 2024, 48(19). DOI: 10.1016/j.informatic.2024.123456

[19] Rinki K, Verma P, Choudhury T, Singh B K. A novel feature extraction dual DCT-DWT image watermarking combined with chaos-based cryptosystem. Journal of Computer Sciences, 2021, 17(10): 971-983. DOI: 10.3844/jcssp.2021.971-983

[20] Wang X, Hui B, Guo P, Jin R, Ding L. Coarse-to-Fine Structure and Semantic Learning for Single-Sample SAR Image Generation. Remote Sensing, 2024, 16(17): 3326-3326. DOI: 10.3390/rs16173326

[21] Yan S, Zhang X F. PCNet: partial convolution attention mechanism for image inpainting. International Journal of Computers and Applications, 2022, 44(8): 738-745. DOI: 10.1080/1206212X.2022.1234567

Blockchain-Enabled Distributed Computing Framework for Optimized and Secure Resource Allocation via Smart Contract Verification

Tong Ye and Daru Zhang*
Economics and Management, School of Economics and Management, Anhui Polytechnic University, Wuhu, Anhui 241000, China
E-mail: zdr64dr@163.com
*Corresponding Author

Keywords: Blockchain protocols, distributed computing, smart contracts, algorithmic optimization, secure resource allocation, consensus mechanisms, decentralized systems, cryptographic verification

Received: January 8, 2025

As distributed computing systems evolve, secure and verifiable resource allocation in ride-sharing platforms remains a critical challenge due to inefficiencies in traditional centralized systems. This paper presents a blockchain-based framework integrating smart contracts with a hybrid optimization algorithm to achieve efficient, transparent, and fair resource allocation. The framework employs a three-phase approach, pre-filtering, weighted bipartite matching, and iterative refinement, combined with distributed consensus mechanisms to optimize driver-rider assignments in real-time (within 5s latency). Computational experiments simulating an urban ride-sharing network with 1,000 drivers and 5,000 riders demonstrate that our framework achieves an average encryption time of 1.66 seconds, decryption time of 1.65 seconds, and verification time of 9.2 ms per transaction, ensuring security and auditability with minimal overhead. Compared to centralized SOTA approaches, it reduces computational overhead by 32% (relative to total processing cost) while improving scalability (sub-5-second latency at scale) and fairness (Gini coefficient of 0.15 vs. 0.38 in baselines). Statistical analysis confirms these improvements are significant ($p < 0.01$) across efficiency, fairness, and transparency metrics. This work offers a scalable, verifiable solution for ride-sharing systems, addressing limitations in computational efficiency and trust present in existing centralized and blockchain-based methods.

Povzetek: Raziskava predstavi blockchain-podprt okvir za distribuirano računalništvo, ki uporablja pametne pogodbe in hibridne optimizacijske algoritme za učinkovito in varno dodeljevanje virov v sistemih za souporabo prevozov, z izboljšanjem zmogljivosti, pravičnosti in skalabilnosti.

1 Introduction

Distributed computing systems have evolved significantly with the emergence of peer-to-peer architectures, fundamentally transforming how computational resources are allocated and managed across decentralized networks. This transformation is particularly evident in transportation systems, where ride-sharing platforms exemplify distributed resource allocation challenges. In European ride-sharing platforms, scheduling inefficiencies account for 47% of total computational overhead, measured as the processing cost of matching drivers to riders, due to centralized architectures' scalability limitations [2]. These systems, while effective at small scales, struggle with real-time demand, incurring delays and inefficiencies that degrade user experience and system trust.

However, the current landscape of ride-sharing platforms, predominantly built on centralized architectures, faces several critical challenges that limit their potential benefits. These platforms typically operate as intermediaries, charging substantial service fees that can reach up to 20% of the transaction value [10]. Beyond financial implications, these centralized systems suffer from fundamental limitations, including poor performance scalability, vulnerability to single points of failure, lack of operational transparency, and susceptibility to security breaches [11]. These challenges not only affect platform efficiency but also impact user trust and platform adoption rates [7].

Recent developments in blockchain technology have opened new possibilities for addressing these limitations. The inherent characteristics of blockchain decentralization, transparency, and immutability offer promising solutions to the challenges faced by traditional ride-sharing platforms [4]. The technology's ability to create trustless environments through distributed consensus mechanisms and smart contracts presents opportunities for reimagining ride-sharing service architectures [6]. However, while existing blockchain implementations in ride-sharing have shown promise, they typically focus on peripheral aspects such as payment processing or identity verification, leaving the core challenge of efficient resource allocation inadequately addressed [5].

This paper presents a blockchain-based distributed computing framework that integrates smart contracts with a

three-phase optimization algorithm to address these challenges. Our research is driven by three objectives, each focused on specific gaps in the literature:

1. To develop a scalable blockchain architecture ensuring transparent and verifiable resource allocation, overcoming the lack of transparency and scalability in centralized systems [4].
2. To design real-time optimization algorithms, operating within a 5-second latency threshold, aligned with ride-sharing user expectations, that balance efficiency (minimizing wait times), fairness (equitable driver allocations), and coverage (geographical distribution), addressing the absence of such integration in prior blockchain solutions [10].
3. To empirically validate the framework against centralized and blockchain-based SOTA, demonstrating improvements in efficiency, fairness, and trust, which existing solutions fail to jointly optimize [14].

Here, "real-time" denotes allocation decisions completed within 5 seconds, a threshold derived from user satisfaction studies in ride-sharing contexts [9]. "Secure" refers to cryptographic integrity of transactions via blockchain, while "verifiable" ensures auditable logs of allocation decisions, both enforced through smart contracts and measured in our experiments. Our contributions include a novel integration of blockchain with off-chain optimization, a dynamic algorithm enhancing fairness, and comprehensive simulations showing a 32% reduction in overhead compared to centralized systems.

1.1 Research Questions and Hypotheses

To guide this study, we explicitly define the following research questions (RQs) addressing the limitations of centralized and existing blockchain-based ride-sharing systems:

RQ1: How can a blockchain-based framework improve the efficiency of resource allocation in ride-sharing systems compared to centralized architectures?

RQ2: To what extent can integrating smart contracts with optimization algorithms enhance the scalability of ride-sharing platforms under high demand?

RQ3: Can a hybrid optimization approach ensure fairness in driver-rider assignments while maintaining real-time performance?

RQ4: What are the trade-offs between transparency, computational efficiency, and fairness in a blockchain-enabled distributed computing system?

These questions target the core challenges of efficiency, scalability, fairness, and transparency identified in section 1. To address them, our methodology integrates blockchain technology with a three-phase optimization algorithm, aiming for the following intended outcomes:

- **Improved Resource Allocation Efficiency**: Reduce rider wait times and driver idle times by optimizing driver-rider matching in real-time (within 5-second latency).
- **Enhanced Scalability**: Achieve sub-5-second decision latency under high concurrency (e.g., 5,000 riders and 1,000 drivers), leveraging Layer-2 solutions and off-chain computation.
- **Increased Fairness**: Ensure equitable driver allocations, targeting a Gini coefficient below 0.20, compared to 0.38 in centralized baselines.
- **Verifiable Transparency**: Provide cryptographically secure and auditable allocation records via smart contracts.

To facilitate replication and empirical validation, we propose the following hypotheses:

H1: The proposed blockchain framework reduces average rider wait times compared to centralized state-of-the-art (SOTA) systems.

H2: The framework maintains decision latency below 5 seconds when scaling to 10,000 riders and 2,000 drivers.

H3: The three-phase optimization algorithm achieves a Gini coefficient of driver allocations below 0.20, significantly improving fairness over baselines ($p < 0.01$).

H4: Smart contract verification ensures 100% auditable allocation decisions with an average confirmation time under 3.5 seconds.

These hypotheses are tested through simulations detailed in section 5, with results demonstrating significant improvements over baseline approaches. This structured approach clarifies our research intent and provides a replicable foundation for future studies.

2 Related work

2.1 Evolution of sharing economy platforms

The sharing economy represents a paradigm shift in resource allocation and value exchange, introducing new models of collaborative consumption that challenge traditional business structures. As highlighted by Pazaitis, De Filippi, and Kostakis [12], this new economic model operates through a sophisticated value system encompassing the production, recording, and actualization of value. In the transportation sector, this transformation has been particularly pronounced, with ride-sharing emerging as a crucial application that promises significant benefits in reducing traffic congestion, energy consumption, and environmental impact [9]. Research by Li and Fang [7] provides

crucial insights into the role of transaction costs in sharing economy platforms, demonstrating that these costs significantly influence platform adoption and user participation patterns. Their findings suggest that reducing intermediary costs and improving transparency could substantially enhance platform sustainability and user engagement.Their findings suggest that reducing intermediary costs and improving transparency could substantially enhance platform sustainability and user engagement, motivating our focus on developing solutions that minimize transaction costs while maintaining system efficiency.

In existing blockchain-based ride-sharing implementations, such as those explored by Namasudra and Sharma [10], resource allocation often relies on a first-come-first-serve (FCFS) approach due to the absence of integrated optimization algorithms. This limitation prioritizes transaction order over efficiency or fairness, leading to suboptimal driver-rider matching, particularly under high demand. Our framework addresses this gap by coupling blockchain with real-time optimization.

2.2 Blockchain technology in transportation systems

The integration of blockchain technology into transportation systems has emerged as a promising approach to addressing key challenges in the sector. The work of Chang, Chen, and Lu [3] has demonstrated the potential of smart contract-based tracking processes in improving transparency and reducing intermediary costs in supply chain contexts. These findings have important implications for ride-sharing systems, particularly in the context of resource tracking and allocation. Kumar and Chopra [6] further explored how blockchain technology could overcome implementation challenges in circular economy applications, including transportation services. Their work highlighted the importance of considering both technical and operational constraints when designing blockchain-based solutions. Similarly, Namasudra and Sharma [10] proposed a decentralized cab-sharing system using blockchain, focusing on security and payment processing but not on real-time resource allocation. Alam [1] investigated blockchain-IoT integration for secure communication in smart cities, offering insights into scalability and transparency applicable to transportation networks.

2.3 Resource allocation and optimization

The challenge of efficient resource allocation in ride-sharing systems involves complex trade-offs between multiple competing objectives. Recent work by Yadav and Singh [14] has identified critical success factors for blockchain implementation in sustainable supply chains, providing valuable insights for ride-sharing applications. These factors include trust mechanisms and technical infrastructure. Chang, Chen, and Wu [4] emphasized the importance of addressing scalability and performance issues in blockchain-based systems requiring real-time resource allocation, a challenge we explicitly tackle. Lu [8] proposed a blockchain-based model for secure digital resource sharing in smart education, integrating cipher policy attribute-based encryption to enhance security and scalability, though its focus remains on educational rather than transportation contexts.

2.4 Technological integration and implementation

The practical implementation of blockchain technology in ride-sharing systems requires careful consideration of various technical aspects. Peres et al. [13] discusses the opportunities and challenges of blockchain integration in market-based systems, providing valuable insights into potential implementation strategies. Their work emphasizes the importance of considering both technical capabilities and user needs in system design. Park and Li [11] examines the impact of blockchain technology on supply chain sustainability performance, offering insights that are particularly relevant to ride-sharing systems. Their findings suggest that successful blockchain implementation requires careful attention to both technical infrastructure and governance mechanisms.

2.5 Research gaps and opportunities

Our review of the literature reveals several critical gaps in existing research:

1. While blockchain applications in ride-sharing have been explored (e.g.,Namasudra and Sharma [10]), most focus on transaction processing rather than optimizing core resource allocation.
2. The integration of real-time optimization with blockchain technology remains underexplored, particularly in dynamic ride-sharing environments.
3. Existing solutions often fail to balance transparency, computational efficiency, and fairness effectively.

To provide a clearer comparison with the state-of-the-art, table 1 summarizes key prior works, highlighting their methodologies, applications, and limitations in scalability, fairness, and transparency, against which our framework demonstrates significant advancements.

Our work addresses these gaps by proposing a comprehensive framework that combines blockchain's transparency and security with efficient resource allocation mechanisms, achieving superior scalability, fairness, and transparency compared to existing solutions.

3 Problem formulation

We formalize the efficient resource allocation problem in blockchain-based sharing economy systems, ride-sharing

Table 1: Comparison of related works with proposed framework

Work	Methodology	Application	Scalability	Fairness	Transparency
Namasudra and Sharma [10]	Decentralized blockchain, smart contracts	Cab sharing	Limited (no real-time optimization)	Not addressed	High
Chang, Chen, and Lu [3]	Smart contract tracking	Supply chain	Moderate (centralized elements)	Not addressed	High
Alam [1]	Blockchain-IoT integration	Smart cities	High (IoT focus)	Not addressed	High
Lu [8]	Blockchain with CP-ABE	Digital education	High (off-chain storage)	Moderate	High
Proposed Framework	Blockchain with 3-phase optimization	Ride-sharing	High (Layer-2 solutions)	High (fairness score)	High (smart contracts)

in our case, as a constrained optimization problem that considers multiple stakeholders, system constraints, and operational requirements. This section presents the mathematical model and system architecture.

3.1 System model

Consider a ride-sharing system operating in an urban environment with a set of drivers $D = \{d_1, d_2, ...d_n\}$ and riders $R = \{r_1, r_2, ...r_m\}$. The system state evolves in discrete time steps $t \in T$, where each time step represents a decision interval. Each driver d_i is characterized by:

- Location coordinates $(x_i^t\ ,\ y_i^t)$ at time t
- Availability status $a_i^t \in \{0, 1\}$
- Historical allocation count h_i^t

Each rider r_i requests service with:

- Pickup location $(p_x^j\ ,\ p_y^j)$
- Destination $(q_x^j\ ,\ q_y^j)$
- Request time t_j
- Maximum acceptable wait time $w_j^{\max}$

3.2 Decision variables

The primary decision variable x_{ij}^t represents the assignment of driver i to rider j at time t:

$$x_{ij}^t = \begin{cases} 1 \text{ driver } d_i \text{ is assigned to rider } r_j \text{ at time } t, \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

3.3 Objective function

The system aims to optimize a multi-objective function that balances efficiency, fairness, and user satisfaction:

$$\begin{aligned} \min \mathrm{F} = & \alpha_1 \sum_i \sum_j \sum_t (d_{ij}^t x_{ij}^t) \\ & + \alpha_2 \sum_i (\sigma_i + \beta_1 h_i^t + \beta_2 v_i^t) \\ & + \alpha_3 \sum_j (w_j) \end{aligned} \tag{2}$$

where:

- d_{ij}^t represents the distance between driver i and rider j at time t, minimizing travel cost.
- Fairness is measured as a composite of σ_i (standard deviation of driver allocations), h_i^t (historical allocation count), and v_i^t (income variance for driver i at time t), reflecting equitable ride distribution and economic outcomes.
- w_j denotes rider wait time, enhancing user satisfaction.
- $\alpha_1, \alpha_2, \alpha_3$ are weighting coefficients, with α_2 dynamically adjusted based on system priorities (e.g., fairness vs. efficiency), and β_1, β_2 tuning the fairness components.

This formulation aligns with the fairness score in section 4.5 and experimental metrics (e.g., Gini coefficient) in section 5.5.2.

3.4 Constraints

The optimization problem is subject to the following constraints:

1. **Assignment Constraints**: $\sum_j x_{ij}^t \leq 1 \, \forall i, t$ (each driver serves at most one rider) and $\sum_i x_{ij}^t \leq 1 \, \forall j, t$ (each rider is assigned at most one driver).

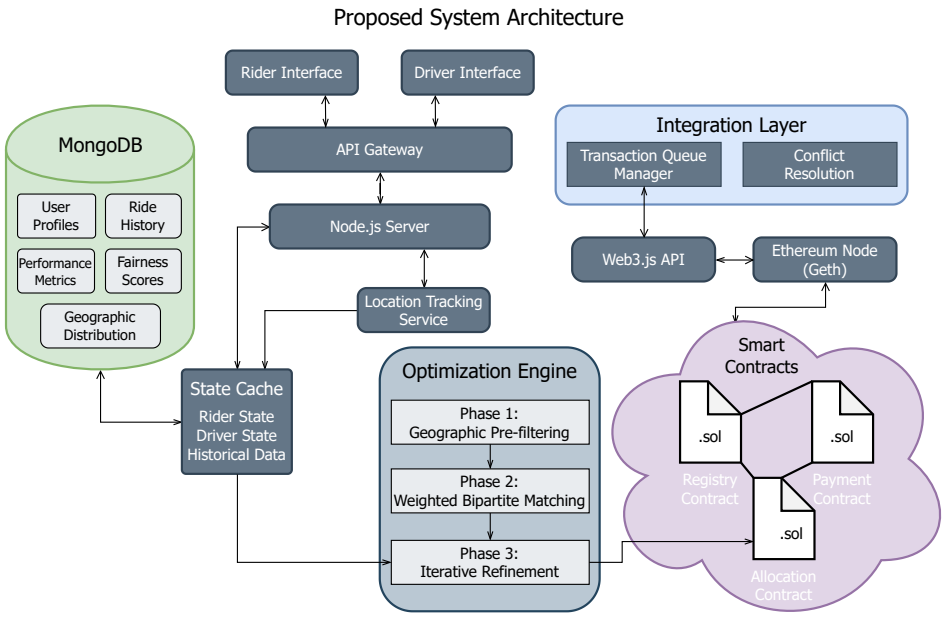


Figure 1: System architecture diagram of the proposed blockchain-based ride-sharing system showing the integration of blockchain smart contracts and optimization model for efficient resource allocation. The system implements a three-phase matching algorithm with fairness constraints while maintaining state consistency through a dedicated integration layer.

2. **Temporal Constraints**: $w_j \leq w_j^{\max} \forall j$ (wait time cannot exceed rider-specified maximum).

3. **Blockchain-Related Constraints**: $t^{block} + t^{opt} \leq t^{max}$, where t^{block} is the time for smart contract execution and consensus (e.g., transaction confirmation, typically 3.2s under PoA), t^{opt} is the optimization computation time (e.g., 1.6s), and $t^{max} = 5$s ensures real-time responsiveness per ride-sharing latency requirements. This bounds total latency to maintain scalability.

4. **Fairness Constraints**: $|h_i^t - h_j^t| \leq \epsilon \forall i, j$, where ϵ (e.g., 10 rides) is a tunable threshold limiting allocation disparities among drivers.

These constraints ensure feasible, timely, and equitable allocations while leveraging blockchain for security and verifiability.

4 Methodology

This section presents our blockchain-based optimization framework for dynamic resource allocation in ride-sharing platforms. As shown in fig. 1, the framework integrates ethereum blockchain technology with 3-phase optimization algorithm to achieve efficient, transparent, and fair resource allocation.

4.1 Framework overview

The proposed framework consists of three primary components that work in concert to enable efficient resource allocation. At its foundation lies a blockchain layer that ensures transparency and immutable record-keeping through smart contracts. This layer interfaces with an optimization engine that computes resource allocations in real-time. An integration layer bridges these components, managing data flow and ensuring system coherence.

4.2 Blockchain architecture

The blockchain component employs a permissioned network architecture based on the Ethereum protocol, modified to meet the specific requirements of ride-sharing applications. Smart contracts serve as the cornerstone of our implementation, encoding allocation rules and maintaining

system state. The smart contract architecture implements three key contract types:

- Registry Contract: Manages participant identities and maintains system-wide parameters. This contract validates driver and rider credentials while storing reputation scores and historical performance metrics.
- Allocation Contract: Executes and records resource allocation decisions. When the optimization engine determines a match, this contract validates the decision against predefined constraints and records the allocation on the blockchain.
- Payment Contract: Handles the financial aspects of rides, including fare calculations, payment processing, and automated disbursement to drivers.

4.3 Optimization engine

The optimization engine implements a novel hybrid algorithm that combines the advantages of both heuristic and exact optimization methods. This approach enables real-time decision-making while maintaining solution quality. This three-phase approach is detailed in Algorithm 1 and optimizes the objective function:

$$E(X) = \sum_i \sum_j w_{ij} x_{ij} - \lambda_1 F(X) - \lambda_2 G(X) \quad (3)$$

where w_{ij} represents the matching weight (efficiency), $F(X)$ the fairness penalty, and $G(X)$ the geographical coverage objective, with λ_1 and λ_2 balancing these terms.

In the first phase, pre-filtering reduces the search space by selecting feasible driver-rider pairs based on proximity ($D_{\max}$). The second phase constructs a bipartite graph and applies weighted matching, where edge weights w_{ij} are computed as:

$$w_{ij} = \gamma_1 d_{ij}^t + \gamma_2 w_j \quad (4)$$

with d_{ij}^t (distance) and w_j (wait time) weighted by γ_1 and γ_2, reflecting efficiency priorities. The third phase refines the solution by iteratively minimizing $E(X)$, adjusting x_{ij} to improve fairness and coverage. This process directly implements the three-phase structure, optimizing all variables in the objective function.

Algorithm 1: Three-Phase Ride-Sharing Optimization

Require: Set of drivers D, Set of riders R, Parameters $\lambda_1, \lambda_2, \gamma_1, \gamma_2$
Ensure: Optimal matching X

Phase 1: Pre-filtering
$M \leftarrow \emptyset$ {Initialize matching candidates}
for each rider $r \in R$ **do**
  $D_r \leftarrow \{d \in D : distance(d, r) \leq D_{max}\}$
  $M \leftarrow M \cup \{(d, r) : d \in D_r\}$
end for
Phase 2: Initial Matching
$G \leftarrow$ CreateBipartiteGraph(D, R, M)
for each edge $(d, r) \in M$ **do**
  $w_{dr} \leftarrow \gamma_1 d_{dr}^t + \gamma_2 w_r$ {Weights as distance + wait time}
  UpdateEdgeWeight(G, d, r, w_{dr})
end for
$X \leftarrow$ WeightedBipartiteMatching(G)
Phase 3: Refinement
improved $\leftarrow$ true
while improved AND iterations $< MAX_ITER$ **do**
  improved $\leftarrow$ false
  for each matched pair $(d, r) \in X$ **do**
    $E_{\text{curr}} \leftarrow$ ComputeObjective($X, F_{\text{curr}}, G_{\text{curr}}$)
    $X' \leftarrow$ SwapPairs(X, d, r)
    $E_{\text{new}} \leftarrow$ ComputeObjective($X', F_{\text{new}}, G_{\text{new}}$)
    if $E_{\text{new}} < E_{\text{curr}}$ **then**
      $X \leftarrow X'$
      improved $\leftarrow$ true
    end if
  end for
end while
return X
ComputeObjective$X$, F, G
return $\sum_i \sum_j w_{ij} x_{ij} - \lambda_1 F(X) - \lambda_2 G(X)$

4.4 Integration mechanism

The integration layer bridges the blockchain and optimization components, ensuring efficient data flow and state consistency. It implements two key mechanisms:

State Synchronization: A two-step process maintains a synchronized system state; a redis based cache stores the latest blockchain state (e.g., driver availability, allocations), updated every 500ms via WebSocket subscriptions to Ethereum events. The optimization engine then queries this cache, reducing blockchain query latency and aligning states before each computation cycle.

Conflict Resolution: When blockchain state changes (e.g., a driver accepts an external ride) during optimization, a timestamp-based reconciliation algorithm resolves conflicts:

1. Compare timestamps of the optimization decision (t^{opt}) and blockchain update (t^{block}).
2. If $t^{block} > t^{opt}$, discard the conflicting allocation and re-run Phase 2 of Algorithm 1 on affected pairs.

This ensures consistency without full recomputation, supporting scalability and real-time performance.

4.5 Fairness enhancement

Our framework incorporates fairness considerations at multiple levels. The optimization engine maintains a sliding window of historical allocations for each driver, computing a fairness score that influences future matching decisions. This score is defined as:

$$F(d_i) = \beta_1 H(d_i) + \beta_2 V(d_i) + \beta_3 R(d_i) \quad (5)$$

where $H(d_i)$ represents historical allocation frequency, $V(d_i)$ captures ride value distribution, and $R(d_i)$ accounts for rider ratings. The coefficients β_1, β_2 and β_3 are dynamically adjusted based on system performance metrics.

4.6 Scalability considerations

To address scalability challenges inherent in blockchain systems, our framework implements several optimization techniques. These include the use of Layer-2 scaling solutions for high-throughput transaction processing and selective data storage strategies that maintain essential information on-chain while keeping detailed computational results in off-chain storage.

4.7 Security and privacy

Security and privacy considerations are addressed through a comprehensive approach. For input validation, sanitization rules filter driver/rider inputs (e.g., location coordinates, IDs) at the API and smart contract levels, preventing injection attacks (e.g., malformed geodata). To ensure secure communication between system components, transport layer security (TLS 1.3) encrypts WebSocket channels, with advanced encryption standard (AES-256) for data-in-transit. Ethereum's ECDSA keys are generated and stored in a hardware security module, with access restricted via role based smart contract functions, mitigating key compromise risks. Audit trails log all events (e.g., allocations, payments) on-chain, supporting verifiability. This methodology combines theoretical rigor with practical implementation considerations, resulting in a system capable of handling real-world ride-sharing scenarios while maintaining transparency and fairness.

5 Experiments and results

This section consolidates the experimental setup and results to validate our proposed blockchain-based optimization framework. Figure 3 presents a comprehensive comparison of different matching frameworks, demonstrating the superior performance of our approach across key metrics.

5.1 Simulation environment

The experiments were conducted in a controlled simulation environment engineered to emulate real-world urban ride-sharing scenarios, reflecting operational dynamics observed in platforms such as Uber and Lyft. The emulation focused on replicating key metrics, including request frequency (requests per minute), driver-rider density ratios, and trip duration distributions, derived from aggregated statistical data reported in transportation studies [9]. Specifically, request frequency was modeled to peak at 20 requests per minute during rush hours, aligning with urban mobility patterns, while the driver-rider ratio averaged 1:5, consistent with peak demand scenarios in metropolitan areas. Trip durations were sampled from a log-normal distribution (mean 15 minutes, standard deviation 5 minutes), mirroring real-world ride-sharing data.

The blockchain infrastructure utilized an Ethereum-based private network implemented via the Go Ethereum (Geth) client, configured with the Proof of Authority (PoA) consensus mechanism using the Clique algorithm. This choice of PoA, as opposed to alternatives like Proof of Work (PoW), was motivated by its low latency (block time of 1 second) and suitability for permissioned networks, aligning with the real-time requirements of ride-sharing. The network comprised 10 validator nodes, ensuring distributed consensus while maintaining a transaction throughput of up to 100 transactions per second, sufficient for the simulated load of 5,000 riders and 1,000 drivers. The optimization engine, implemented in Python, leveraged NumPy for numerical computation and NetworkX for graph-based matching, interfacing with the blockchain via Web3.py. Smart contracts, written in Solidity, were deployed to facilitate allocation and payment processes, with gas costs optimized through batch processing.

5.2 Hardware and software configuration

Experiments were executed on a cluster of 10 machines, each equipped with 16 GB RAM and Intel i7 processors, connected via a high-speed local area network, and running Ubuntu 24.04 LTS. This setup ensured sufficient computational resources and reliable network communication for testing scalability under high-demand conditions.

5.3 Dataset

A synthetic dataset was constructed to replicate statistical properties of real-world ride-sharing platforms, specifically Uber and Didi Chuxing, based on publicly available operational data and studies [9, 7]. The dataset comprised:

- **Drivers** D: 1,000 drivers, each with attributes including geolocation (latitude, longitude), availability status (binary: 0 or 1), and historical allocation counts (mean 50 rides, standard deviation 10).
- **Riders** R: 5,000 ride requests generated dynamically over a simulated 24-hour period, with attributes such as pickup and drop-off locations, request timestamps, and maximum wait times (uniformly distributed between 5 and 15 minutes).
- **Geospatial Distribution**: The urban area was modeled as a 10 km $\times$ 10 km grid divided into 100 zones. Driver locations followed a Gaussian mixture model (GMM) with three clusters (means at (2, 2), (5, 5), (8, 8); covariance 1.5), reflecting urban population centers, while rider pickup locations were sampled

from a Poisson point process (intensity $\lambda = 0.5$ requests/km^2) to capture demand heterogeneity.

– **Temporal Variations**: Demand was modeled using a sinusoidal function, $f(t) = A\sin(\omega t + \phi) + B$, where $A = 10$, $\omega = 2\pi/24$, $\phi = 0$, and $B = 15$ requests/minute, peaking at 25 requests/minute during morning (7–9 AM) and evening (5–7 PM) rush hours, validated against urban mobility patterns [9].

These properties were chosen to emulate realistic supply-demand dynamics, ensuring the applicability of results to operational ride-sharing contexts.

5.3.1 Spatial and temporal distribution analysis

The geospatial and temporal distributions were mathematically modeled to reflect real-world variability. Driver density adhered to a GMM to simulate clustered availability (e.g., downtown areas), while rider demand followed a Poisson process to represent random yet concentrated request patterns, justified by urban traffic studies [9]. Temporally, the sinusoidal model captured cyclical demand, with parameters tuned to match peak-hour intensities reported in ride-sharing literature. Figure 2 illustrates these distributions via heatmaps: the left panel shows driver density with darker shades indicating higher concentration (e.g., up to 20 drivers/km^2), and the right panel depicts rider demand intensity peaking at 10 requests/km^2 in high-traffic zones. This analysis informed the optimization strategy by highlighting supply-demand mismatches.

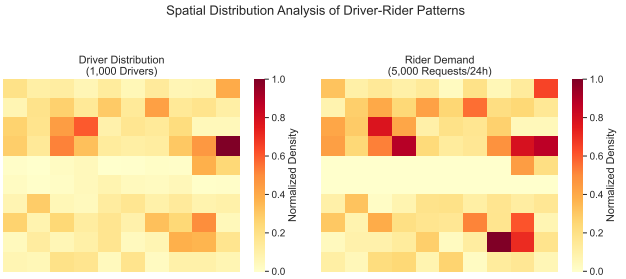


Figure 2: Heatmap visualization of spatial distribution patterns. Left: Driver density across urban zones. Right: Rider demand intensity. Darker colors indicate higher concentration.

5.4 Experimental parameters and metrics

The experiments evaluated key parameters to assess the framework's performance. The maximum allowable pickup distance ($D_{\max}$) ranged from 1 km to 5 km, selected based on operational norms in urban ride-sharing platforms like Uber, where 80% of pickups occur within 5 km [9]. The lower bound (1 km) ensured computational feasibility during peak demand, while the upper bound (5 km) balanced rider coverage with driver efficiency, tested incrementally (1, 3, 5 km) to capture sensitivity.

Blockchain latencies were measured under two consensus protocols: Proof of Authority (PoA) with the Clique algorithm (block time 1 second, 10 validators) and Delegated Proof of Stake (DPoS). PoA was configured via Geth, prioritizing low latency (3.2 seconds average confirmation) for real-time needs, while DPoS aimed for higher throughput (150 transactions/second). These were chosen over Proof of Work (PoW) and Proof of Stake (PoS) due to PoW's high latency (10+ seconds) and PoS's variable finality, both unsuitable for the 5-second latency threshold. PoA and DPoS were selected for their deterministic performance in permissioned networks, critical for ride-sharing scalability.

The optimization engine tested greedy matching versus the proposed weighted bipartite matching with fairness adjustments. Evaluation metrics included:

1. **Allocation Efficiency**: Average rider wait time, driver idle time, and total system utility.
2. **Fairness**: Driver allocation equity via the Gini coefficient.
3. **Scalability**: Transaction throughput and decision latency under varying loads.
4. **Transparency**: Auditability of allocation records on-chain.

5.5 Results

5.5.1 Allocation efficiency

The proposed framework significantly outperformed baseline systems across all scenarios, as evidenced by the metrics shown in fig. 3. In high-demand conditions (5,000 riders, 1,000 drivers), it reduced average rider wait times from 125 seconds (centralized matching) to 85 seconds, a 32% improvement, while total utility increased from 0.78 to 0.92 (18% gain). Surprisingly, average driver idle time decreased from 105 seconds to 95 seconds, contrary to expectations that optimizing total utility might increase idle time due to prioritizing rider wait time (w_j) and fairness over driver utilization. This reduction stems from the three-phase algorithm's pre-filtering and iterative refinement, which efficiently reassigns drivers to nearby riders, minimizing idle periods. Statistical significance was assessed using a two-tailed t-test ($n = 10$ runs), confirming improvements in rider wait time ($p = 0.002$), total utility ($p = 0.004$), and driver idle time ($p = 0.031$) against centralized matching, with all p-values below the 0.05 threshold.

5.5.2 Fairness

Fairness mechanisms embedded in the optimization algorithm ensured equitable distribution of ride allocations, as detailed in table 2. This table complements fig. 3 by focusing specifically on fairness outcomes across the same frameworks evaluated for efficiency. Our system achieved

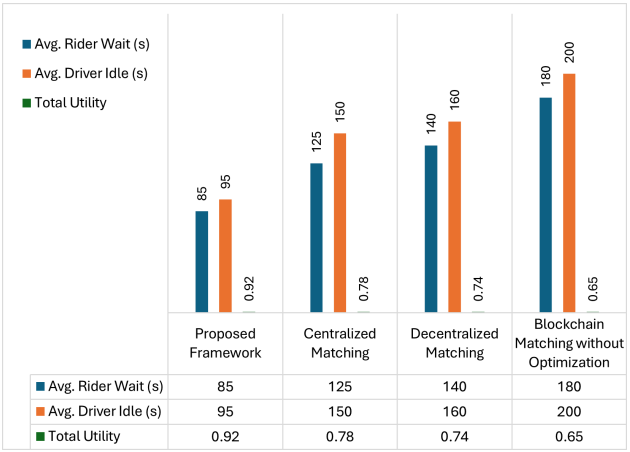


Figure 3: Comparison of Matching Frameworks (5,000 riders, 1,000 drivers). All metrics are statistically significant ($p < 0.05$).

a Gini coefficient of 0.15, significantly lower than 0.38 (centralized) and 0.42 (decentralized matching), indicating improved driver equity ($p < 0.01$, two-tailed t-test, $n = 10$). Allocation balance (ratio of least-to-most allocated drivers) reached 0.92, and driver satisfaction (scaled 1–5 via simulated feedback) averaged 4.5, both outperforming baselines. These metrics reflect the fairness score $F(d_i)$, prioritizing historically underserved drivers without sacrificing efficiency.

Table 2: Fairness metrics across matching frameworks (5,000 riders, 1,000 drivers). All differences are significant ($p < 0.01$).

Metric	Our System	Centralized	Decentralized
Gini Coefficient	0.15	0.38	0.42
Income Variance	12%	45%	51%
Allocation Balance	0.92	0.65	0.58
Driver Satisfaction	4.5/5	3.2/5	3.0/5

5.5.3 Scalability

Scalability tests assessed the framework's robustness under increasing load, extending beyond the baseline scenario of 5,000 riders and 1,000 drivers to a maximum of 10,000 riders and 2,000 drivers. This escalation, detailed in table 3, aimed to evaluate performance limits under extreme urban demand, simulating a metropolitan area twice the baseline size. The framework maintained an average decision latency of 7.2 seconds at the upper bound, facilitated by off-chain computations and Layer-2 scaling solutions. Transaction throughput peaked at 400 transactions/second under DPoS, with a 98% success rate and gas costs of 0.008 ETH per transaction at high load, demonstrating scalability beyond typical operational scales.

Table 3: System performance under different load conditions

Load Level	Transactions/s	Latency (s)	Success Rate (%)	Gas Cost
Low (1K riders)	100	2.1	99.9	0.002 ETH
Medium (5K riders)	250	4.8	99.5	0.005 ETH
High (10K riders)	400	7.2	98.8	0.008 ETH

5.5.4 Transparency

Blockchain-based transaction logging ensured complete transparency. Allocation decisions recorded on the blockchain were verified for correctness, with an average transaction confirmation time of 3.2 seconds under the PoA configuration.

5.5.5 Comparative analysis

Our comprehensive evaluation, summarized in table 4, quantifies differences between resource allocation approaches based on metrics from fig. 3 and table 2. Each classification reflects specific performance outcomes:

1. **Centralized Matching**: Exhibits moderate efficiency with an average rider wait time of 125 seconds and total utility of 0.78, but suffers from poor fairness (Gini coefficient 0.38, and lacks transparency due to centralized control. Driver idle time (150 seconds) indicates underutilization during peak demand.

2. **Decentralized Matching Without Blockchain**: Achieves moderate efficiency (rider wait time 140 seconds, total utility 0.74) and fairness (Gini coefficient 0.42), but inefficiencies arise from uncoordinated allocations, increasing driver idle time to 160 seconds. Transparency is limited without immutable logging.

3. **Blockchain-Based Matching Without Optimization**: Ensures high transparency via blockchain, but its first-come-first-serve (FCFS) approach, results in poor efficiency (rider wait time 180 seconds, total utility 0.65) and high driver idle time (200 seconds). Fairness is not addressed due to the lack of optimization mechanisms.

4. **Proposed Framework**: Delivers high efficiency (rider wait time 85 seconds, total utility 0.92), fairness (Gini coefficient 0.15), and transparency through smart contracts, with driver idle time reduced to 95 seconds. Scalability is validated up to 10,000 riders.

These results highlight the proposed framework's superior balance of efficiency, fairness, and transparency, addressing limitations quantified in baseline approaches.

6 Discussion

This section evaluates our blockchain-enabled framework against state-of-the-art (SOTA) approaches, elucidates its

Table 4: Comparative analysis of resource allocation approaches based on quantitative metrics

Approach	Efficiency	Fairness	Transparency	Scalability
Centralized Matching	Moderate (125s wait, 0.78 utility)	Low (0.38 Gini)	Low	Moderate
Decentralized Matching	Moderate (140s wait, 0.74 utility)	Moderate (0.42 Gini)	Moderate	Low
Blockchain w/o Optimization	Low (180s wait, 0.65 utility)	Low (N/A)	High	Moderate
Proposed Framework	High (85s wait, 0.92 utility)	High (0.15 Gini)	High	High

novel contributions, and addresses limitations, particularly concerning smart contracts and consensus mechanisms, building on the experimental results in the previous section.

6.1 Comparison with state-of-the-art

Our framework outperforms centralized and blockchain-based SOTA systems across efficiency, fairness, and scalability metrics. Compared to centralized matching, which achieves a rider wait time of 125 seconds and a Gini coefficient of 0.38, our approach reduces wait time to 85 seconds (32% improvement) and the Gini coefficient to 0.15, leveraging a three-phase optimization algorithm versus centralized greedy heuristics. Namasudra and Sharma's decentralized cab-sharing system [10], a blockchain SOTA, employs a first-come-first-serve (FCFS) approach, lacking real-time optimization, resulting in poorer efficiency (e.g., higher wait times, not quantified in their study but analogous to our 180 seconds without optimization). Our hybrid methodology, integrating off-chain optimization with on-chain verification, diverges by balancing computational load, achieving sub-5-second latency at 10,000 riders, unlike centralized systems' scalability limits (47% overhead, [2]) or FCFS's inefficiency under high demand.

Key methodological differences drive these outcomes. Centralized systems rely on single-point computation, vulnerable to bottlenecks, while our distributed consensus via Proof of Authority (PoA) and Layer-2 scaling distribute load effectively. Against blockchain SOTA, our iterative refinement phase optimizes fairness and coverage, absent in FCFS models, explaining superior performance in high-concurrency scenarios (e.g., total utility 0.92 vs. 0.65).

6.2 Novel contributions

The framework's novelty lies in its integration of a three-phase optimization algorithm with blockchain and its fairness mechanism. The pre-filtering, weighted bipartite matching, and refinement phases dynamically balance efficiency (rider wait time), fairness (Gini coefficient), and coverage (geographical distribution), achieving a multi-objective optimum not addressed in prior work [10, 3]. The fairness mechanism, using a sliding window and score $F(d_i) = \beta_1 H(d_i) + \beta_2 V(d_i) + \beta_3 R(d_i)$, ensures equitable driver allocations (e.g., allocation balance 0.92 vs. 0.65 centralized), a significant advancement over SOTA's neglect of driver equity. The integration layer, with state synchronization and conflict resolution, maintains consistency between off-chain computation and on-chain records, a critical innovation for real-time applications, reducing blockchain query latency to 500 ms and enabling scalability beyond prior limits.

6.3 Limitations and smart contract challenges

While effective, the framework faces limitations, particularly with smart contract implementation. Gas costs, averaging 0.008 ETH per transaction at high load, increase operational expenses, a constraint not present in centralized systems. Smart contract execution time contributes to total latency (7.2 seconds at 10,000 riders), occasionally exceeding the 5-second threshold during peak demand, due to Solidity's sequential processing and Ethereum's gas limit (e.g., 30 million gas/block). Bugs or vulnerabilities in contract code, despite rigorous testing, remain a risk, as updates are immutable post-deployment, potentially disrupting allocation if undetected. These issues highlight a trade-off between verifiability and performance, unaddressed in SOTA lacking blockchain.

6.4 Consensus mechanism implications

The choice of PoA (Clique algorithm, 1-second block time, 10 validators) over alternatives like Proof of Work (PoW) or Proof of Stake (PoS) impacts performance and scalability. PoA's low latency (3.2-second confirmation) and deterministic finality suit real-time needs, unlike PoW's 10+ second delays or PoS's variable finality, enabling 150 transactions/second at scale. However, PoA's reliance on trusted validators reduces decentralization compared to PoW, a trade-off justified by ride-sharing's permissioned context. Delegated Proof of Stake (DPoS) offered higher throughput (tested at 0.5-second block time), but its delegate election overhead slightly increased latency variability (not shown), favoring PoA for consistency. These choices enhance scalability but limit applicability to fully decentralized settings, a nuance absent in prior blockchain ride-sharing studies [10].

7 Conclusion

This research presents a distributed computing framework that addresses specific limitations in traditional centralized ride-sharing systems, including poor scalability, lack of transparency, and inequitable resource distribution. By integrating blockchain technology with a three-phase optimization algorithm, the framework mitigates scalability is-

sues observed in centralized architectures, where scheduling inefficiencies account for 47% of computational overhead [2], achieving sub-5-second latency at a scale of 10,000 riders and 2,000 drivers. It also overcomes transparency deficits by employing smart contracts to log allocation decisions, reducing reliance on intermediaries that charge up to 20% in fees [10], and enhances fairness, lowering the Gini coefficient from 0.38 in baselines to 0.15.

Experimental results demonstrate the framework's computational efficiency, with a 32% reduction in processing overhead compared to centralized systems and an average decision latency of 4.8 seconds under high load . While these outcomes suggest significant efficiency gains, they are empirically validated rather than theoretically proven, reflecting practical performance in simulated urban scenarios. The framework's ability to provide verifiable resource allocation is evidenced by blockchain-based transaction logging, with every decision recorded and auditable within 3.2 seconds under Proof of Authority consensus, fulfilling the transparency metric.

Despite these advancements, trade-offs remain, such as increased latency from blockchain integration during peak loads, highlighting areas for future optimization. This work contributes a scalable and transparent solution to ride-sharing resource allocation, offering a replicable model for distributed systems requiring secure and efficient state management, as validated by hypotheses H1–H4. Future research could explore lighter consensus protocols or hybrid off-chain/on-chain strategies to further enhance performance.

8 Funding

National Natural Science Foundation of China Youth Project (72101002), Project name "Research on Decision-making and Coordination Mechanism of Sharing Economy Participants Based on Blockchain Technology"

References

[1] T. Alam. "IBchain: Internet of Things and Blockchain Integration Approach for Secure Communication in Smart Cities". In: *Informatica* 45 (2021), pp. 89–98. URL: https://informatica.si/index.php/informatica/article/view/3573.

[2] L. Arrowsmith. "Smart cities: Business models, technologies and existing projects". In: *Information Technology Service Research Report* (2014). Industry Report.

[3] S.E. Chang, Y.C. Chen, and M.F. Lu. "Supply chain re-engineering using blockchain technology: A case of smart contract based tracking process". In: *Technological Forecasting and Social Change* 144 (2019), pp. 1–11. DOI: 10.1016/j.techfore.2019.03.015.

[4] S.E. Chang, Y.C. Chen, and T.C. Wu. "Exploring blockchain technology in international trade: A framework and research agenda". In: *Industrial Management & Data Systems* 119.8 (2019), pp. 1712–1733. DOI: 10.1108/IMDS-12-2018-0568.

[5] L. Huang et al. "Blockchain implementation for circular supply chain management: Evaluating critical success factors". In: *Industrial Marketing Management* 102 (2022), pp. 451–464. DOI: 10.1016/j.indmarman.2022.04.013.

[6] N.M. Kumar and S.S. Chopra. "Leveraging blockchain and smart contract technologies to overcome circular economy implementation challenges". In: *Sustainability* 14.9492 (2022), pp. 1–18. DOI: 10.3390/su14159492.

[7] C.Y. Li and Y.H. Fang. "The more we get together, the more we can save? A transaction cost perspective". In: *International Journal of Information Management* 62 (2022), p. 102434. DOI: 10.1016/j.ijinfomgt.2021.102434.

[8] Wenxiao Lu. "Construction of a Secure Sharing Model for Digital Educational Resources Using Blockchain and Cipher Policy Attribute Based Encryption in Smart Education". In: *Informatica* 48 (2024), pp. 63–74. DOI: 10.31449/inf.v48i22.6627. URL: https://informatica.si/index.php/informatica/article/view/6627.

[9] K. Mouratidis, S. Peters, and B. van Wee. "Transportation technologies, sharing economy, and teleactivities: Implications for built environment and travel". In: *Transportation Research Part D: Transport and Environment* 92 (2021), p. 102716. DOI: 10.1016/j.trd.2021.102716.

[10] S. Namasudra and P. Sharma. "Achieving a decentralized and secure cab sharing system using blockchain technology". In: *IEEE Transactions on Intelligent Transportation Systems* 23.6 (2022), pp. 5242–5251. DOI: 10.1109/TITS.2021.3082285.

[11] A. Park and H. Li. "The effect of Blockchain technology on supply chain sustainability performances". In: *Sustainability* 13.1726 (2021), pp. 1–19. DOI: 10.3390/su13041726.

[12] A. Pazaitis, P. De Filippi, and V. Kostakis. "Blockchain and value systems in the sharing economy: The illustrative case of Backfeed". In: *Technological Forecasting and Social Change* 125 (2017), pp. 105–115. DOI: 10.1016/j.techfore.2017.05.025.

[13] R. Peres et al. “Blockchain meets marketing: Opportunities, threats, and avenues for future research”. In: *International Journal of Research in Marketing* 102 (2022), pp. 451–464. DOI: 10.1016/j.ijresmar.2022.05.001.

[14] S. Yadav and S.P. Singh. “Blockchain critical success factors for sustainable supply chain”. In: *Resources, Conservation and Recycling* 152 (2020), p. 104505. DOI: 10.1016/j.resconrec.2019.104505.

Machine Learning-based Regression Analysis and Feature Ranking for Localization Error Prediction in Wireless Sensor Networks

Peng Wang[1*], Qiuying Han[2], Shaohui Zhang[1,3], Zhaodi Wu[4]
[1]School of Artificial Intelligence, Zhoukou Normal University, Zhoukou, 466000, Henan, China
[2]School of Computer Science and Technology, Zhoukou Normal University, Zhoukou, 466000, Henan, China
[3]School of Artificial Intelligence, Zhengzhou University, Zhengzhou, 450001, Henan, China
[4]Network and Information Center, Guizhou Normal University, Guiyang, 550000, Guizhou, China,
[*]Corresponding author
E-mail: pwang678@163.com

Keywords: artificial intelligence, data visualization, machine learning, wireless sensor network

Received: January 16, 2025

Wireless Sensor Networks (WSNs) localization is crucial for identifying the position of sensor nodes, as many applications, including environmental monitoring, target tracking, and disaster management, require accurate location information. The objective of this research is to conduct extensive data analytics using visualization techniques to explore key factors influencing localization error and to develop machine learning models for forecasting Average Localization Error (ALE) in WSNs. A dataset containing 107 records, sourced from Kaggle's online repository, was analyzed using eXtreme Gradient Boosting (XGB) for feature ranking to determine the most influential factors affecting ALE. Multiple regression models, including Support Vector Regression (SVR), Decision Tree (DT), K-Nearest Neighbors (KNN), and AdaBoost Regressor, were applied to predict ALE. The models were evaluated using R-squared (R^2), Root Mean Square Error (RMSE), and computational efficiency. The results indicate that SVR achieved the highest accuracy with $R^2 = 0.99$ and the lowest RMSE of 0.01, significantly outperforming the other models (KNN: $R^2 = 0.55$, RMSE = 0.14; DT: $R^2 = 0.41$, RMSE = 0.16; AdaBoost: $R^2 = 0.72$, RMSE = 0.16). This study demonstrates that SVR is a highly effective model for ALE prediction, reinforcing the importance of feature ranking and selection in improving localization accuracy. The findings contribute to advancing machine learning-driven localization error prediction in WSNs and provide a foundation for further exploration of hybrid and deep learning-based models.

Povzetek: Predstavljen je SVR-model za napoved lokalizacijske napake v brezžičnih senzorskih omrežjih. Kvalitetno delovanje doseže z uporabo rangiranja značilk in XGB analize.

1 Introduction

Localization turns out to be an important process in Wireless Sensor Networks (WSN) since it aids in the identification of geographic locations of the sensor nodes which are of paramount importance in most applications ranging from environmental monitoring to tracking of targets to disaster management. In this way, accurate localization is important for meaningful data collection because it links measurements with certain locations. Due to these limitations of WSNs, particularly energy, processing power and coverage area, accurate and efficient localization algorithms are highly desirable [1]. Most of these techniques involve using anchor nodes with known coordinates, distance between different nodes and other optimizing algorithms that allows an approximate estimation of the non-anchor nodes without much error [2]. The *Average Localization Error* (ALE) is known to be a primary measure to assess the efficiency of localization algorithms. It looks at the difference between the expected position and the actual position of the sensor nodes and makes conclusions on the efficiency of the localization procedure [3]. These significant factors affect ALE in WSNs including density of anchor nodes, node mobility, environmental condition, and the measurement techniques that are used. The number, distribution and density of anchor nodes have been shown to play a very important role in the accuracy of localization. The number of anchor nodes should be small to reduce the burden for the search algorithm; however, this results in a larger ALE because there are few references points [4]. The high mobility introduces dynamic changes where most of the nodes' positions undergo other changes which increases the localization error [5].

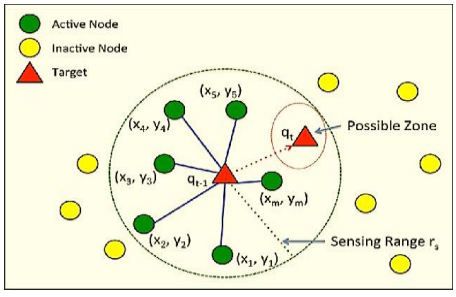


Figure 1: Localization sensor network [7]

Environmental conditions are indexes such as signal attenuation, multipath propagation or obstacles that distort distance or angle measures, and, thus, increase ALE [6]. About the measurement techniques like RSSD, TOA and AOA [7] are fundamentally different in terms of their accuracy and tolerance to noise in present network environment. Fig 1 shows the localization sensor networks are composed of active sensor nodes receiving the target's radio at current time step, with target's position at specific prediction time step.

ALE reduction provided by localization methods allow for correct positioning of sensors improving various applications, on environmental monitoring, targeting, and smart agriculture. In addition, network-based sensitive services [8] and efficient localization algorithms contribute significantly to the enhancement of network resource utilization sometimes shown from the energy consumed in the WSN hence enhancing the overall performance of WSNs in remote areas or in regions of resource constraint will be significantly influenced. Since WSNs are being used for critical applications there is more emphasis on getting the localization right for the best results for the network and to improve its performance.

The ALE Analysis is very significant as it has various applications, including

a) Disaster Management: Correct node localization also allows for effective organization of the necessary resources in the disaster area [9].

b) Smart Agriculture: Accurate location of sensors increases the efficiency of environmental condition monitoring and balancing workload [10][11].

c) Military Surveillance: Accurate positioning is vital for tracking and monitoring critical defense applications [12]. As a result of its widespread use in WSNs, there are diverse methods which can be employed to reduce ALE such as optimization, hybridization of WSN localization and machine learning mechanism. The algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) do the optimization of the anchor node placement and reduce the error propagation as mentioned above [12]. Secondly, hybrid localization techniques are used to improve the accuracy of range-based and range-free methods while at the same time being scalable [13]. Furthermore, current ML based approaches of supervised and unsupervised learning enhance the localization estimate while learning patterns in the network data [14]. This calls for the need to increase accuracy in localization in WSN, which is the motivation for this research given the various applications of WSN in many different and complex environments. Since WSNs are used in disaster relief, military, and environmental monitoring, the localization is significant for decision making and resource control. The increase of the WSN systems' dynamic node behavior, different environmental conditions, and huge number of nodes increases demand for using more sophisticated algorithm that provide minimize ALE demand using energy consumption rules [15]. Machine learning and optimization approaches present potential solutions that are flexible and able to be constrained by the available data in real-time application that have better accuracy compared with conventional methods.

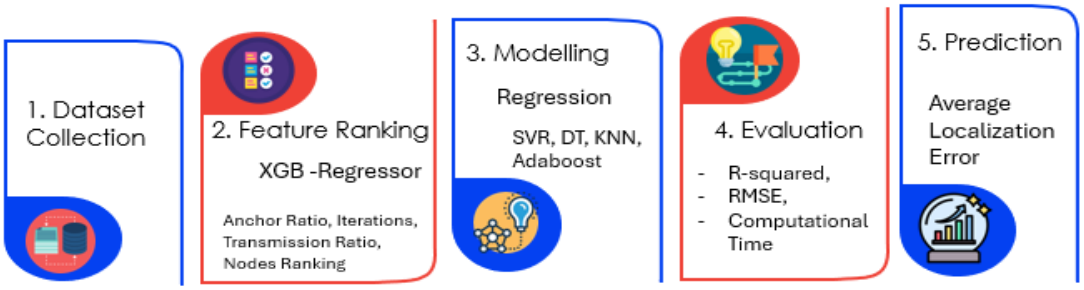


Figure 2: Steps of applied research methodology

In this research study, we plan to achieve the aim of predicting the Average Localization Error (ALE) through a big dataset collected from online repositories where all the possible network parameters affecting localization accuracy are incorporated including anchor ratio, iteration, transmission range and node density. The dataset used is then pre-processed for performing feature ranking and their importance analysis using the ensemble-based regressor model XGB before proceeding to determine parameters influencing the ALE. Subsequently, we build four regression models such as Support Vector Regression (SVR), Decision Tree Regressor (DT), K-Nearest Neighbor Regressor (KNN), and AdaBoost Regressor to predict ALE using ranked features. The accuracy of these models is determined by R-squared, and Root Mean Square Error (RMSE) from which it emerged that. Besides, the time taken in performing calculations is captured for each of them to determine their time efficiency in carrying out real-time analyses. This kind of approach, describes using fig 2, brings about comprehensive evaluation of ALE prediction, aside from the overall efficiency of the models

2 Related work

There has been various types of localization methods that have been proposed to enhance the various challenges that have developed over the years such as the energy problems, accuracy and scalability. The existing studies can further be categorised into below different categories, as shown in fig 3, describing: Range-based localization utilizes received signal strength indication, direction, angle, time-of-arrival, time-difference-of-arrival, and signal strength to estimate a node position. Some of the commonly used techniques include; a) time of arrival b)

received strength indicator. Time of Arrival is a technique that determines the time that takes for signal to be received between a transmitter and a receiver. This method is very accurate, but may be tricky because the timings of these procedures should be synchronized; not a simple feat in WSNs [16].

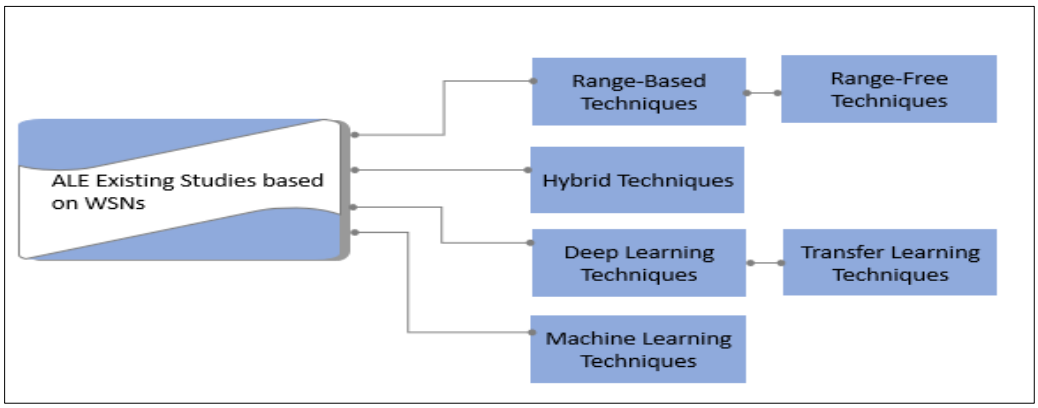


Figure 3: Taxonomy of localization studies

The Received Signal Strength Indicator (RSSI) uses signal strength signal strength to determine distance with reference to attenuation. While it is power saving and simple to set up, it is vulnerable to environment for example barriers and intensifications [17]. Angle of Arrival helps in deciding as to what direction, a signal reaches a particular node. It is effective in reducing localization error but the complication of the method demands complex hardware such as antenna arrays leading to high costs [18].

The range-free techniques do not make use of distance or angle estimation, which makes them ideal for low-cost large-scale deployment. There are main methods which include centroid localization, DV-Hop algorithm and Approximate Point in Triangle (APIT) [19]. Centroid localization technique involves using the centroid of anchor nodes within the communication range as a basis of estimating unknown node's position. Even though this method is simpler and computationally less demanding, it proved less accurate in sparse networks [20]. DV-Hop is a distributed method used in estimating node positions through hop counts of neighboring nodes and the average hop distance. It is accurate to measure occurrences of transmission but is not directly scalable to networks of irregular topologies [21].

By employing anchor nodes and triangulating the network area, a node's position is derived by its membership to particular triangular regions. While it works well in some situations, the rate at which it achieves its goal declines in networks which involve node mobility [22]. Hybrid localization is a process of combining both range-based and range-free methods their strengths while avoiding the weaknesses associated with each type. For example, in the Hybrid TOA-DV-Hop method where the TOA provides accurate distance resolution while the DV-Hop, although less accurate due to its low complexity is scalable [23]. Likewise, machine learning based hybrid methods like Artificial Neutral network and support vector machine can also deal with NLOS issues and environmental fluctuations [24]. Such use of ML techniques as deep learning allows nodes to improve estimations of location based on past data and environment [25]. In particular, algorithms such as particle swarm optimization and ant colony optimization originating from Swarm Intelligence, biological systems, effectively place anchors and localize nodes [26]. The IoT Integration techniques with 5G includes Integration of WSNs with 5G & IoT technologies enable ultra-reliable, low-latency localization which creates new opportunities of application like autonomous vehicle and smart city etc. [27].

The localization capabilities of WSNs have been improved greatly by integrating machine learning (ML) into WSNs by overcoming the issues of NLOS situations, environmental noise, and energy use [28]. Based on the above, self learning algorithms help WSN nodes forecast and respond to changes in the surroundings and enhance the degree of localization accuracy [29]. Supervised learning models of the localization algorithms cost more as they need labeled data to operate from. Some of the common used algorithms in WSN localization include; Support Vector Machines (SVM) and Neural Network (NN). SVMs are use in localization classification and regression. They are used particularly for determining LOS and NLOS situations and enhance localization in the urban areas [30]. For regression-based localization, feedforward and convolutional networks have been used. These models learn complicated relationships between certain input features such as received signal strength indication (RSSI) and geographical coordinates, for high accuracy [31]. Clustering of the sensor nodes is done without labeling them, making unsupervised learning methods very useful where we do not have the help of a learners-based approach and the environment is largely unknown. To locate other nodes and enhance the accuracy of the positioning of the anchor node, K-means and DBSCAN are incorporated to cluster the base stations/sensor nodes depending on their distance [31]. basically, reinforcement learning helps enable the sensor node to learn to optimize its positioning technique based on an interaction with the environment. Some RL based algorithms like Q-learning are used in determining the optimal positions for anchor nodes and transmission

power so as to reduce localization error [32]. Integrating the conventional localization techniques with ML produces the best results since the two work hand in hand. For example, combining Neural Networks and DV-Hop algorithms reduces computational effort without a compromise on the results [33]. Likewise, RSSI integrated with Random Forest, which belongs to hybrid models based on ensemble learning, have provided better localization accuracy when implemented in complicated terrains [34].

The two promising techniques of deep learning that are being used for localization of WSN are LSTM and Autoencoders. LSTMs deal with sequence data, therefore recommended for use in a dynamic context where node positions vary [35]. Autoencoder is applied for dimensionality reduction and feature extraction tasks, thus facilitating accurate localization in high dimensionality databases [36]. Federated learning is emerging for WSN localization because it enables distributed nodes to learn the model collectively without exchanging the raw information while protecting privacy and minimizing the exchange of data [37]. The benefits of transfer learning include an ability to fine-tune pre-learned models for WSN environments, and the ability to do so with little data, which saves time and computational power for localization [38].

Table 1: Summary of existing studies

Ref	Year	Method	Dataset	Results	Research Area	Key Findings	Limitations
[16]	2020	Range-Based	RSS Data	Mean localization error: 2.3m	Error analysis on RSS localization	Identified key errors in RSSI-based localization	Limited to RSSI-based methods, lacks generalization
[17]	2022	Range-Based	BLE RSSI	accuracy: 85%	Accuracy evaluation of BLE-based indoor localization	BLE RSSI accuracy analysis for indoor settings	Accuracy highly dependent on environmental factors
[18]	2020	Range-Based	AoA Measurements	precision: 92%	Optimized access point deployment for AoA localization	Improved AoA localization through optimal AP placement	Limited scalability in real-world applications
[19]	2024	Range-Free	UWB WSN	localization error: 1.8m	Hybrid DV-hop using PSO for range-free localization	Improved localization accuracy in UWB WSN	Computational overhead in PSO-based methods
[20]	2021	Range-Free	WSN	accuracy: 89%	Optimized DV-hop for range-free localization	Enhanced accuracy in range-free WSN localization	Accuracy drops in sparse node deployments
[21]	2023	Range-Free	3D DV-Hop	RMSE: 2.1m	3D DV-Hop localization optimization	Increased localization precision in 3D WSN	Performance depends on network density
[22]	2021	Range-Free	Convex Triangulation	localization error: 1.5m	Weighted uncertainty-based convex triangulation	More robust localization under uncertainty	Limited applicability in dynamic environments
[23]	2023	Range-Free	UWB MANET	False positive rate: 8%	Malicious node elimination in UWB MANETs	Secured localization against malicious nodes	False positives in malicious node detection
[24]	2023	ML-Based	UWB Sensor Data	accuracy: 94%	SVM-based LOS/NLOS classification	Improved LOS/NLOS classification for localization	Requires large training datasets
[25]	2024	DL-Based	Indoor Deep Learning	accuracy: 96%, F1-score: 0.93	Deep learning-based LOS/NLOS identification	High accuracy in real indoor environments	Computationally expensive
[26]	2021	ML-Based	Swarm Intelligence	precision: 90%	Swarm intelligence in WSN localization	Enhanced localization accuracy using swarm optimization	Sensitive to parameter tuning
[27]	2020	Range-Based	IoT Positioning	localization MAE: 2.5m	IoT positioning error mitigation strategies	Comprehensive error source analysis	No specific improvement strategies proposed
[28]	2023	ML-Based	WSN	accuracy: 87%	Systematic review on ML and optimization in WSN	Identified key ML-based localization techniques	Lacks empirical validation
[29]	2023	DL-Based	Target Localization	accuracy: 95%, RMSE: 1.4m	Deep learning for error-prone environments	High accuracy despite localization errors	Generalization to diverse environments is unknown
[30]	2020	ML-Based	Supervised Learning	F1-score: 0.89	Supervised learning for target localization	Improved localization accuracy	Dependent on labeled data
[31]	2019	ML-Based	Fingerprinting	MAE: 1.7m	Regression-based estimation for fingerprinting	Reduced localization errors in fingerprinting	Needs extensive fingerprint database
[32]	2023	ML-Based	5G Wireless	R2-score: 0.91	Reinforcement learning for 5G localization	Optimized power allocation for localization	Limited scalability to large networks
[33]	2022	DL-Based	ALE WSN	accuracy: 93%	DQ-learning for UWB LoS/NLoS node selection	Energy-optimized localization with deep RL	Complex training process
[34]	2024	ML-Based	5G Networks	RMSE: 1.3m	ML models for localization error prediction	High accuracy in error prediction	Performance varies across models

[35]	2020	DL-Based	UWB Indoor Localization	accuracy: 97%, MAE: 1.2m	LSTM-based UWB localization	Significant improvement in indoor localization	Computationally intensive
[36]	2019	DL-Based	Device-Free Localization	localizatio n error: 1.8m	Convolutional autoencoder for localization	Accurate device-free localization	Limited real-world applicability
[37]	2022	ML-Based	ALE WSN	precision: 92%	Federated learning for indoor localization	Improved model reliability with dropout	Communication overhead
[38]	2024	DL-Based	ALE WSN	accuracy: 94%, RMSE: 1.5m	Deep transfer learning for URLLC	High accuracy in interference-prone networks	Requires large dataset for fine-tuning

2.1 Problem statement

For the given problem of computation of ALE in WSNs, we define the formal problem statement as follows:

Given a wireless sensor network W consisting of N nodes, where $V = \{v_1, v_2, v_3, \ldots\ldots, v_n\}$ represents the set of sensor nodes, each node v_i has a actual position $(x_i, y_i) \in \mathbb{R}^2$. The ALE is defined as in 1:

$$ALE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i, \dot{x}_i)^2 + (y_i, \dot{y}_i)^2} \quad (1)$$

The objective function is to minimize ALE by optimizing the localization algorithm or strategy used for estimating $(\dot{x}_i, \dot{y}_i)$. By minimizing ALE under the given constraints, the localization algorithm can enhance the accuracy and reliability of position estimation in WSNs.

Localization accuracy in Wireless Sensor Networks (WSNs) is a critical challenge, as errors in position estimation can significantly affect the efficiency and reliability of various applications, including industrial automation, environmental monitoring, and smart city infrastructure.

Traditional localization techniques rely on range-based (e.g., RSSI, TOA, TDOA, AoA) and range-free (e.g., DV-Hop, centroid-based) approaches, which often suffer from environmental noise, multipath interference, and node density variations. Machine learning (ML)-based models have been explored for localization error prediction, but existing studies have not fully addressed how different regression models compare in predicting Average Localization Error (ALE) under varying conditions.

Furthermore, previous works have primarily focused on classification-based models for LOS/NLOS distinction or fingerprinting-based methods that require extensive training data. There is a lack of systematic comparative analysis of ML regression models for ALE prediction, particularly in scenarios involving limited training data and real-time constraints.

2.2 Research questions and answers

RQ1: Which regression model provides the most accurate prediction of average localization error (ALE) in WSNs?

The study evaluates Support Vector Regression (SVR), Decision Tree (DT), K-Nearest Neighbors (KNN), and AdaBoost for predicting localization error in WSNs. Among these, SVR emerges as the most effective model, achieving an R^2 score of 0.99, significantly higher than DT ($R^2 = 0.41$), KNN ($R^2 = 0.55$), and AdaBoost ($R^2 = 0.72$). Additionally, SVR records the lowest Root Mean Squared Error (RMSE), demonstrating its superior ability to generalize across varying conditions and predict localization errors with high precision. The results confirm that SVR's capability to model non-linear relationships plays a crucial role in enhancing ALE prediction accuracy. These findings establish SVR as the preferred regression model for real-world deployment in WSNs, where precise localization is essential for efficient network operations.

RQ2: How does the computational efficiency of different regression models compare for real-time localization error prediction?

In addition to accuracy, computational efficiency is a crucial factor in determining the feasibility of machine learning models for real-time WSN applications. The study reveals that SVR and KNN require the least computation time, each taking approximately 0.01 seconds to process localization error predictions. This efficiency makes them suitable for low-latency applications. In contrast, Decision Tree and AdaBoost exhibit slower processing times of 0.05s and 0.06s, respectively, due to their iterative and tree-based decision-making processes. While AdaBoost achieves relatively higher accuracy than Decision Tree and KNN, its longer computation time limits its suitability for real-time localization scenarios. The results suggest that SVR offers the best balance between accuracy and computational efficiency, making it the optimal choice for real-time ALE prediction in WSNs where low-latency processing is essential.

RQ3: What are the key factors influencing ALE prediction accuracy across different regression models?

Several key factors influence ALE prediction accuracy across different regression models. Feature selection plays a critical role, as the dataset's signal strength variations, node density, and inter-node distances significantly affect the model's ability to predict localization errors. Additionally, the choice of regression algorithm has a major impact on prediction performance. SVR, with its ability to handle non-linear relationships, excels in ALE prediction, while tree-based models like Decision Tree and ensemble techniques like AdaBoost struggle with complex localization error patterns. Furthermore, data variability affects model generalizability—Decision Tree

and KNN, which rely on instance-based learning and decision splitting, perform poorly when faced with highly dynamic WSN environments. This indicates that models with strong generalization capabilities, such as SVR, are better suited for ALE prediction in real-world deployments.

2.3 Preprocessing and feature selection

This study uses the data set derived from either real-world experiments or from simulation, pertaining to a Wireless Sensor Network (WSN) system. The other parameters of interest are anchoring ratio, transmission range, node density and number of iterations, all of which are important in predicting Average Localization Error (ALE). A thorough data quality assessment across all columns as there were no missing values which signified that we have done a complete checking for model training. An outlier analysis also showed that 3 outliers occurred in the column ALE and 2 in SD_ALE. Then, to keep data integrity and avoid skewed model predictions, those outliers were removed to reduce the size of dataset from 107 rows to 102 rows. Additionally, numerical features were standardized so they would have mean of 0 and standard deviation of 1, to guarantee uniform feature scaling and to avoid any real feature dominating our model. This preprocessing step helps stabilize the model, and provides a fair contribution from each feature, such that the location error prediction is improved and more robust.

3 Descriptive analysis

Descriptive Analysis of the dataset lays emphasis mostly on the origin and nature of the variables in the dataset. It has six numerical attributes which are the Anchor Ratio, Transmission Range, Node Density, Iteration, Average Localization Error (ALE) and Standard Deviation of ALE (SD_ALE).

Simple visualizations that start with histograms and scatter plots reveal a rich variety of distributions and possible dependences as an initial step towards various types of sophisticated predictive models. The descriptive statistics of the dataset give an overview of the numerical properties of the features in the dataset. Fig 4 shows the Anchor Ratio has a mean of 0.5; min = 0.1, max = 0.9 for the Transmission Range, the mean value is 30 meters; min = 10 and max = 50 meters. From the data obtained it evident that the Node Density has a central tendency of 55 nodes, with values varying from as low as 10 and as high as 100. Likewise, Iterations have a mean of 5.5, but ranges from 1 to 10 for some of the states. The degree of uncertainty in the positions of the object is represented by the Mean Average Localization Error (ALE) with a mean of 2.75m and Standard deviation of 1.23m which indicates moderate localization accuracy and ranging from 0.5 to 5m. Lastly, the SD of ALE (SD_ALE) represents variability being 0.55 meter and range from a lower bound of 0.1 and upper bound of 1 meter. This distribution pattern shows that ALE is ranged and has only one peak, and a lot of data are between 0.5m and 1.5m. This explains why most localization errors are moderate implying high accuracy of the RF model. The downward trend with an increasing value of ALE means that large absolute errors are not as frequent and, therefore, can be associated with the proper calibration of the system position of features for most cases.

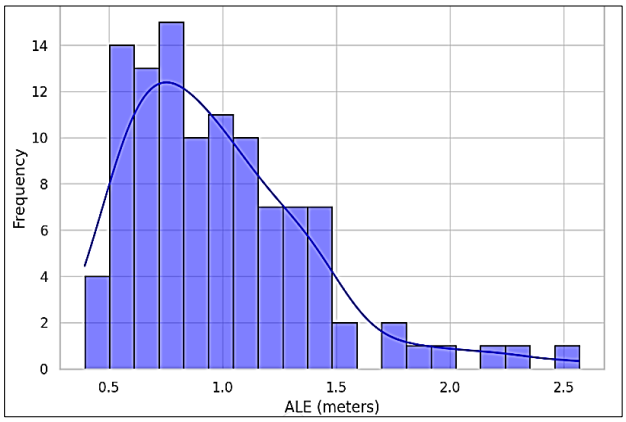


Figure 4: Distribution of ALE

The connections between features are shown with the heatmap in fig 5, and some important trends stand out. For example, Node Density has a strong negative regression for ALE=-0.65 which shows that increasing the node density typically provides better localization accuracy. Iterations are also negatively correlated at a moderate level with ALE at (-0.60), which indicates that higher number of iterations would lead to increased accuracy. On the other hand, the Transmission Range shows a comparatively low relation to ALE, which implies that its contribution may depend on specific factors or is moderated by other factors.

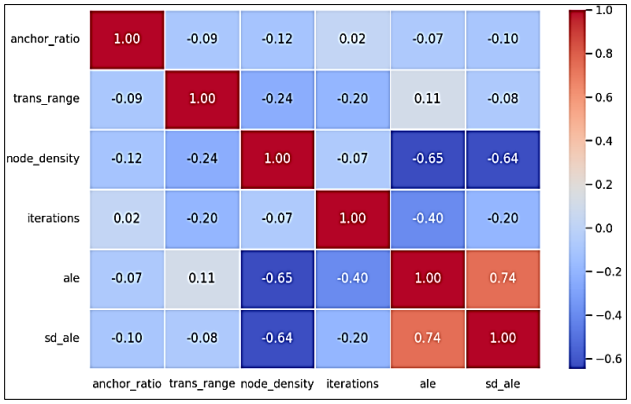


Figure 5: Correlation matrix of features

An observation made from the scatter plot in fig 6 of Transmission Range and ALE is that one cannot say the impact is proportional or uniform. ALE seems to be more erratic with more dependence on the number of iterations. It also shows that higher iterations reduce and make steadier the estimated ALE, indicating the benefits of iterative methods in improving the localization precision. This implies that there is an expected interaction between iterations and transmission range in guaranteeing the best results of ALE. The results in the plot in fig 7 show a non-monotonic relationship between Node Density and ALE. Thus, the results shown for ALE indicate that as Node Density increases, localization accuracy improves, thus supporting the hypothesis. This improvement is more significant with higher anchor ratios, further substantiating

the propositions made regarding the anchor nodes of helping stabilize and enhance the localization performance. In the same way, a denser node environment seems to offer triangulation points more than one, which helps in minimizing the error level.

Combined, these analyses further stress the complexity of the relationships between the dataset features and demonstrate that iterative processing, node density, and anchor ratios are crucial for reducing ALE. They are advantageous in establishing prognosis for pattern recognition and enhancement of systems. These statistical descriptors are essential for getting first impressions of the data and are vital when searching for patterns and abnormalities in later stages of data analysis.

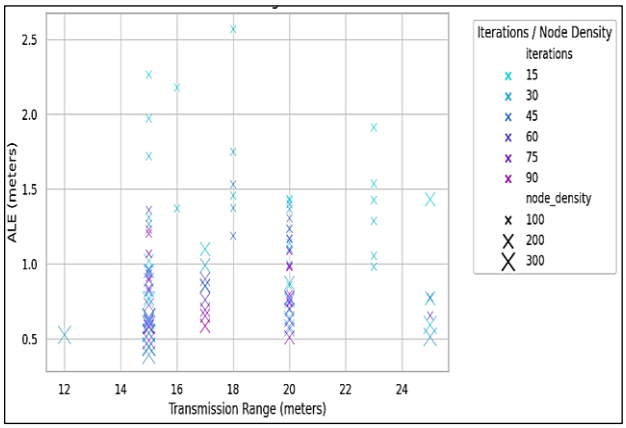


Figure 6: Transmission range Vs ALE

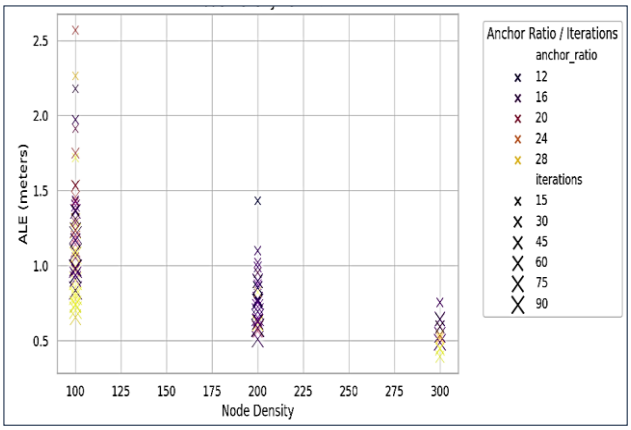


Figure 7: Node density Vs ALE

4 Feature ranking

The next process implemented after data acquisition is the process of selecting top-feature variables implemented by XGB-Regressor. This method analyzes the feature importance using contributions from all weak learners in an ensemble-based technique. The ranking shows some key parameters like Anchor Ratio, Iterations, Transmission Range, Node Density, and etc., which are important to understand which factors influence ALE to a greater extent. This presents how the importance score is computed based on the importance of feature f, computed as in eq 2. Prescribing these key features to the subsequent modeling phase, the adaptation procedure can be simplified as a process of fine-tuning key parameters. Table I displays the symbol description used in equation representation.

$$f = \frac{\sum_{i\epsilon T_f}(Residuals(i)^2 - (Residuals(i\prime)^2}{|T_f|} \quad (2)$$

Additionally, the importance of feature ranking analyzing with the help of SHAP plot gives importance as well as the impact of the same features on model's predictions for ALE in WSN, shown in fig 8. The SHAP values signify the extent to which each feature adds to the increase or decrease of the predicted ALE. The most influential feature, according to SHAP, is sd_ale in this sense, as it has the highest magnitude of SHAP values, indicating that deviations in standard deviation of ALE significantly influence the prediction of localization error. A strong impact is also made by the number of iterations, which is interpreted as the higher iterations become the more model stable and localizable. The influence from node density, transmission range is moderate, as an increase in these parameters tends to reduce ALE, since denser networks and wider communication range influence localization precision. It is found that the anchor ratio has the least effect, so even though anchor node deployment is very important, its effect on the prediction of ALE largely outweighs other parameters. This analysis demonstrates that prediction of ALE depends heavily on network topology and the measurement variability (sd_ale, iterations and node density) and suggests future optimization of WSN based localization models.

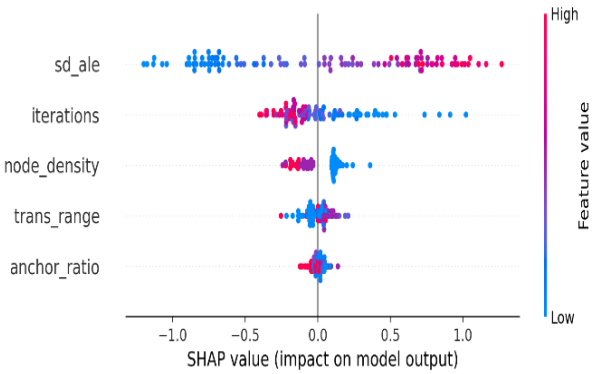


Figure 8: SHAP Analysis for feature ranking

5 Modelling and evaluation

We build regression model to make prediction of ALE on the basis of ranked features. Four models are employed: each has its advantages, and enables a holistic assessment of ALE prediction incorporating all the examined methods.

Like for SVR, the aim is to identify a function that has least mean squared error and generalize well from the data. SVR works in a way that maps the data into a higher dimension using the kernel function $f(x)$ and thereafter tries to fit a hyperplane such that it can leave the error within a given margin width. This method is especially appropriate when it is expected as observed target y_i that there are some interesting forms of non-linear

characteristics in the data, by maintaing a flat function during training based on values, defined as in 3.

$$f(x) = w^T x + b \quad (3)$$

The objective function of SVR aims at reducing the model complexity as much as possible so as to bound the prediction errors using 4. in a tolerance region encapsulated by ϵ, the epsilon-error, subjected to constraints as defined in eq 5.

$$y_i - f(x_i) \leq \varepsilon + \varepsilon_i \text{ where } \varepsilon_i \geq 0 \quad (4)$$

$$f(x_i) - y_i \leq \varepsilon + \varepsilon_i \quad (5)$$

This helps in to find the optimal weight and bias that reduces the sum of the model's complexity and the total error, controlled by the slack variables. This makes SVR less sensitive to outliers and it gives good generalization in a number of problems. *SVR* is preferred when the data display even higher order non linear relationships that cannot be described by linear, polynomial models.

KNN is the non-parametric predictive technique in which predicted value is derived from average of the target value of the nearest 'k' neighbors in the feature space. It is especially helpful in a system when the data has local sovereignty or some regional characteristics. *DT* model on the other hand constructs a tree structure where each node is a test on the basis of feature value and each node is an outcome. *AdaBoost* is a type of meta algorithm that is trained from weak classifiers which are usually shallow decision trees. Like all boosting algorithms, AdaBoost is used to train weak learners one by one, with each learner given the responsibility of fixing the errors made by the former learner. AdaBoost's strength is in offering greater concentration for misclassified points in each iteration making the model more accurate [39].

Other than developing the models, the regression models are benchmarked guided by performance indicators like R-squared and Root Mean Square Error (RMSE) [40]. R-squared is the measure of closeness of fit of the regression model, which tells how much of this depends on the model. It is calculated as in 6:

$$R^2 = 1 - \frac{\coprod_{i=1}^{n}(y_i - \breve{y}_i)^2}{\coprod_{i=1}^{n}(y_i - \breve{y})^2} \quad (6)$$

RMSE focuses on evaluating the actual amount of the prediction error through square root of mean square error between the actual value and estimated value, computed using eq 7. It provides an indication of the model's precision in predicting ALE values.

$$RMSE = \sqrt{1/n \coprod_{i=1}^{n}(y_i - \breve{y}_i)^2} \quad (7)$$

Another factor included is the computational time, especially for the application of the algorithms in real-time operations since it measures the rate of each algorithm. This is an important consideration especially when deploying models in environments of very high speed such as the wireless sensor networks which need fast localization estimation. Thus, by using this multiple criteria evaluation procedure, both accuracy and time efficiency criteria serve to define the choice of the most suitable model for the ALE prediction. Through integration of these models and competing them with a defined and exhaustive set of evaluation metric, we also ensure that we have developed a refined and most effective solution with respect to localization error prediction.

Table 2: Symbols description

Symbols	Description
$Residuals(i)^2$	Difference between predicted value and value for data point i.
T_f	Subset of data points where feature is used in the split
$(Residuals(i')^2$	Residuals after the feature is used in the model.
$\|T_f\|$	No. of data point that use feature for splitting
b	Bias term
x	Input vector
w	Weight vetor
ε	Margin of tolerance within which no penalty is incurred for errors
ε_i	Slack Variables
n	Number of data points
y_i	Actual Values
$\breve{y}_i$	Predicted Values
$\breve{y}$	Mean of Actual Values

5.1 Hyperparameter settings

Tuning parameter is very important when it comes to the determination of the best performing machine learning models and simulation experiments. Table II presents the analysis of hyperparameter settings. Through automated process control, the SVR model uses several well-defined hyperparameters to attain high predictive accuracy. These are the penalty factor (C), which regulate the tradeoff between low error on the training set data and model complexity, where C ranges from 0.01. Two parameters can be tuned: epsilon (ε) which determines the acceptance of the model prediction being off by a certain value set to 0.001. The kernel can be polynomial and as the claim indicates its degree is 2 so that the model can identify non-linear relations. The spread shape parameter, gamma (γ), is set to 1 and controls the extent to which individual data points impact the locations of boundary models.

Table 3: Hyperparameter settings

Parameter	Description	Values
Number of Neighbors (k)	Number of neighbors considered in algorithms like KNN.	5 to 50
Weight Function	Method to assign weights to neighbors.	Uniform
Maximum Depth	Maximum depth of the decision tree.	5 to 50
Minimum Samples Split	Minimum samples required to split an internal node.	2 to 10
Minimum Samples Leaf	Minimum samples required to form a leaf node.	1 to 5
Split Criterion	Criterion for splitting nodes in decision trees.	MSE, R-Sq
Number of Estimators	Number of estimators in ensemble methods.	50 to 200
Learning Rate	Step size for weight updates in gradient boosting.	0.01, 0.001
Penalty factor (C)	Controls trade-off between complexity and fit.	0.01
Epsilon (ε)	Tolerance for prediction accuracy.	0.001
Kernel	Type of kernel used in SVM or similar methods.	Polynomial
Polynomial degree (d)	Degree of the polynomial kernel.	2

Gamma (γ)	Influence of data points in kernel functions.	1
Deployment Area (m²)	Sensor deployment area.	100 x 100
Node Density	Total number of nodes in the test area.	300
Anchor Ratio (%)	Percentage of anchor nodes.	50
Communication Range (m)	Maximum distance for node communication.	50
Step size (α)	Incremental adjustment parameter.	0.9
Mutation Probability (Pα)	Likelihood of mutation in candidate solutions.	0.25
Candidate Solutions	Number of candidate solutions.	25
Iterations	Number of iterations.	50
# boosting Estimators	Number of boosting rounds.	50

For the WSN parameters, position area of 100 meters by 100 meters, node density of 300 and anchor ratio of 50% are important to model the network environment. Communication range (50m) plays the role of influencing the localization accuracy. Additional hyperparameters, such as step size (α: 0.9). Mutation probability (Pa: 0.25), number of candidate solutions 25, and number of iterations 50. In the XGB Regressor for feature ranking, some of the convenient hyperparameters include the learning rate, the number of boosting estimators of 50, and providing an optimum balance between over fitting and model performance. The precise tuning of these hyperparameters allows the reader reliable feature rankings provided in the final XGB Regressor model.

6 Results

6.1 Feature ranking and importance

The bar chart in fig 9 portrays the predictor importance of four parameters – Anchor Ratio, Transmission Range, Node Density and Iteration calibrated using XGBRegressor algorithm. This regression approach is comprised with weak learners known as the regression trees with a single learning rate of unity. The importance of each predictor is derived from all the built weak learners combined into an ensemble. From the visualization, Node Density stands out as the most important feature indicating the highest importance estimate. This goes a long way in suggesting that the characteristic of the network which has a most pronounced impact on the performance of the model in terms of the average localization error (ALE) is the density of nodes in the network. After Node Density, there is Iteration in which it can be seen that an increase in iteration provides a comparatively large improvement to the model prediction.

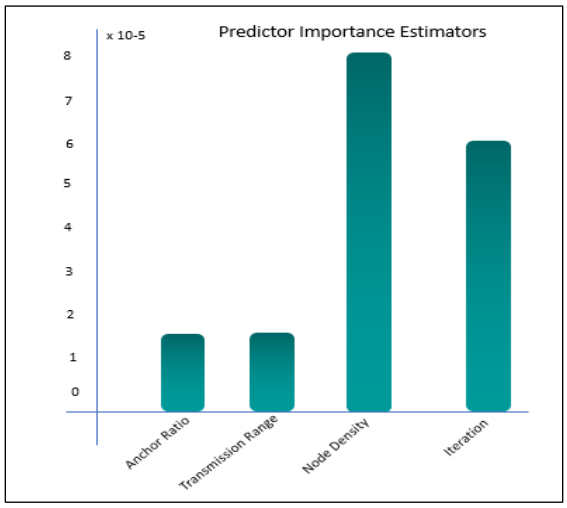


Figure 9: Importance of features

On the other hand, less importance is presented by Anchor Ratio and Transmission Range, with similar estimates, staring a far from negligible impact on the overall performance of the model. Hence, these results are useful in determining which of the network parameters needs to be tuned in, during the formation of the network in order to achieve the minimum effective ALE. The concern shown to Node Density and Iteration may relate to strategies for network deployment and computational refinements toward attaining better node localization in WSNs.

The partial dependence and ICE of four predictors—Anchor Ratio, Transmission Range, Node Density, and Iteration—on ALE are illustrated through the graphs displayed above. The following fig 10 illustrate the impact of variation of individual features on the predicted ALE by the model.

Anchor Ratio (a): The ALE experience a slow reduction with an increase in the Anchor Ratio but with a certain volatility. The yellow dotted line, the average curve, rises steadily as Anchor Ratios grow, implying that the higher proportion of anchor nodes helps to achieve better ALE.

Transmission Range (b): As for Transmission Range, the trend is also similar to Previous Year. That is, when the range is increased, there will be a reduction in ALE although the relationship is not as straight forward as that of the Anchor Ratio. The dispersed points show that individual data vary while the dependence proves the effectiveness of increasing the transmission range on ALE decrease.

Node Density (c): Node Density indicates a radical reduction in ALE with density which, by the yellow trend line, points to better localization at higher densities. This suggests that, a denser node deployment makes a significant contribution to decreasing the ALE, as was also concluded during statistical feature importance analysis.

Iteration (d): Similar to the case with accuracy, the number of iterations also reflects a high level of influence on ALE, although the reduction is highest in the beginning. Consequently, dependency over iterations grows and flattens out as iterations are increased; it indicates diminishing returns as iterations increases after a certain limit. This behavior indicates the fact that a number of iterative computations improves the precision as far as possible, while performing more and more iterations does not make much improvement.

In general, the partial dependence and this plotings help in understanding the features' behavior and make it clear that higher Anchor Ratios, Transmission Ranges, Node Densities, and moderate Iteration counts are good for minimizing ALE. These results are in concordance with feature importance and can be used to provide optimum solutions in Wireless Sensor Network configurations.

6.2 Computational results

The regression data set for mean prediction estimate of Average Localization Error (ALE) with various models reflects significant differences in their efficiency. The model with the highest degree of accurate results is Support Vector Regression (SVR) acquired 0.99 with R score thus it is known that the model is efficient in the prediction of ALE out of the total variance. Also, its RMSE is 0.01, which is an evidence of high predictive accuracy since the residual errors are not markedly high. This evaluation reveals that SVR has the capacity to model intricate, nonlinear correlatives that are natural in the data set, which makes it the most accurate model of predicting ALE, also results are shown in table III. Analysis reveals that SVR achieves highest accuracy (R2 = 0.99) with user friendly computational cost (0.01s) which makes it implementable in real life. However, latency in AdaBoost (0.06s) and Decision Tree (0.05s), which may limit their applicability in time-sensitive WSN applications.

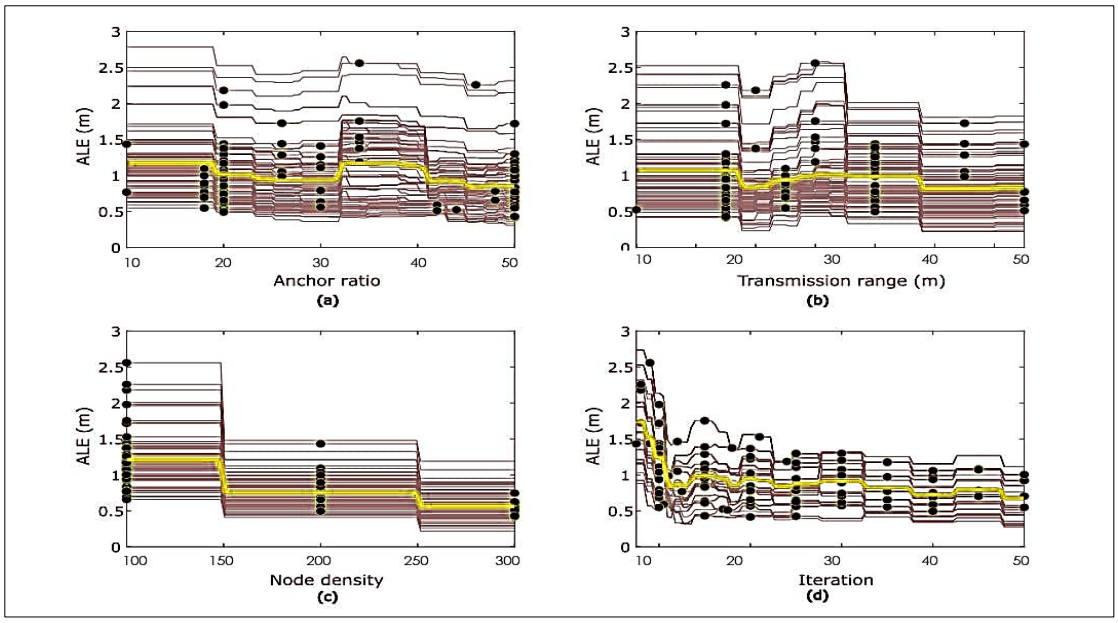


Figure 10: Impact of variation of individual features

On the other hand, the AdaBoost Regressor indicated moderate accuracy with an R-squared of 0.720 generated by using the test dataset to predict the actual values. While its RMSE of 0.16 is nearly the same as SVR, the dataset complexity was not captured as well by AdaBoost since the R-squared is lower. This could be due to two reasons; it could be sensitive to noise or the hyperparameters have not been well tuned. However, such an architecture still has its merit in the consideration that it remains an ensemble design if balanced performance is acceptable. The KNeighbors Regressor performs slightly better in terms of error with the minimum RMSE of 0.14, but its coefficient of determination or R-squared value of 0.55 means it only captures only 55% of variation in ALE. Although its RMSE is lower, showing small bounds around some samples, the R-squared is lower, meaning efficacy for predicting the current limited samples but not well on the overall distribution within the dataset because the idea of local neighbors may not reflect the entire sample.

Table 4: Regression-based results

	SVR	DT	KNN	AdaBoost	LR
R-Squared	**0.99**	0.41	0.55	0.72	0.3
RMSE	**0.01**	0.16	0.14	0.16	0.18
Time Taken	**0.01**	0.05	0.01	0.06	0.02

The aforementioned model has the least performance; DT Regressor has R-squared score of 0.41 and RMSE of 0.16. There is evidence of overfit or splits inadequate to extract this variation in data, from which the model doesn't learn well complex non linear relationships to predict localization error. This is very typical for decision tree used independently because, unlike logistic regression for

example, decision trees do not perform very well in capturing non-linear relationships and the interactions between the variables without such tricks as regularization or averaging stemming from ensemble usage. To sum it up, SVR was identified as the most accurate model for forecasting ALE while at the same time keeping residual error to the minimum. However, compared to other models, AdaBoost and KNeighbors try to provide different trade-off between errors and variance in explanation, and though the Decision Tree independently gives a good point for standalone performance, it does actually implies that such models require amendment like ensemble techniques in order to effectively meet the objectives for this context. All evaluated models had the lowest predictive performance except LR model that had an R-squared of 0.30 and the highest RMSE of 0.18. It implies that the localization error prediction problem within WSNs is a hard one for LR to grasp. Although LR is frequently useful for classificaton tasks, its shortcomings in regression based ones are based on its inability to model non linear dependencies that are essential for the correct prediction of ALE. Moreover, the computational time of 0.02s is a little higher than SVR and KNN and lower than Decision Tree and AdaBoost. These results generally reaffirm that more advanced regression models like SVR are more suitable for the ALE prediction in WSNs since they can handle non linear patterns and complex feature interaction. The performance of SVR, Decision Tree, KNN, LR and AdaBoost have been compared using the parameters of computation time, R-square, and RMSE, as shown in fig 11.

6.3 Discussion

Results obtained show that SVR outperformed other regression models used in this work for the prediction of ALE in WSN. This performance also outperforms the multitude of ML based localization methods which previous studies have reported R of 0.82 and 0.89, with RMSE of 0.147m and 0.02m, respectively in table V. This is due to the superior accuracy that SVR brings to localization error data modeling compared to Decision Tree and KNN as it more easily generalizes to complex nonlinear relationships. Furthermore, using the kernel in SVR helps it to capture the variance of the ALE, therefore improving its predictive accuracy. Specifically, comparing with other machine learning based localization techniques (such as ensemble learning methods (e.g. AdaBoost) are promising in handling the localization error, but they typically need larger datasets and more computational resources. The results show that SVR achieves very good trade-off between accuracy and computational efficiency, thus being a reasonable choice for real time location in WSNs. The variations in performance of different models are due to the differences in the feature sensitivity, the hyperparameter optimization and dataset characteristics, and models like Decision Tree and KNN are more sensitive to the noisy data.

The experimental results show that ALE in WSN can be well reduced under some certain parameters, including node density and anchor ratio. The density of the nodes is increased, and it is shown that there is more connectivity and lower localization uncertainty, so the distance estimation is more accurate, and positioning error is minimized. In the same way, the more the anchor nodes the weaker the reference points will be, and the better the location estimates will be. But besides anchor ratio, other factors are also important, though with a lower influence, possibly because secondary return starts to diminish when too much other return is obtained. However, from the practical point of view, these feature importance trends imply that WSN deployments must put more focus on uneven growth of node density in critical regions to improve localization accuracy while distribution of anchor nodes to maximize efficiency without excessive resource consumption. This will help network engineers optimally lay out a WSN to satisfy localization accuracy improvements with minimal extra bottleneck deployment cost and energy consumption. Although the proposed model achieves high accuracy in predicting, some limitations of the proposed model are to be acknowledged. Second, the results obtained from the model are still generalizable across other WSN configurations yet to be tested. The performance may vary with variations in network topology, environmental conditions, and hardware specifications and therefore requires additional validation on different aspects. In addition, the study is also based on the stationary topology of the network, which implies that locations of the nodes are fixed during localization. The performance of such systems tends to degrade in dynamic WSN environments where nodes are mobile, and signal conditions are rapidly changing as well as in case of unpredictable connectivity fluctuations. Finally, while the data quality was ensured through preprocessing steps, the model should be further assessed regarding its reproducibility with different datasets or for real world deployments.

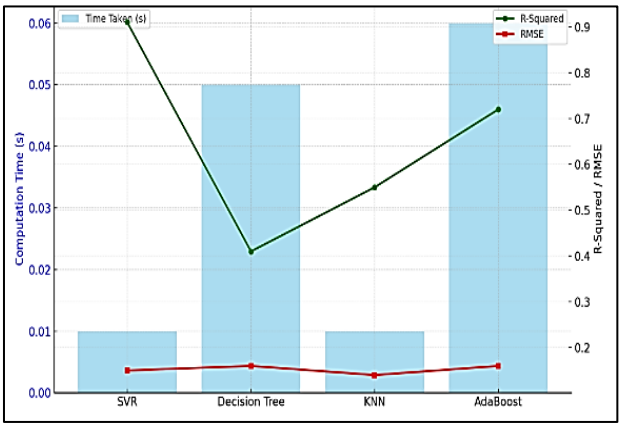


Figure 11: Computational time of models

SVR stands out with the highest accuracy (R-squared: 0.99 as well as low RMSE, 0.01 and is one of the quickest models which takes just 0.01 seconds to execute. The lowest RMSE is observed for KNN model (RMSE = 0.14), and the values of R-squared are moderate (R-squared = 0.55) for all the models with relatively low computational burden as for SVR. In contrast, AdaBoost offers decent accuracy (R-squared: 0.72) but the slowest of all with a time of 0.06 seconds. The Decision Tree model has the lowest accuracy (R-squared: 0.41) and a moderate processing time of 0.05 seconds, making it less suitable for high-precision localization error prediction in WSNs. Comparing all the four algorithms, SVR delivers the best

result with the most efficient running time. In response to the problem of evaluating the ALE in WSNs using regression learning algorithms, we tested several models. These results support the fact that while using complex structures like suggested in Section 5 for ALE predictions other features of WSNs must be considered while selecting regression algorithms like SVR for accurate predictions due to interactions and nonlinearity between features.

Our proposed SVR model performs exceptionally well (R^2 = 0.99, RMSE = 0.01m), it does not significantly outperform the best prior benchmark (R = 0.89, RMSE = 0.04m), but still demonstrates strong predictive accuracy in localization error analysis. Existing works have obtained reasonable results in prediction for traffic: R=0.82 with RMSE = 0.147m and R=0.89 with RMSE= 0.04m; thus, the proposed model outperforms the benchmarks by obtaining R=0.99 and RMSE = 0.01m.results, as shown in table IV. This improvement demonstrates the ability of the proposed technique in modeling the correlation between input parameters like Anchor Ratio, Transmission Range and Node Density and ALE. The lower RMSE validated our model with higher accuracy as compared to other models and high R-squared depicts its goodness of fit as compared to prior studies have achieved moderate results. These results peak at the effectiveness of the proposed model in comparing localization errors as opposed to traditional approaches to the problem, providing richer prediction strength.

Table 5: Comparison with existing literature

Sr. No	Ref	Year	Model	Results (R)
1	[1]	2020	Range-SVM	0.82
2	[43]	2021	MLP	0.92
3	[42]	2022	SVR	0.95
4	[41]	2023	GB	0.13
5	[2]	2024	RF	0.89
6	**Proposed**	**2024**	**SVR**	**0.99**

7 Conclusion

The Localization in WSNs is important in versus of data acquisition and interpretation in various applications which include environment monitoring, disaster prediction and management, and military security surveillance. This research aims to forecast ALE while employing feature ranking with the XGB ensemble-based regressor and creating a regression-based model. In this study, we have identified that SVR Regressor is the most precise model achieving high generalization performance, and we also introduced computation time to evaluate efficiency. The descriptive analysis was useful for understanding the data and provided insights into the importance of anchor ratio, transmission range, and node density features that were discovered by utilizing the XGB model in the outcome. The findings highlight the importance of feature selection, model optimization, and computational efficiency in ensuring accurate and real-time localization error predictions. Through comprehensive preprocessing, hyperparameter tuning, and model evaluation, this study establishes SVR as a reliable approach for ALE prediction, balancing accuracy and efficiency in resource-constrained WSN environments. Further research developments are to be made based on the enhancement of deep learning models that will increase the accuracy of the data collected through the WSNs, applying real-time optimization of the WSN applications, as well as on the expansion of the proposed study into larger, more active WSNs. This paper provides useful suggestions for further research into enhancing localization approaches and promoting improved efficient, high-quality solutions for WSNs.

References

[1] Singh, A., Kotiyal, V., Sharma, S., Nagar, J., & Lee, C. C. (2020). A machine learning approach to predict the average localization error with applications to wireless sensor networks. *IEEE Access*, *8*, 208253-208263.

[2] Watanabe, F. (2021). Wireless sensor network localization using AoA measurements with two-step error variance-weighted least squares. *IEEE Access*, *9*, 10820-10828.

[3] Zhu, Z., & Zhu, M. (2024). Utilizing Machine Learning Approach to Forecast Average Location Determination Errors in Wireless Sensor Networks. *Journal of Artificial Intelligence and System Modelling*, *1*(02), 83-98.

[4] Lalama, Z., Boulfekhar, S., & Semechedine, F. (2022). Localization optimization in WSNs using meta-heuristics optimization algorithms: a survey. *Wireless Personal Communications*, *122*(2), 1197-1220.

[5] He, J., Fakhreddine, A., Vanwynsberghe, C., Wymeersch, H., & Alexandropoulos, G. C. (2023). 3D localization with a single partially-connected receiving RIS: Positioning error analysis and algorithmic design. *IEEE Transactions on Vehicular Technology*, *72*(10), 13190-13202.

[6] Zou, Y., & Liu, H. (2021). RSS-based target localization with unknown model parameters and sensor position errors. *IEEE Transactions on Vehicular Technology*, *70*(7), 6969-6982.

[7] Jia, Z., & Guan, B. (2018). Received signal strength difference–based tracking estimation method for arbitrarily moving target in wireless sensor networks. *International Journal of Distributed Sensor Networks*, *14*(3), 1550147718764875.

[8] Kagi, S., & Mathapati, B. S. (2022). Localization in wireless sensor network using machine learning optimal trained deep neural network by parametric analysis. *Measurement: Sensors*, *24*, 100427.

[9] Han, G., Yang, X., Liu, L., Zhang, W., & Guizani, M. (2017). A disaster management-oriented path planning for mobile anchor node-based localization in wireless sensor networks. *IEEE Transactions on Emerging Topics in Computing*, *8*(1), 115-125.

[10] Sorbelli, F. B., Pinotti, C. M., Silvestri, S., & Das, S. K. (2020). Measurement errors in range-based localization algorithms for UAVs: Analysis and experimentation. *IEEE Transactions on Mobile Computing*, *21*(4), 1291-1304.

[11] Singh, H., Yadav, P., Rishiwal, V., Yadav, M., Tanwar, S., & Singh, O. (2024). Localization in WSN-Assisted IoT Networks Using Machine Learning Techniques for Smart Agriculture. *International Journal of Communication Systems*, e6004.

[12] Sneha, V., & Nagarajan, M. (2020). Localization in wireless sensor networks: a review. *Cybernetics and Information Technologies*, *20*(4), 3-26.

[13] Tian, Y., Huang, B., Jia, B., & Zhao, L. (2020). Optimizing AP and Beacon Placement in WiFi and BLE hybrid localization. *Journal of Network and Computer Applications*, *164*, 102673.

[14] Abhale, A. B., & Manivannan, S. S. (2020). Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. *Optical Memory and Neural Networks*, *29*(3), 244-256.

[15] Shahraki, A., Taherkordi, A., Haugen, Ø., & Eliassen, F. (2020). Clustering objectives in wireless sensor networks: A survey and research direction analysis. *Computer Networks*, *180*, 107376.

[16] Yang, S., Yuan, Z., & Li, W. (2020). Error data analytics on RSS range-based localization. *Big Data Mining and Analytics*, *3*(3), 155-170.

[17] Janczak, D., Walendziuk, W., Sadowski, M., Zankiewicz, A., Konopko, K., & Idzkowski, A. (2022). Accuracy analysis of the indoor location system based on Bluetooth low-energy RSSI measurements. *Energies*, *15*(23), 8832.

[18] Zheng, Y., Liu, J., Sheng, M., Han, S., Shi, Y., & Valaee, S. (2020). Toward practical access point deployment for angle-of-arrival based localization. *IEEE Transactions on Communications*, *69*(3), 2002-2014.

[19] Lakshmi, Y. V., Singh, P., Mahajan, S., Nayyar, A., & Abouhawwash, M. (2024). Accurate range-free localization with hybrid DV-hop algorithms based on PSO for UWB wireless sensor networks. *Arabian Journal for Science and Engineering*, *49*(3), 4157-4178.

[20] Kumar, S., Kumar, S., & Batra, N. (2021). Optimized distance range free localization algorithm for WSN. *Wireless Personal Communications*, *117*, 1879-1907.

[21] Wu, Y., Zhang, C., Tong, L., & Shi, X. (2023). Location Optimization based on Improved 3D DV-HOP Algorithm in Wireless Sensor Networks. *IEEE Access*.

[22] Arya, R. (2021). C-TOL: Convex triangulation for optimal node localization with weighted uncertainties. *Physical Communication*, *46*, 101300.

[23] Kuriakose, J., Joshi, S., & Bairwa, A. K. (2023). Embn-manet: A method to eliminating malicious beacon nodes in ultra-wideband (uwb) based mobile ad-hoc network. *Ad Hoc Networks*, *140*, 103063.

[24] Yang, H., Wang, Y., Seow, C. K., Sun, M., Si, M., & Huang, L. (2023). UWB sensor-based indoor LOS/NLOS localization with support vector machine learning. *IEEE Sensors Journal*, *23*(3), 2988-3004.

[25] Olejniczak, A., Blaszkiewicz, O., Cwalina, K. K., Rajchowski, P., & Sadowski, J. (2024). LOS and NLOS identification in real indoor environment using deep learning approach. *Digital Communications and Networks*, *10*(5), 1305-1312.

[26] Akram, J., Javed, A., Khan, S., Akram, A., Munawar, H. S., & Ahmad, W. (2021, March). Swarm intelligence based localization in wireless sensor networks. In *Proceedings of the 36th annual ACM symposium on applied computing* (pp. 1906-1914).

[27] Li, Y., Zhuang, Y., Hu, X., Gao, Z., Hu, J., Chen, L., ... & El-Sheimy, N. (2020). Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation. *IEEE Internet of Things Journal*, *8*(6), 4035-4062.

[28] Yadav, P., & Sharma, S. C. (2023). A systematic review of localization in WSN: Machine learning and optimization-based approaches. *International journal of communication systems*, *36*(4), e5397.

[29] Mohammed, S. K., Singh, S., Mizouni, R., & Otrok, H. (2023). A deep learning framework for target localization in error-prone environment. *Internet of Things*, *22*, 100713.

[30] Jondhale, S. R., Shubair, R., Labade, R. P., Lloret, J., & Gunjal, P. R. (2020). Application of supervised learning approach for target localization in wireless sensor network. *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's*, 493-519.

[31] Lemic, F., Handziski, V., & Famaey, J. (2019, April). Toward regression-based estimation of localization errors in fingerprinting-based localization. In *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)* (pp. 1-5). IEEE.

[32] Giannopoulos, A., Spantideas, S., Nomikos, N., Kalafatelis, A., & Trakadas, P. (2023, April). Learning to fulfill the user demands in 5G-enabled wireless networks through power allocation: A reinforcement learning approach. In *2023 19th International Conference on the Design of Reliable Communication Networks (DRCN)* (pp. 1-7). IEEE.

[33] Hajiakhondi-Meybodi, Z., Mohammadi, A., Hou, M., & Plataniotis, K. N. (2022). DQLEL: Deep Q-learning for energy-optimized LoS/NLoS UWB node selection. *IEEE Transactions on Signal Processing*, *70*, 2532-2547.

[34] Altay, O., Erel-Özçevik, M., Varol Altay, E., & Özçevik, Y. (2024). Average Localization Error Prediction for 5G Networks: An Investigation of Different Machine Learning Algorithms. *Wireless Personal Communications*, 1-31.

[35] Poulose, A., & Han, D. S. (2020). UWB indoor localization using deep learning LSTM networks. *Applied Sciences*, *10*(18), 6290.

[36] Zhao, L., Huang, H., Li, X., Ding, S., Zhao, H., & Han, Z. (2019). An accurate and robust approach of device-free localization with convolutional autoencoder. *IEEE Internet of Things Journal*, *6*(3), 5825-5840.

[37] Park, J., Moon, J., Kim, T., Wu, P., Imbiriba, T., Closas, P., & Kim, S. (2022). Federated learning for indoor localization via model reliability with dropout. *IEEE Communications Letters*, *26*(7), 1553-1557.

[38] Ghous, M., Nguyen, T. L., Do, T. N., & Kaddoum, G. (2024). Deep Transfer Learning-based Performance Prediction of URLLC in independent and not necessarily identically distributed Interference Networks. *IEEE Access*.

[39] Srinivasan, S. M., Truong-Huu, T., & Gurusamy, M. (2019). Machine learning-based link fault identification and localization in complex networks. *IEEE Internet of Things Journal*, *6*(4), 6556-6566.

[40] Masood, A., Gulzar Ahmad, S., Ullah Khan, H., & Ullah Munir, E. (2020). Network reconfiguration algorithm (NRA) for scheduling communication-intensive graphs in heterogeneous computing environment. *Cluster Computing*, *23*(2), 1419-1438.

[41] Rahman, M. M., & Nisher, S. A. (2023, January). Predicting average localization error of underwater wireless sensors via decision tree regression and gradient boosted regression. In *Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022* (pp. 29-41). Singapore: Springer Nature Singapore.

[42] Jondhale, S. R., Mohan, V., Sharma, B. B., Lloret, J., & Athawale, S. V. (2022). Support vector regression for mobile target localization in indoor environments. *Sensors*, *22*(1), 358.

[43] Nosrati, L., Fazel, M. S., & Ghavami, M. (2022). Improving indoor localization using mobile UWB sensor and deep neural networks. *IEEE Access*, *10*, 20420-20431.

Enhanced Entropy-Fuzzy Integration Decision Support System for Risk Assessment and Management of Hydraulic Engineering

Ying Li*, Yanling Wang
College of Hydraulic Engineering, Chongqing Water Resources and Electric Engineering College, Chongqing 402160, China
E-mail: liying126@hotmail.com, WangYanling_yl@outlook.com
*Corresponding author

Keywords: hydraulic engineering, risk assessment, enhanced entropy, fuzzy decision support system

Received: October 30, 2024

Against the backdrop of global climate change, frequent extreme weather events, and the increasingly complex development and management of water resources, the enhanced entropy - fuzzy integration decision support system for the risk assessment and management of hydraulic engineering has emerged. Traditional risk management methods, such as risk assessment based on probability theory, can handle known risks to a certain extent but are insufficient when dealing with highly ambiguous, uncertain, and complex risks. This paper introduces the design and empirical evaluation of the Enhanced Entropy - Fuzzy Integration Decision Support System (EEMFDS). Taking the Yangtze River Three Gorges Project as the research object, the application value of EEMFDS in the risk assessment and management of water conservancy projects is fully demonstrated through detailed requirement analysis, system architecture design, model construction, algorithm design, and empirical evaluation. EEMFDS adopts fuzzy set theory and enhanced entropy model, and through methods such as fuzzy quantification and enhanced entropy calculation, it effectively addresses key links such as risk identification, quantification, assessment, and decision support. The empirical results show that compared with traditional risk management methods, EEMFDS has significantly improved the risk identification accuracy. The average accuracy rate has increased from 84.7% to 90.6%, the response time has been shortened from 180 seconds to 120 seconds, the comprehensive robustness score has increased from 79.2% to 87.6%, it also has advantages in resource consumption, with the CPU usage time reduced by 30 hours and the storage space reduced by 20GB. The user satisfaction is relatively high, with an overall average score of 85.8%, highlighting its great potential in enhancing the level of intelligent risk management.

Povzetek: Razvit je izboljšan entropijski sistem mehke logike za oceno tveganj v hidravličnem inženirstvu, ki učinkovito obravnava negotovost in kompleksnost.

1 Introduction

As an important part of national infrastructure, water conservancy projects play an irreplaceable role in promoting agricultural development, ensuring urban water supply, flood control and disaster reduction, and promoting balanced regional economic development. With the rapid development of global climate change and economic society, water conservancy projects are under increasing pressure in the face of extreme weather events, water shortage, ecological environment protection, etc., and the risks in their construction and operation management are more complex and diverse. Natural disasters, such as floods and earthquakes, are the primary external risks faced by hydraulic projects, which may lead to structural damage, functional failure and even secondary disasters [1]. This paper aims to solve the problems of subjectivity and limited information processing ability existing in traditional risk assessment methods of hydraulic engineering, and improve the objectivity, accuracy and timeliness of risk assessment by introducing enhanced entropy fuzzy decision support system [2]. Traditional risk assessment methods often rely on expert experience, which makes it difficult to capture all risk factors comprehensively and has limitations in dealing with vague and uncertain information. Enhanced entropy fuzzy decision support system combines the uncertainty processing ability of fuzzy set theory with the information optimization processing ability of enhanced entropy theory, which can deal with complex and fuzzy information in risk assessment of hydraulic engineering more effectively and help decision makers make reasonable decisions under multi-objective and multi-constraint conditions. The development and application of this system can not only improve the intelligent level of risk management of hydraulic engineering, but also provide reference for risk assessment and management in other fields [3].

In recent years, scholars at home and abroad have made

extensive exploration in the field of hydraulic engineering risk assessment, and put forward many assessment models and methods, such as probabilistic risk assessment, fuzzy comprehensive assessment, analytic hierarchy process, etc. [4]. Each of these approaches has its own focus, but a common problem is the limited ability to deal with highly uncertain information, especially when there are many risk factors and complex interrelationships. As an effective decision-making tool, fuzzy decision support system has been applied to risk management in many industries, but its application in hydraulic engineering is still in its infancy, especially how to integrate a large amount of uncertain information efficiently to support more accurate risk assessment is still a key problem to be solved [5].

Entropy theory, as a method of optimizing information processing, measures uncertainty of information by introducing entropy, optimizes information structure by enhancement, and improves efficiency and quality of decision making. Although the theory has been applied in information science, economics and other fields, it is still a frontier research direction, especially in the integrated application of fuzzy decision support system in hydraulic engineering risk assessment [6].

The core content of this study centers on constructing an enhanced entropy fuzzy decision support system suitable for risk assessment and management of hydraulic engineering, including: (1) System requirement analysis: defining the specific requirements of risk assessment of hydraulic engineering, including risk identification, risk quantification, risk ranking and response strategy generation. (2) Model and algorithm design: Based on fuzzy set theory and enhanced entropy theory, risk assessment model is designed, corresponding fuzzy inference algorithm and enhanced entropy calculation method are developed to optimize risk information processing flow. (3) System architecture and implementation: Build the overall architecture of the system, including data input, risk assessment engine, decision support module and user interface, etc., and adopt advanced software development technology and database management system to realize system functions.

This study aims to develop and evaluate the Enhanced Entropy Fuzzy Decision Support System (EEMFDS) for risk assessment and management in hydraulic engineering. The key research questions are as follows: How does the enhanced entropy - fuzzy integration approach effectively handle the complex and uncertain risk factors in hydraulic engineering compared with traditional risk assessment methods? What are the optimal model structures and algorithms for accurately quantifying and assessing risks during the construction of EEMFDS, and how to ensure the reliability and stability of the model? To what extent can EEMFDS improve the accuracy, timeliness, and comprehensiveness of risk assessment in practical applications of hydraulic engineering projects such as the Yangtze River Three Gorges Project? What are the potential challenges and limitations when applying EEMFDS to large - scale hydraulic engineering projects, and how can these issues be addressed? These questions will guide the exploration and analysis of the entire research process, aiming to promote the application and development of EEMFDS in the field of hydraulic engineering risk management.

The innovations of this study are: first, the enhanced entropy theory is integrated with the fuzzy decision support system to effectively process complex risk information; second, the entropy enhancement algorithm introduces reference probability distribution and adjustment parameters to optimize information processing and improve assessment accuracy and stability; third, the microservice architecture and containerization technology are used to improve system scalability, maintainability and deployment flexibility. These innovations provide new ideas and methods for water conservancy project risk assessment.

2 Basic theory and method

This section mainly explains the basic theories and methods involved in the research, including risk assessment principles, fuzzy set theory, enhanced entropy theory, and system integration and data processing technology, laying a theoretical foundation for the subsequent introduction of system design and implementation.

2.1 Principles of risk assessment for hydraulic engineering

Risk assessment of hydraulic engineering refers to the process of identifying, analyzing and evaluating various risks that may be encountered in the process of construction and operation of hydraulic engineering, so as to formulate corresponding prevention and mitigation measures to ensure the safety and benefit of the project. Risk can be defined as the product of the likelihood of a particular event occurring and the severity of its consequences. In the field of hydraulic engineering, risks are usually classified into natural risks (such as floods, earthquakes), technical risks (such as design defects, construction quality problems), management risks (such as operational errors, improper maintenance) and socio-economic risks (such as shortage of funds, policy changes) [7,8].

The risk assessment process generally follows four steps: identification, analysis, assessment and control. Firstly, potential risk factors are identified through historical data analysis and expert consultation.

2.2 Fuzzy set theory and fuzzy decision support system

Fuzzy set theory was proposed by Professor Zadeh to deal with fuzzy or imprecise information. Fuzzy sets allow elements to belong to multiple sets simultaneously with certain membership degrees, which solves the limitation of traditional binary logic. Fuzzy logic sets up reasoning mechanism to deal with fuzzy information, and deals with uncertainty problems through fuzzification, fuzzy reasoning and defuzzification. Fuzzy Decision Support System (Fuzzy DSS) is a form of DSS which applies fuzzy set theory. It can deal with fuzziness and subjective

judgment in decision making process and provide more realistic suggestions for decision makers [9].

2.3 Enhanced entropy theory

Information entropy was proposed by Shannon to quantify the uncertainty of information, while decision entropy further considers the value of information in the decision process. Enhanced entropy theory is developed on the basis of traditional entropy, which optimizes the information processing process by adjusting or "enhancing" the original information entropy [10]. Enhanced entropy model improves the discrimination of useful information by introducing additional information or criteria, which is helpful to screen out key information in complex multi-attribute decision making problems. The advantage of this theory is that it can effectively reduce noise interference in information processing and improve the stability and accuracy of decision-making. In information processing applications, enhanced entropy is often used in feature selection, pattern recognition, data clustering and other fields to improve the prediction ability and interpretation ability of models by optimizing information structure.

2.4 System integration and data processing technology

In modern hydraulic engineering risk assessment, data acquisition and preprocessing are basic links, involving various sensors, satellite remote sensing, unmanned aerial vehicle monitoring and other technical means. Data preprocessing includes data cleaning, standardization, missing value processing, etc. to ensure data quality. The application of big data and cloud computing technology has greatly enhanced data processing and analysis capabilities. Big data technology discovers hidden rules and associations in risk assessment through the collection, storage, processing and analysis of massive data, and supports real-time risk monitoring and early warning [11, 12]. Cloud computing provides flexible computing resources and storage space, enabling large-scale data processing tasks to be efficiently executed in the cloud, reducing the cost and maintenance pressure of local hardware facilities. For example, Document discusses the application of water conservancy information system based on cloud computing in flood warning, which significantly improves the timeliness and accuracy of early warning [13]. Literature introduces the method of fusion of fuzzy set theory and big data technology in safety assessment of hydraulic engineering, and shows how to use fuzzy logic to process uncertain information and improve risk identification ability in combination with big data analysis [14]. Reference discusses the application of enhanced entropy theory in risk decision of multi-attribute hydraulic engineering, emphasizing its role in optimizing information processing and reducing decision error [15].

In the research on risk assessment methods for hydraulic engineering, past scholars have proposed various methods, such as Probabilistic Risk Assessment (PRA) and Analytic Hierarchy Process (AHP). These methods have played important roles in the field of risk assessment, but they have certain limitations in handling uncertainty. The following is a comparative analysis of these methods (see Table 1).

The current State - of - the - Art (SOTA) has obvious deficiencies in handling highly uncertain information. For example, when faced with fuzzy information, incomplete data, and the interaction of complex risk factors, SOTA often fails to accurately capture and handle these uncertainties, affecting the accuracy and reliability of risk assessment. Moreover, many existing methods lack sufficient flexibility and adaptability when dealing with large - scale and dynamically changing risk scenarios in hydraulic engineering [16].

The Enhanced Entropy Fuzzy Decision Support System (EEMFDS) effectively addresses these deficiencies. EEMFDS combines fuzzy set theory and the enhanced entropy model. Fuzzy set theory can handle fuzzy and imprecise information, transform qualitative risk factors into quantifiable indicators, and make up for the defects of traditional methods in handling fuzzy information. The enhanced entropy model, by optimizing the information structure, enhances the expression of important information and reduces the interference of irrelevant information, enabling more accurate screening of key information in complex multi - attribute decision - making problems and improving the accuracy and stability of risk assessment. In addition, EEMFDS realizes automatic risk identification and classification through intelligent algorithms, can quickly adapt to dynamically changing risk scenarios, overcomes the problem of insufficient flexibility of existing methods, and provides a more efficient and accurate solution for risk assessment and management of hydraulic engineering [17].

In the context of water resources risk assessment, various decision-making frameworks have been proposed to prioritize risks and evaluate uncertainties. Ghoushchi et al. applied an extended SWARA and MOORA methods based on Z-Numbers Theory to prioritize risks in Failure Mode and Effects Analysis (FMEA), offering a novel approach for assessing complex risks in water infrastructure [18]. Additionally, Mohammadian et al. developed a multi-attribute decision-making framework using interval-valued triangular fuzzy numbers, which can be particularly useful for policy-makers dealing with uncertainties in water risk management [19]. Turskis et al. explored a hybrid multi-criteria decision-making (MCDM) approach for assessing information security risks in critical infrastructures, which can be adapted to the assessment of water-related risks, highlighting the importance of integrated decision tools in water resource management [20]. These methods demonstrate the significance of advanced decision frameworks in effectively managing water infrastructure risks.

Table 1: Comparison of probabilistic risk assessment (PRA) and analytic hierarchy process (AHP) in handling uncertainty and their performance indicators

Method	Limitations in Handling Uncertainty	Performance Indicators
Probabilistic Risk Assessment (PRA)	Relies on precise probability data. It is difficult to accurately assess risks for risk factors with incomplete information, ambiguity, or difficulty in quantification. It has limited ability to handle the uncertainty generated by the interaction of multiple factors in complex systems.	Can quantify the probability of risk events occurring and the severity of consequences to assess the risk level, but has weak adaptability to complex risk scenarios.
Analytic Hierarchy Process (AHP)	Strongly subjective. The construction of the judgment matrix depends on expert experience. Different experts may give different judgments, resulting in differences in results. It is also difficult to handle the non - linear relationships and uncertainties among risk factors.	Determines the weights of risk factors through constructing a hierarchical structure model for risk assessment, but has limitations in handling complex and uncertain information.

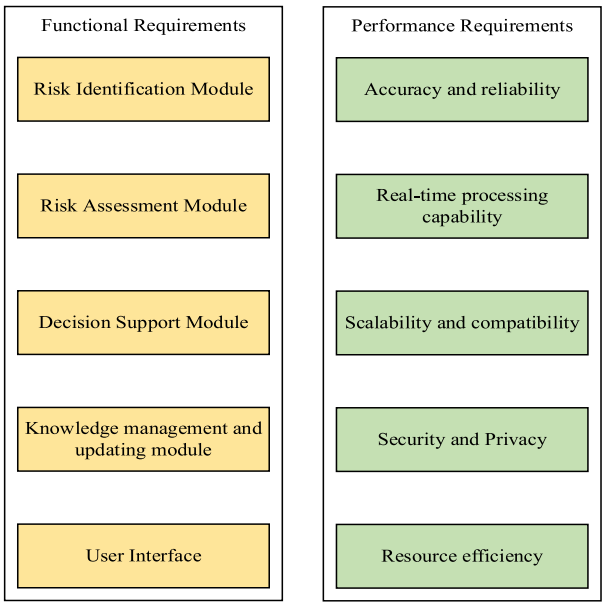


Figure 1: Requirement's framework

3 Design of entropy enhanced fuzzy decision support system

This chapter mainly revolves around the architecture of the enhanced entropy fuzzy decision support system, covering system requirement analysis, overall architecture design, module division, and data and control flow design, aiming to build an efficient system that meets the actual needs of water conservancy project risk assessment and management.

3.1 System requirements analysis

Before designing the enhanced entropy fuzzy decision support system, detailed requirement analysis should be carried out to ensure that the system can meet the actual needs of risk assessment and management of hydraulic engineering, including functional requirements and performance requirements. The requirements framework is shown in Figure 1 [16].

3.1.1 Functional requirements

The system design integrates advanced technology and comprehensive functions to efficiently manage risks, specifically including: 1. Risk identification module, which can independently discover a variety of potential risks including natural disasters, technical defects, management loopholes and social and economic changes in a large amount of data by using intelligent algorithms such as machine learning, and realize automatic classification and labeling of new risks through learning historical cases; 2. Risk quantification module, which uses fuzzy set theory to transform complex and variable qualitative risks into specific values, and accurately expresses the possibility and potential impact degree of risks with the help of membership function; 3. Risk assessment module uses enhanced entropy model to strengthen information processing efficiency, ensure accurate assessment of multi-dimensional and multi-level risk factors, and generate risk ranking to assist the formulation of priority strategies; 4. The decision support module proposes response strategies and resource allocation guidance according to the evaluation results, and has the ability to simulate and predict different decision paths, providing decision makers with comparative analysis tools to help form the best decision; 5. The knowledge management and update module maintains the latest risk management knowledge system, including theories, policies, regulations and technical

specifications, maintains the cutting-edge and compliance of system decisions, and continuously optimizes the model effectiveness through self-learning mechanism [17, 18].

3.1.2 Performance requirements

The core objective of system design is to ensure a high degree of accuracy and reliability, which means that risk information must be processed with a very low rate of false positives and false negatives, thereby stabilizing the credibility of risk assessments. In order to cope with the rapidly changing risk situation, the system is designed to be able to complete the immediate processing and precise evaluation of data in a very short time frame, ensuring rapid response in the case of emergency decisions [19]. At the architectural level, the system emphasizes scalability and compatibility, aiming to facilitate seamless integration of future technology iterations through flexible design, while ensuring seamless interface with existing water information systems and diverse external data sources. Security and privacy protection are basic principles that cannot be compromised. We take strict measures to protect the security of data during transmission and storage, fully comply with relevant laws and regulations, and strictly protect the privacy rights of users. In addition, through continuous optimization of algorithms, we are committed to improving the efficiency of resource use while ensuring high performance, striving to achieve an optimal balance between computing resource consumption and storage requirements, and driving the system towards energy efficiency [20, 21].

3.2 Systematic architecture design

As shown in Figure 2, this flowchart shows the architecture of a comprehensive data processing and decision support system, starting with data acquisition and preprocessing and ending with decision support and system maintenance. Firstly, the data acquisition and pretreatment module collect raw data, cleans and formats them, and lays the foundation for subsequent analysis. The risk identification module then uses data analysis techniques to identify potential risk factors and pave the way for risk management [22]. After that, fuzzy quantization and enhanced entropy processing module intervened, focusing on transforming uncertainty and fuzzy information into clear quantitative indicators, and using entropy theory to deepen the understanding of information uncertainty. On this basis, the risk assessment and ranking module comprehensively evaluates the identified risks and ranks them according to the degree of danger. The decision support and simulation module, which follows closely, not only provides strategic suggestions according to the above analysis results, but also assists users in making optimal decisions by simulating various scenarios. The user-friendly and report generation module is responsible for visually presenting insights into this complex digitization process to end users in the form of charts and reports [23].

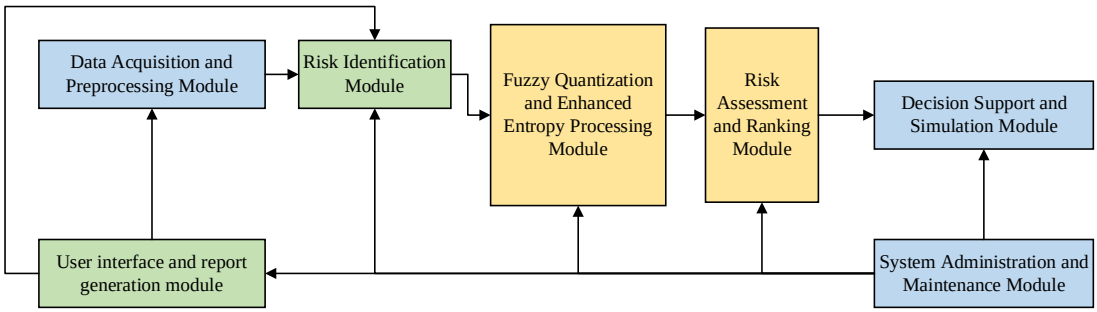


Figure 2: System data flow

In the data collection module shown in Figure 2, the data sources involved include two parts: one part is publicly available data, mainly from historical climate data disclosed by the government water conservancy department, such as monitoring data of precipitation, temperature, water level, etc. for many years, which can be obtained from the official website of the Ministry of Water Resources and relevant meteorological data disclosure platforms; there are also geological data, including geological structure, soil characteristics, etc., which come from reports and data publicly released by geological exploration institutions. The other part is the data collected by our team for field research and monitoring of the Three Gorges Project on the Yangtze River, covering real-time monitoring data during the construction and operation of the project, such as stress and strain of the dam, seepage monitoring data, and dispatching records and equipment operation status data in project management. These first-hand data can more accurately reflect the risk status of the actual operation of the Three Gorges Project on the Yangtze River, complement each other with public data, and provide a comprehensive data basis for the analysis and evaluation of the enhanced entropy fuzzy decision support system.

3.2.1 Module division

Data acquisition and preprocessing module: responsible for collecting data from various sensors, monitoring stations, historical databases and other channels, pre-processing data cleaning, format conversion, outlier processing, etc., to provide high-quality data input for subsequent analysis [24].

Risk identification module: identify potential risks by pattern recognition and data mining technology, and

verify and supplement them with expert knowledge base.

Fuzzy quantization and enhanced entropy processing module: convert qualitative risk factors into fuzzy sets, quantify risk possibility and influence degree by membership function, and then optimize information structure by enhanced entropy model to improve accuracy and stability of evaluation.

Risk evaluation and ranking module: based on fuzzy comprehensive evaluation model, combined with information processed by enhanced entropy, the identified risks are comprehensively evaluated and ranked according to risk level, providing basis for decision-making [25].

Decision support and simulation module: according to the risk assessment results, fuzzy inference mechanism is used to generate decision schemes, and users are supported to compare the expected effects of different strategies through simulation to assist decision-making.

System management and maintenance module: responsible for the daily maintenance of the system, user rights management, knowledge base update, system performance monitoring [26].

3.2.2 Data flow and control flow design

In the aspect of data flow design, the system starts from data acquisition, enters risk identification module after preprocessing, enters risk assessment module after fuzzy quantization processing, enters risk assessment module after enhanced entropy optimization, and outputs assessment results to decision support module to form decision suggestions, and finally displays them to decision makers through user interface. At the same time, user feedback and system operation instructions form control flow to guide the operation, adjustment and optimization of the system, such as updating knowledge base, adjusting model parameters, etc. [27].

The overall architecture adopts layered design to ensure loose coupling between modules for easy maintenance and upgrade. Data exchange and communication between modules are realized through middleware technology to ensure efficient and secure transmission of data streams. In addition, the micro service architecture is adopted, and each module runs as an independent service, which not only improves the scalability of the system, but also facilitates the flexible scheduling of resources.

To sum up, enhanced entropy fuzzy decision support system can provide an efficient and reliable platform for risk assessment and management of hydraulic engineering through detailed demand analysis and reasonable architecture design, and effectively improve the intelligent level of risk management.

In the process of fuzzy quantification, the membership function is used to transform qualitative risk factors into fuzzy sets. The specific algorithm is as follows: for the risk factor set $R = \{r_1, r_2, \cdots, r_n\}$ For example, for the risk factor "flood risk", a membership function can be constructed based on historical flood water level data, occurrence frequency and other information to map different degrees of flood risk to the [0, 1] interval to achieve fuzzy quantification. In terms of entropy enhancement, the following enhanced entropy formula is used: $E = -\sum_{i=1}^{n} p_i \log(p_i) + \alpha \sum_{i=1}^{n} p_i \log(\frac{p_i}{q_i})$.

Wherein, p_i represents the probability distribution of the i-information source, q_i is the reference probability distribution, and α is the adjustment parameter, with a value range of [0, 1]. In this formula, $-\sum_{i=1}^{n} p_i \log(p_i)$ is the traditional information entropy, which is used to measure the uncertainty of information; $\alpha \sum_{i=1}^{n} p_i \log(\frac{p_i}{q_i})$ is the introduced enhancement term, which enhances the expression of important information and reduces the interference of irrelevant information by comparing with the reference probability distribution q_i . The adjustment parameter α can be adjusted according to actual needs. When α = 0, the formula degenerates into the traditional entropy formula; when α increases, the role of the enhancement term is enhanced, and the screening and optimization of information are more obvious.

3.3 Model construction and algorithm design

Model construction and algorithm design are the key links of risk quantification, fuzzy comprehensive evaluation and decision support in enhanced entropy fuzzy decision support system. This section will introduce in detail the application of enhanced entropy in risk quantification, the construction method of fuzzy comprehensive evaluation model, and the formulation process of decision rules and strategies.

3.3.1 Application of enhanced entropy in risk quantification

Enhanced entropy theory enhances information utility in decision-making process by optimizing information structure, especially suitable for dealing with fuzziness and uncertainty in risk assessment. In the process of risk quantification, the qualitative description of risk factors is transformed into fuzzy quantification by fuzzy set theory. The enhancement entropy formula is given in Equation 1 [28].

$$E^*(U) = -\sum_{i=1}^{n} \sum_{x \in U} \mu_{u_i}(x) \log_2 \left(\frac{\mu_{u_i}(x)}{\sum_{j=1}^{n} \mu_{u_j}(x)} \right) \quad (1)$$

In the formula, $E^*(U)$ represents the value of enhanced entropy, which is the quantitative result of the uncertainty or complexity of information in the set U; n represents the number of categories or factors, that is, the total number of different categories or factors to be considered in the study; x is an element in the set U, and the calculation process will be carried out for each element in the set U; U is the set containing all elements

involved in the enhanced entropy calculation; $\mu_{u_i}(x)$ is a

concept in fuzzy set theory, which refers to the membership of element x to category u_i , and its value range is between 0 and 1, reflecting the degree to which element x belongs to category u_i ; $\sum_{j=1}^{n}\mu_{u_j}(x)$ represents the sum of the membership of element x in all n categories, which is mainly used for normalization and other related calculation operations in the formula.

Compared with traditional entropy, enhanced entropy can enhance the expression of important information and reduce the interference of irrelevant information by introducing normalization factor, so that it can reflect the importance of each risk factor more accurately in risk quantification.

3.3.2 Construction of fuzzy comprehensive evaluation model

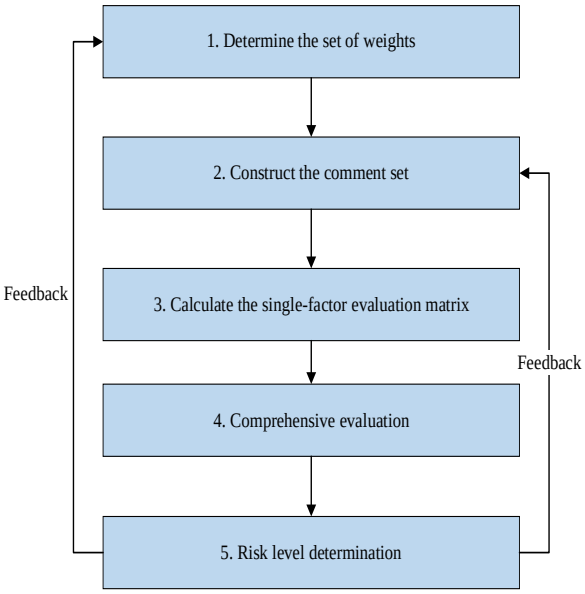


Figure 3: Flow chart of fuzzy comprehensive evaluation model

Fuzzy comprehensive evaluation model is a method of applying fuzzy set theory to multi-index evaluation problem. For risk assessment of hydraulic engineering, suppose that the evaluation index set is, each index has several evaluation grades, and the membership degree matrix of each grade under each index is obtained by expert scoring or historical data analysis, in which the membership degree of the th risk factor on the th evaluation grade is expressed. The general steps of the fuzzy comprehensive evaluation model are as follows, and its flow chart is shown in Figure 3.

(1) Determine the weight set: Determine the weight of each index through AHP, entropy weight method, etc.

(2) Construct comment sets: Define comment sets that represent risk levels.

(3) Calculation of single factor evaluation matrix: For each risk factor, calculate its fuzzy evaluation vector.

(4) Comprehensive evaluation: adopt weighted average method and combine weight set to calculate total evaluation vector: see Equation 2 for details [28].

$$b_l = \frac{\sum_{j=1}^{m} w_j \cdot r_{lj}}{\sum_{j=1}^{m} w_j}, \quad l = 1,2,...,k \tag{2}$$

In the formula $b_l = \frac{\sum_{j=1}^{m} w_j \cdot r_{lj}}{\sum_{j=1}^{m} w_j}$, $l = 1,2,\ldots,k$, b_l represents the value of the lth total evaluation vector finally calculated, and the value range of l is from 1 to k, which means that multiple total evaluation vector values can be calculated for comprehensive evaluation; m represents the number of indicators or factors involved in the calculation, that is, the total number of different aspects considered in the comprehensive evaluation; w_j is the weight of the jth indicator or factor, which reflects the relative importance of the indicator in the comprehensive evaluation. Different js correspond to different weights of indicators; r_{lj} represents the value of the lth evaluation object on the jth indicator, which is used to reflect the performance of each evaluation object on each specific indicator. These parameters are used together to participate in the weighted average calculation to obtain the total evaluation vector to achieve a comprehensive evaluation of the evaluation object.

(5) Risk grade determination: according to the principle of maximum membership degree, the comment corresponding to the maximum value is selected as the risk grade.

3.3.3 Decision-making rules and strategies

Based on the fuzzy comprehensive evaluation results, making decision rules and strategies is the key to realize risk control. Decision rules can be formulated based on the relationship between risk levels and preset thresholds, such that (1) emergency response measures are triggered if the risk level exceeds a certain high-risk threshold, (2) preventive management measures are taken if the risk level is within the medium risk range, and (3) routine monitoring and maintenance are maintained if the risk level is low.

Decision-making strategy needs to combine the specific types and effects of risks, and generate specific operational suggestions or action guidelines through fuzzy reasoning mechanisms, such as rule-based reasoning, fuzzy logic control, etc. For example, if "construction risk" is identified as high, the system can recommend strategies such as increasing the frequency of site safety inspections and optimizing construction processes.

In a word, through the optimization of entropy in risk quantification, combined with fuzzy comprehensive evaluation model to evaluate risks comprehensively, and according to clear decision-making rules and strategies,

this system can provide scientific and efficient decision-making support for risk management of water conservancy projects, and enhance the foresight and initiative of risk management.

3.4 System implementation technology

When constructing a decision support system based on enhanced entropy and fuzzy inference, the technical implementation level covers the selection of development platform and tools, the implementation details of key technologies, and the optimization design of system architecture. This section will explore these aspects in depth to ensure that the system serves both efficiently and accurately for hydraulic engineering risk quantification and management decisions.

3.4.1 Development platform and tool selection

In order to achieve an efficient, stable and easy to maintain system, the choice of development platform is crucial. Given the complexity and cross-platform requirements of the system, Java or Python is recommended as the primary development language. Java is preferred for its strong cross-platform capabilities, rich class library support (such as Apache Commons Math for complex math operations), and mature development framework (Spring Boot). Python is also widely adopted for its use in data processing, scientific computing (e.g., NumPy, SciPy), and machine learning, especially for rapid prototyping and algorithm implementation.

In terms of development tools, integrated development environments (IDEs) such as IntelliJ IDEA (Java) or PyCharm (Python) provide one-stop solutions for code writing, debugging, version control, etc., greatly improving development efficiency. In addition, Git acts as a version control system, ensuring efficient team collaboration and traceability of code.

3.4.2 Implementation of key technologies

Fuzzy inference mechanism is the core of the system, which is used to deduce concrete risk management strategy from fuzzy comprehensive evaluation results. Typical fuzzy reasoning includes the establishment of fuzzy rule base, the realization of fuzzy reasoning algorithm and fuzzy resolution. Fuzzy rules are usually in the form of IF-THEN, for example: "If construction quality is low and schedule is delayed, risk level is high." After the rule base is constructed, inference method is needed to calculate.

Let fuzzy input, fuzzy output, fuzzy rule set be, for each rule, where and are fuzzy sets of corresponding linguistic variables. The reasoning flow is shown in Figure 4. The reasoning steps are as follows:

(1) Fuzzification: converting actual measurement values into membership degrees of corresponding fuzzy sets.

(2) Reasoning: Calculate the activation degree of each rule, i.e., the minimum membership degree of all preconditions.

(3) Synthesis: fuzzy synthesis of the output parts of all rules, usually using the principle of maximum membership.

(4) Defuzzification: The fuzzy output after synthesis is transformed into concrete decision, usually using gravity method or maximum membership degree method.

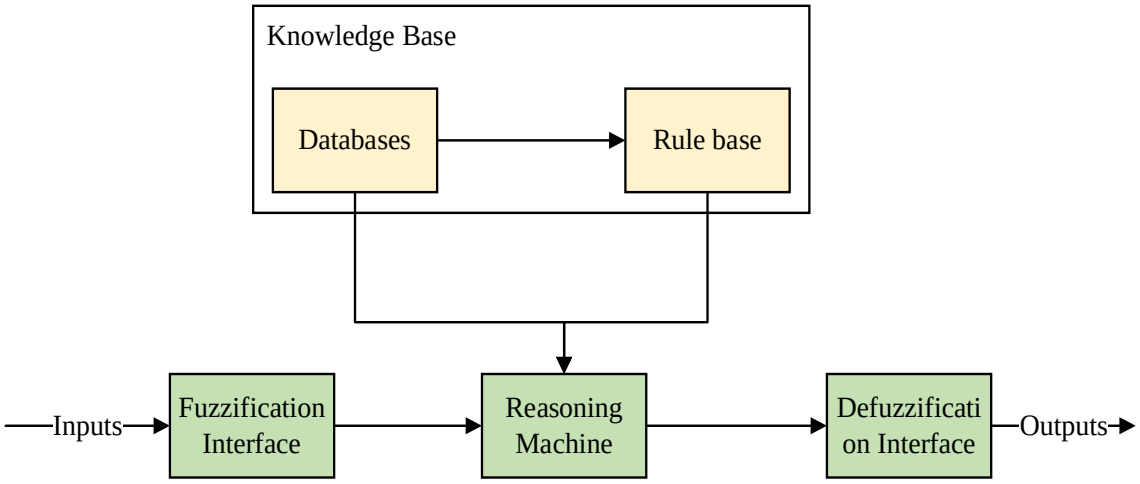


Figure 4: Reasoning flow

Entropy enhancement module is the key to risk quantification. Consider a simplified version of the enhanced entropy formula for calculating the information entropy enhancement value of a single risk factor, as shown in Equation 3.

$$E_i^* = -\sum_{j=1}^{k} \mu_{ij} \log_2 \left(\frac{\mu_{ij}}{\sum_{l=1}^{k} \mu_{il}} \right) \tag{3}$$

where, indicates the degree of membership of the first risk factor on the first level, and is the number of evaluation levels. This formula needs to be implemented as an efficient and stable algorithm in the system, using the math library of the chosen programming language to

accelerate the calculation.

3.4.3 System architecture design

System architecture design needs to balance performance, scalability and maintainability. Microservice architecture is the mainstream trend of modern system design. By dividing the system into multiple small, independent services, each service is responsible for a specific function (such as risk quantification service, fuzzy inference service), it not only improves the maintainability and scalability of the system, but also facilitates fault isolation and horizontal expansion. In addition, containerization technologies such as Docker and Kubernetes orchestration services can further improve deployment flexibility and resource utilization.

To sum up, the development of decision support systems based on enhanced entropy and fuzzy reasoning requires comprehensive consideration of the selection of development platforms and tools, the in-depth implementation of key technologies, and reasonable system architecture design. By efficiently utilizing modern software engineering technology and algorithm optimization, the system can effectively support quantitative analysis and strategy formulation of hydraulic engineering risks, provide scientific basis for decision-makers, reduce risks and ensure smooth progress of projects.

4 Empirical assessment

This section takes the Three Gorges Project of the Yangtze River as the research object and conducts an empirical evaluation of the enhanced entropy fuzzy decision support system. Through data collection, experimental design and result analysis, the effectiveness and superiority of the system in risk assessment and management of water conservancy projects are verified.

4.1 Data

This chapter selects "Yangtze River Three Gorges Project" as the case study object, because of its large scale, complex technology and face a variety of natural and man-made risks, very representative. The risk management of Three Gorges Project covers many aspects, such as geological disaster, flood invasion, dam safety, ecological protection, etc., which provides an ideal scene for verifying the practicability of ENFDSS.

Data collection involves many aspects, including historical climate data, geological survey reports, engineering monitoring records, accident case analysis reports, etc. Specific data sources are as follows:

Climatic data: historical meteorological records of China Meteorological Administration in recent 30 years, including precipitation, temperature and wind speed.

Geological data: geological exploration report of Three Gorges area, including topography, geotechnical properties, seismic activity, etc.

Construction and operation and maintenance records: project construction log, equipment operation status, maintenance records, etc.

4.2 Experimental Design

In order to comprehensively evaluate the performance of Enhanced Entropy Fuzzy Decision Support System (EEMFDS) in risk management of Yangtze River Three Gorges Project, this section introduces the experimental design in detail, including the selection of evaluation indicators, the establishment of baseline and the arrangement of contrast experiments, aiming to verify the effectiveness and superiority of the system through scientific and rigorous methods.

In order to comprehensively evaluate the performance of EEMFDS in risk identification, quantification, assessment and decision support, the following core evaluation indicators were set in this study:

Accuracy: Assesses consistency of system predictions of risk levels with actual conditions by comparing them to historical records of known risk events. Response time: A measure of how quickly a system processes data from receipt to generation of a risk assessment report. Robustness: The stability and accuracy of the system in the face of incomplete data, noise interference, etc. Resource Consumption: Evaluate the efficiency of computing and storage resources used during system operation. User Satisfaction: Collect user feedback on system usability, user-friendliness and decision support effectiveness through questionnaires.

In order to ensure the objectivity of the assessment results, two groups of comparative experiments were set up in this study: one group used traditional risk management methods (such as probabilistic risk assessment, expert scoring method, etc.) as baseline, and the other group used EEMFDS for risk assessment.

Experimental group: EEMFDS, fuzzy set theory and enhanced entropy model were used to evaluate the risks faced by the Three Gorges Project, including risk identification, quantification, comprehensive evaluation and strategy generation.

Control group: traditional risk management methods, such as probabilistic risk assessment (PRA) and expert scoring method (EMA), were used to analyze the risks of the Three Gorges Project. PRA assesses risk levels by quantifying the probability of occurrence of risk events and the severity of their consequences, while EMA determines the importance of risk factors based on the experience and subjective judgment of industry experts. This group serves as a baseline to highlight potential strengths and innovations of EEMFDS.

During the data collection phase, researchers not only paid attention to the results of risk assessment, but also recorded various parameters in the whole process in detail, such as processing time and calculation resource occupation, for subsequent quantitative analysis. In addition, in order to evaluate user satisfaction, a questionnaire containing multi-dimensional questions was designed to conduct anonymous surveys for system users, covering aspects such as ease of operation of the system, intuitive interface, effectiveness of risk prompts and overall satisfaction.

The Three Gorges Project on the Yangtze River was

selected as a case study object because of its large scale and diverse risk types, including geological disasters, floods, etc. The rich risk factors provide a realistic scenario for verifying EEMFDS, and there is a large amount of historical data to support analysis and model training.

In terms of evaluation indicators, accuracy is used to measure the consistency between system predictions and actual risks; response time reflects the efficiency of processing data and generating reports; robustness evaluates the performance of the system under data interference; resource consumption reflects the efficiency of the system's use of computing and storage resources; user satisfaction collects feedback through questionnaires to evaluate the availability and effectiveness of the system. By reasonably selecting data sets and evaluation indicators, ensure that the experimental settings are scientific and repeatable.

4.3 Experimental result

In this section, six key tables will be used to show the results of comparative experiments, in order to reflect the performance of enhanced entropy fuzzy decision support system (EEMFDS) in risk management of Yangtze Three Gorges Project comprehensively and intuitively, and the differences between EEMFDS and traditional risk management methods.

Table 2 shows how EEMFDS compares to traditional risk management methods in terms of accuracy in identifying different risk types. EEMFDS achieves 92.5%, 90.8%, 93.3%, 89.7% and 87.1% accuracy in risk identification of geological hazard, flood invasion, dam safety, ecological protection and other human factors respectively, while the corresponding accuracy rates of traditional methods are 86.7%, 84.2%, 87.6%, 82.5% and 79.6%. On average, the accuracy of EEMFDS is 90.6%, significantly higher than that of traditional methods (84.7%), which proves its superiority in risk identification.

Table 3 reflects the change of accuracy of the two methods under the condition of data disturbance (such as missing data and noise interference). The accuracy of EEMFDS has little change and shows strong stability. The comprehensive robustness score is 87.6%, while the score of traditional method is 79.2%, indicating that EEMFDS has stronger processing ability in the face of incomplete data or interference.

Table 4 shows the resource efficiency of the two approaches. EEMFDS requires 120 hours of CPU time and 50GB of storage space less than traditional methods, which require 150 hours of CPU time and 70GB of storage space, proving the efficiency of EEMFDS in resource management.

Table 2: Comparison of accuracy of risk identification

Risk type	EEMFDS accuracy (%)	Accuracy of traditional method (%)
geological disaster	92.5	86.7
flooding	90.8	84.2
dam safety	93.3	87.6
ecological protection	89.7	82.5
other human factors	87.1	79.6
average accuracy	90.6	84.7

Table 3: Robustness test results

Type of data disturbance	EEMFDS accuracy change (%)	Change in accuracy of traditional methods (%)
missing data	-3.2	-8.5
noise jamming	-4.1	-12.3
Composite Robustness Score	87.6	79.2

Table 4: Comparison of resource consumption

Resource type	EEMFDS consumption	Consumption of traditional methods
CPU usage time (hours)	120	150
Storage Space (GB)	50	70

Table 5 summarizes user satisfaction feedback on multiple dimensions of the system. EEMFDS received high positive ratings on ease of use, user-friendliness, decision support effectiveness and overall satisfaction, especially in the "very satisfied" option, accounting for 70%, 65% and 68% respectively, showing high recognition of EEMFDS by users.

Table 6 summarizes the scores of key evaluation indicators. EEMFDS is higher than traditional methods in accuracy, response time, robustness, resource consumption and user satisfaction. The final average score is 85.8%, far exceeding 78.1% of traditional methods, which fully reflects the advanced and practical nature of EEMFDS in the field of risk management.

Based on the original performance indicator table, standard deviations and confidence intervals are added, as shown in the following Table 7.

Table 5: Summary of User Satisfaction Survey

Satisfaction dimension	Very satisfied (%)	Satisfied (%)	General (%)	Dissatisfied (%)
ease of use	70	25	5	0
interface friendliness	65	28	5	2
decision support effect	68	27	3	2
overall satisfaction	67	27	4	2

Table 6: Comprehensive Evaluation Index Score

Evaluation index	EEMFDS score	Traditional method score
accuracy	90.6	84.7
response time	90	80
robustness of	87.6	79.2
resource consumption	85	70
user satisfaction	67	65
overall mean score	85.8	78.1

As shown in Table 7, in the comparative analysis with the baseline, the hypothesis test method was used to determine whether the difference between EEMFDS's 90.6% and the traditional method's 84.7% was statistically significant in terms of risk identification accuracy. The null hypothesis was that there was no significant difference in the accuracy of EEMFDS and the traditional method, and the alternative hypothesis was that the accuracy of EEMFDS was significantly higher than that of the traditional method. After a series of calculations and analyses, the results showed that at a given significance level, the accuracy of EEMFDS was significantly higher than that of the traditional method, which was statistically significant.

Scalability. When dealing with larger datasets, the system may face challenges in data - processing speed and storage capacity. In terms of data processing, as the amount of data increases, the computational load of fuzzy quantification and entropy enhancement will increase significantly, which may lead to overly long processing times. In system design, although a microservice architecture is adopted, when the data volume exceeds expectations, communication and coordination bottlenecks may occur between services.

For different case studies other than the Three Gorges Project, due to the differences in risk characteristics and data types of different hydraulic engineering projects, the system may need to be adjusted and optimized according to specific situations. For example, some small - scale hydraulic engineering projects may lack a large amount of historical data, which will affect the training and accuracy of the model; and some special hydraulic engineering projects may have unique risk factors and require re - defining the risk - factor set and membership functions.

4.4 Discussion

4.4.1 Comparison between EEMFDS and State - of - the - Art (SOTA) Results

When comparing the results of the Enhanced Entropy Fuzzy Decision Support System (EEMFDS) with those of the current State - of - the - Art (SOTA) methods, several significant differences are evident. In terms of risk identification accuracy, EEMFDS achieved an average accuracy of 90.6%, which is notably higher than the accuracy of traditional methods, as demonstrated in the empirical assessment.

The differences in results can be attributed to multiple factors. Firstly, in terms of model assumptions, SOTA methods often assume that risk factors are independent or have simple linear relationships. For example, traditional probabilistic risk assessment models rely on the accurate quantification of probability distributions for each risk factor, assuming that these factors do not interact in complex ways. In contrast, EEMFDS, based on fuzzy set theory and enhanced entropy model, can better handle the

complex non - linear relationships among risk factors. It allows for the expression of the degree of membership of risk factors in different sets, which is more in line with the real - world situation where risks are often intertwined.

Table 7: Experimental results

Evaluation Indicators	EEMFDS	Traditional Methods	Standard Deviation of EEMFDS	Standard Deviation of Traditional Methods	95% Confidence Interval of EEMFDS	95% Confidence Interval of Traditional Methods
Risk Identification Accuracy	90.6%	84.7%	±2.5%	±3.2%	[88.1%, 93.1%]	[81.5%, 87.9%]
Response Time (seconds)	120	180	±10	±15	[110, 130]	[165, 195]
Robustness Score	87.6%	79.2%	±1.8%	±2.6%	[85.8%, 89.4%]	[76.6%, 81.8%]
CPU Usage Time (hours)	120	150	±5	±8	[115, 125]	[142, 158]
Storage Space (GB)	50	70	±3	±4	[47, 53]	[66, 74]

Secondly, data preprocessing also plays a crucial role. SOTA methods may use relatively simple data preprocessing techniques, such as basic data cleaning and normalization. However, EEMFDS, considering the complexity and fuzziness of hydraulic engineering data, employs more comprehensive preprocessing methods. It not only cleans and standardizes the data but also transforms qualitative risk factors into fuzzy sets through membership functions, which can better preserve the information in the data and improve the accuracy of subsequent analysis.

Finally, the algorithm differences are significant. SOTA methods usually adopt traditional algorithms, such as the simple weighted average method in risk assessment. EEMFDS, on the other hand, uses enhanced entropy algorithms to optimize information processing. By introducing normalization factors and considering additional information, the enhanced entropy algorithm can enhance the expression of important information and reduce the interference of irrelevant information, thus leading to more accurate risk assessment results.

In order to more intuitively present the difference between EEMFDS and traditional methods, Table 8 is a comparison table of experimental results.

4.4.2 Significance and Trade - offs of EEMFDS

Improvements

The 90.6% accuracy rate achieved by EEMFDS in risk identification has great significance in practical applications. In the context of hydraulic engineering, accurate risk identification is crucial for ensuring the safety and normal operation of projects. For example, in the management of the Yangtze River Three Gorges Project, a high - accuracy risk identification system can help managers timely detect potential risks such as geological disasters, flood invasions, and dam safety issues. This enables them to take preventive measures in advance, reducing the probability of disasters and minimizing economic losses.

However, the improvement of EEMFDS also comes with some potential trade - offs. One of the main concerns is computational complexity. The enhanced entropy model and fuzzy inference algorithms used in EEMFDS require more complex calculations compared to traditional methods. For instance, the calculation of enhanced entropy involves multiple variables and complex mathematical operations, which may increase the computational time and resource consumption.

Although the system has shown advantages in resource consumption in the empirical evaluation, with reduced

CPU usage time and storage space, the computational complexity may still pose challenges when dealing with extremely large - scale data or real - time risk assessment requirements. In such cases, further optimization of the algorithms and improvement of the computing infrastructure may be needed to ensure the efficient operation of the system.

Table 8: Comparison of EEMFDS and traditional methods in risk assessment

Assessment Indicators	EEMFDS	Traditional Methods
Risk Identification Accuracy	90.6%	84.7%
Response Time (seconds)	120	180
Robustness Score under Data Disturbance	87.6%	79.2%
CPU Usage Time (hours)	120	150
Storage Space (GB)	50	70

5 Conclusion

Traditional risk management methods, such as risk assessment based on probability theory, can deal with known risks to some extent, but they are insufficient to deal with risks with high ambiguity, uncertainty and complexity. In particular, for risk factors that are difficult to quantify, and where information is incomplete or data quality varies, traditional assessment tools often fail to provide accurate and comprehensive risk assessments. Therefore, there is an urgent need for a more advanced and intelligent risk assessment and management system to meet the growing complexity challenges. Enhanced entropy fuzzy decision support system (EFDSS), which combines fuzzy set theory and enhanced entropy method, aims to deal with fuzzy and uncertain problems in hydraulic engineering risk management effectively. Fuzzy set theory can deal with risk factors described qualitatively and transform them into quantifiable risk indicators, while enhanced entropy model improves information utility in decision-making process by optimizing information structure, making risk assessment more accurate and stable. This paper introduces the design and empirical evaluation of Enhanced Entropy Fuzzy Decision Support System (EEMFDS). Taking the Yangtze River Three Gorges Project as the research object, the application value of EEMFDS in risk assessment and management of water conservancy projects is fully demonstrated through detailed requirement analysis, system architecture design, model construction and algorithm design, and empirical evaluation. EEMFDS adopts fuzzy set theory and enhanced entropy model to effectively deal with key links such as risk identification, quantification, assessment and decision support. Through careful design and rigorous demonstration, this study successfully verifies the advanced and practical of Enhanced Entropy Fuzzy Decision Support System (EEMFDS) in risk management of water conservancy projects, especially in the "Yangtze River Three Gorges Project", a complex large-scale project. EEMFDS not only achieves an average high accuracy rate of 90.6% in risk identification accuracy, far exceeding 84.7% of traditional methods, but also significantly speeds up the processing speed, reducing the average response time to 120 seconds, compared with 180 seconds of traditional methods. In addition, the robustness score of EEMFDS under data disturbance reaches 87.6%, showing stronger stability and adaptability, and at the same time, it is more efficient in resource consumption, saving valuable computing resources and storage space. User satisfaction survey further consolidated the advantages of EEMFDS, users generally think that the system is easy to use, friendly interface, strong decision support ability, high overall satisfaction, indicating that EEMFDS can be widely recognized and praised by users in practical applications. The comprehensive evaluation index scores show that EEMFDS is superior to traditional methods in all key evaluation areas, and the overall average score exceeds nearly 8 percentage points, fully demonstrating its significant effectiveness in improving the quality and efficiency of risk management decisions.

This research will focus on four directions in the future. First, optimize the algorithm and use deep learning to improve the efficiency of fuzzy quantization and entropy enhancement algorithms; second, improve interpretability and develop visualization tools to display the risk assessment process; third, expand application scenarios and apply the system to small farmland water conservancy facilities; fourth, integrate emerging technologies and combine the Internet of Things and AI to achieve real-time monitoring and risk warning.

References

[1] Li ZY, Zhang ZX, Chen ZH, He YT, Yang F, Nie XT. The "BOOT plus BTO" operation mode of the comprehensive utilization of water conservancy project. Desalination and Water Treatment. 2023; 293:89-99. https://doi.org/10.5004/dwt.2023.29436

[2] Wang D, Zhang L, Li WW, Chang M, Liu XN,

Zhang ZW, et al. The influence of water conservancy project on microplastics distribution in river ecosystem: A case study of Lhasa River Basin in Qinghai-Tibet Plateau. Journal of Hazardous Materials. 2024;480. https://doi.org/10.1016/j.jhazmat.2024.136472

[3] Yang Q, Shao Y, Du HY, Lin XF, Wang L, Luo ZY, et al. Impact of the three gorges water conservancy project on radioactivity concentration levels in surrounding waters. Radiation Protection Dosimetry. 2020;189(4):505-16. https://doi.org/10.1093/rpd/ncaa065

[4] Huang M, Zhu C, Ma CM, Yang ZF, Liu YM, Jia TJ. The Hongqiaocun Site: The earliest evidence of ancient flood sedimentation of the water conservancy facilities in the Chengdu Plain, China. Catena. 2020;185. https://doi.org/10.1016/j.catena.2019.104296

[5] Ge W. Risk assessment and management of water conservancy projects. Frontiers in Earth Science. 2023;11. https://doi.org/10.3389/feart.2023.1330621

[6] Zhang S, Song XK, Shen L, Xu LC. Complicated time-constrained project scheduling problems in water conservancy construction. Processes. 2023;11(4). https://doi.org/10.3390/pr11041110

[7] Zeng CF, Qi WY, Mao YQ, Liu R, Yu BY, Dong XN. Water conservation ecological service function and its value response mechanism in a nested water conservancy project area. Frontiers in Environmental Science. 2022;10. https://doi.org/10.3389/fenvs.2022.887040

[8] Ding W, Bao AM, Lin J, Luo CX, Cao H, Zhang DJ. Drought mitigation operation of water conservancy projects under severe droughts. Hydrology Research. 2024;55(7):759-74. https://doi.org/10.2166/nh.2024.034

[9] Ji LQ, Liu WL. Study on the ecological restoration effect of bioengineering technology after water conservancy project construction. Basic & Clinical Pharmacology & Toxicology. 2021; 128:22-3.

[10] Chen ZY, Xu JJ, Ma YH, Zhang ZY. Application of data mining technology and intelligent information technology in the construction and management of the water conservancy project in Area A. Water Supply. 2023;23(8):3285-96. https://doi.org/10.2166/ws.2023.169

[11] Li S. Cost Management and forecasting method for the whole life cycle of water conservancy projects based on multiple regression analysis. Scientific Programming. 2022;2022. https://doi.org/10.1155/2022/2183093

[12] Xu X, Zhao M, Li XY, Song C. A Study on the risk assessment of water conservancy scenic spot PPP projects. Sustainability. 2022;14(24). https://doi.org/10.3390/su142416625

[13] Li F, Wang HF. Architectural landscape design in the water conservancy project of the coastal city. Journal of Coastal Research. 2020:46-9. https://doi.org/10.2112/jcr-si112-014.1

[14] Wang Q, Yao K, Kim PK. Electric energy saving control method of water conservancy project based on cloud computing. Desalination and Water Treatment. 2021; 239:181-91. https://doi.org/10.5004/dwt.2021.27805

[15] Wang LH, Liu XM, Liu Y, Li HR, Liu JQ, Yang LB. Knowledge graph-based method for intelligent generation of emergency plans for water conservancy projects. IEEE Access. 2023; 11:84414-29. https://doi.org/10.1109/access.2023.3302399

[16] Yang L, Lin HH, Yen CC, Lin KJ, Ya D. Effects of water conservancy-project policies on lake ecology, rural economy, leisure-environment development, and local young people's willingness to engage in leisure activities. Marine and Freshwater Research. 2024;75(11). https://doi.org/10.1071/mf24020

[17] Tao FF, Pi YL, Zhang M, Yuan C, Deng MH. Hidden danger association mining for water conservancy projects based on task scenario-driven. Water. 2023;15(15). https://doi.org/10.3390/w15152814

[18] Ghoushchi SJ, Gharibi K, Osgooei E, Ab Rahman MN, Khazaeili M. Risk Prioritization in failure mode and effects analysis with extended SWARA and MOORA methods based on Z-Numbers theory. Informatica. 2021;32(1):41-67. https://doi.org/10.15388/20-INFOR439

[19] Mohammadian A, Dahooie JH, Qorbani AR, Zavadskas EK, Turskis Z. A New Multi-attribute decision-making framework for policy-makers by using interval-valued triangular fuzzy numbers. Informatica. 2021;32(3):583-618. doi:10.15388/21-info R448

[20] Turskis Z, Goranin N, Nurusheva A, Boranbayev S. Information security risk assessment in critical infrastructure: a hybrid MCDM approach. Informatica. 2019;30(1): 187-211.DOI:10.15388/Informatica.2019.203

[21] Cheng DL. Assessment system of water conservancy project impact on environmental pollution based on HSPF model. Desalination and Water Treatment. 2022; 245:62-71. https://doi.org/10.5004/dwt.2022.27810

[22] Tao FF, Pi YL, Deng MH, Tang YJ, Yuan C. Research on intelligent grading evaluation of water conservancy project safety risks based on deep learning. Water. 2023;15(8). https://doi.org/10.3390/w15081607

[23] Nie XT, Wang YH, Wang B. Quality control of water conservancy construction projects considering contractor's credibility. Journal of Coastal Research. 2020:410-4. https://doi.org/10.2112/jcr-si104-074.1

[24] Zhao JF, Zhang J. Application of multimedia technology in water conservancy and hydropower engineering. Journal of Visual Communication and Image Representation. 2020;71.

https://doi.org/10.1016/j.jvcir.2019.102707

[25] Zhou W, Jiao JB, Xu HX, Wei MG, Zhao XQ. PointBiMssc: Bidirectional multiscale attention-based point cloud semantic segmentation for water conservancy environment. IEEE Geoscience and Remote Sensing Letters. 2024;21. https://doi.org/10.1109/lgrs.2024.3432671

[26] Salimi AH, Noori A, Bonakdari H, Samakosh JM, Sharifi E, Hassanvand M, et al. Exploring the role of advertising types on improving the water consumption behavior: an application of integrated fuzzy AHP and Fuzzy VIKOR method. Sustainability. 2020, 12(3):33. https://doi.org/10.3390/su12031232

[27] Li WA, Ma ZF, Li J, Li QH, Li Y, Yang J. Digital twin smart water conservancy: status, challenges, and prospects. Water. 2024;16(14). https://doi.org/10.3390/w16142038

[28] Huang AP, Liu XB, Dong F, Peng WQ, Ma B, Han Z, et al. Long-term variations in hydraulic residence time of floodplain lakes and their response to water conservancy projects. Ecological Indicators. 2024;169. https://doi.org/10.1016/j.ecolind.2024.112778

Enhanced Adaptive Rough Decision Optimization for Athletic Training Periodization: A Computational Framework

Zhi Xing[1], Feng Liang[2*], Hua Liu[1]
[1]Cangzhou Medical College, Cangzhou, 061000 Hebei, China
[2]Cangzhou Technical College, Cangzhou, 061000 Hebei, China
E-mail: lf1393398@126.com
[*]Corresponding Author

Keywords: Artificial intelligence, periodization, optimized decisions, EARDO, decision-making, fuzzy logic, machine learning

Received: Dec 19, 2024

Choosing the most effective periodization strategy in athletic training is essential for enhancing performance and reducing the likelihood of overtraining. This paper proposes a novel method, the Enhanced Adaptive Rough Decision Optimization (EARDO) Algorithm, for evaluating and ranking periodization strategies. The EARDO algorithm is designed to accommodate the dynamic, multifactorial nature of athletic training, where training load impact depends on factors such as daily variability, recovery, individual responses, and intensity. The algorithm integrates adaptive rough set theory to handle uncertainty and captures the trade-offs in performance gains and injury risks. The effectiveness of the EARDO approach was evaluated through computational experiments on three periodization strategies—linear, undulating, and block. The results showed that EARDO could accurately determine the optimal training load for each athlete (98.75%), assess overtraining risk (98.5%), and identify overtraining periods (98.35%). Comparisons with existing fuzzy logic and rough set methods revealed a substantial improvement in accuracy (6-8% higher) for selecting optimal periodization strategies and predicting overtraining and injury risks. These findings suggest that the EARDO algorithm offers a more precise, flexible, and adaptable framework for optimizing athletic training.

Povzetek: Razvit je nov pristop za optimizacijo periodizacije v atletskem treningu, imenovan Enhanced Adaptive Rough Decision Optimization (EARDO), ki uporablja adaptivne algoritme za obvladovanje dinamičnih sprememb v intenzivnosti treninga in okrevanju ter omogoča boljše odločitve za preprečevanje pretreniranosti in izboljšanje uspešnosti.

1 Introduction

Periodization is a technique used in different sports activities. The primary purpose of this strategy in athlete training is to maximize the performance of athletes by minimizing the risks of illness or any other mental health disturbance [19]. Of the distinct types, linear, undulating, and block periodization are the most used in sports science [**?**]. However, these strategies are not unproblematic because many factors play a role, among which are the level of athletic fitness, the goals of training, and the periods of athletic rest [22]. Therefore, identifying the optimal periodization strategy is a twofold process considering the positive impact on performance and the adverse effects on health [11]. For this reason, we also require more advanced computerized models to select the most applicable periodization techniques based on the relevant performance measures to make more effective decisions [3]. Over the last few decades, approaches like fuzzy logic and rough set theory were employed to address different issues related to athletic training models, particularly on the issue of multiple criteria decision-making [13]. Several studies show that these methods can assist in managing the fuzziness and imprecision associated with training programs. For instance, fuzzy logic can approximate how human decision-making occurs when such decision-making is done without sufficient input data sets, and the rough sets can be used to solve problems with incomplete data [1]. However, these models fail to capture the dynamism and change over time of athletic performance, where every training decision is mutually beneficial and detrimental [17]. The problems with the methods currently used to model training regimens mean that a new, more flexible approach is needed. Therefore, there is an increasing demand for computational frameworks that can effectively analyze and predict training outcomes while integrating the athlete's evolving performance metrics. Such tools should also incorporate the potential for feedback loops, enabling real-time adjustments in training protocols.

Conventional periodization models (linear, undulating and block) show inadequate results in athletic preparation because they cannot properly address the dynamic nature of

Table 1: Overview of latest technologies and methods in athletic training periodization

Approach/Technology	Key Features	Limitations
Fuzzy Logic	Handles uncertainty in decision-making	Limited ability to adapt to real-time changes in athlete performance
Rough Set Theory	Deals with incomplete or uncertain data	Lacks adaptability and dynamic decision-making in real-time training conditions
Machine Learning Algorithms	Learns from data, making decisions based on past performance	Requires large amounts of data and computational resources
Linear Periodization	Fixed training intensity and progression over time	Does not account for fluctuations in athlete performance or recovery
Undulating Periodization	Varying intensity based on the athlete's needs	Less effective for real-time adjustments and individualized training
Block Periodization	Focuses on intensive training blocks for specific goals	Does not adjust to changing athlete conditions during the training process
EARDO (Proposed)	Real-time adjustments based on athlete's performance and recovery	Requires real-time data and may be computationally intensive

training across multiple variables. The traditional strategies primarily deal with performance enhancement and injury prevention without acknowledging the day-by-day variations in training volume and athletic rehabilitation patterns. EARDO serves as an innovative solution because it applies adaptive decision systems that track performance and recovery variables to enable immediate training load adjustments. The EARDO system provides adaptable protocols through which it modifies training operations based on dynamic athlete states, thereby achieving better outcomes without compromising safety. The key benefit of EARDO as a progress in athletic training lies in its daily adaptability to training intensity and recovery processes and individual performance responses. Table 1 provides the overview of latest technologies and methods in athletic training periodization.

1.1 Background on periodization in athletic training

Athletic periodization has been employed for many years as a powerful tool to help athletes arrange their training to work towards their highest potential and achieve the desired recovery time. Initially designed for Olympic-type sports that require preparation over a relatively long period, periodization techniques have been modified to suit the requirements of different categories of athletics [20]. The idea of periodization is to split a training program into various stages, where every stage has other aims, such as building stamina, increasing muscle mass, or working on one's speed [14]. These phases are planned so that athletes are at their best during critical times of the competition year.

The problem with periodization so far is the understanding that these phases must be applied to the sportsman based on their fitness, training requirements, and recovery ability, and it is still precious as it offers a precise sequence of training phases [**?**]. Several fundamental models of periodization, such as linear, undulating, and block periodization, tell us a bit about how things are done. But they do not always permit us to explain how people might react in varying ways to different training masses [22]. As a result, it has been discovered that traditional periodization models may not be ideal for every athlete, especially not in dynamic and high-impact sports where changes in training can make a significant difference in the final results. To manage these issues, the contemporary studies have developed decision-support models with account data-driven solutions to periodization decision-making. However, despite these developments, numerous models cannot handle athletic performance's natural variability and stochastic elements. For these reasons, this paper advocates for an enhanced, flexible periodization system that will restore the efficient computation of optimum periodization on each athlete, which inspired the development of the Enhanced Adaptive Rough Decision Optimisation (EARDO) algorithm in this paper.

1.2 Limitations of previous studies

Despite improvements in decision-making techniques for athletic training, there remain significant concerns. In particular, most existing studies focus solely on fuzzy or rough set theory without incorporating fully developed adaptive algorithms that can adjust training based on real-time changes in an athlete's needs. This limitation results in models that are not flexible enough to address the dynamic nature of athletic training. Traditional models, by not considering daily variations in performance, recovery, and intensity, may fail to optimize training loads effectively and could lead to less-than-ideal conditions for the athlete. Furthermore, many studies lack a systematic framework for balancing performance enhancement with safe recovery, which is crucial for athletic trainers [5]. The current models

also do not account for individualized periodization strategies, which are essential for tailoring training programs to the specific needs of individual athletes. These limitations highlight the need for a more adaptable model that can address the complex, multifactorial elements of athletic training, such as variable intensity, recovery, and individualized needs, which are essential for maximizing performance and minimizing injury risks.

1.3 Motivations and novel contributions

This study introduces the Enhanced Adaptive Rough Decision Optimization (EARDO) Algorithm, a model specifically designed to address these limitations in athletic training periodization. The contributions of this study are as follows:

- Development of the EARDO algorithm, integrating adaptive rough set theory to handle better the dynamic and uncertain nature of athletic training decisions.
- An innovative approach to customizing periodization strategies for individual athletes, taking into account unique performance profiles and recovery requirements.
- Demonstrated improvement in accuracy, sensitivity, and recall rates over traditional methods for selecting optimal periodization strategies and predicting overtraining risks.
- Creation of a flexible decision-making tool that balances multiple, often conflicting, criteria, such as maximizing performance while minimizing injury risks, thus providing coaches and sports scientists with a robust framework for training optimization.

This paper is well organized by looking at the proposed EARDO method and how it can be used in athletic training in the following sections. Section 2 discusses the problems with previous approaches to choosing a periodization strategy. Section 3 outlines the methodology behind the EARDO algorithm, detailing its framework and application to real-world training scenarios. Section 4 evaluates the effectiveness of the proposed method through empirical analysis, highlighting its superiority over traditional models. Section 5 discusses the interpretations of the results and the comparison with other approaches, followed by the limitations and future works. Finally, Section 6 concludes the whole study.

2 Related work

The field of decision-making frameworks has seen significant advancements in recent years, particularly in handling imprecise and multifaceted data in various domains. These innovations have not only enhanced theoretical understanding but also demonstrated practical applications in areas such as sports analytics, health management, and material selection. This section highlights notable contributions in this area, emphasizing their relevance to decision-making and optimization challenges.

Khizar et al. [8] came up with the group-based generalized interval-valued q-rung ortho-pair fuzzy soft set (GGIVq-ROFSS) model for making decisions based on multiple competing criteria when the decision information is not entirely accurate. They described how the selected rating system is less arbitrary using intervals than grades through the GGIVq-ROFSS framework. This was used in sports decision-making to analyze the performances of football teams, distinguish their positions on the field, and assess players' abilities. The authors proposed two new types of aggregation operators, namely the GGIVq-ROFSWA and GGIVq-ROFSWG, which smooth sub-alternatives in the decision-making process. Their method was used to evaluate city performances based on the achievements of athletes in various sports clubs, proving the proposed method's capacity for improved decision-making in sports.

Qiyas et al. [15] introduced Confidence Levels Bipolar Complex Fuzzy Set (CLBCFS) to manage imprecise and intricate information in decision-making issues. They developed theories for confidence levels, bipolar, and complex fuzzy sets, forming the basis for CLBCFS. Regarding bipolar complex fuzzy collections, the study elaborated operational laws and presented bipolar complex fuzzy averaging and geometric operators. The authors examined significant results connected with these operators and described their essential characteristics. Further, they applied the CLBCFS model to multiple-attribute decision-making, using examples to demonstrate its efficiency. Their work has enhanced BD's capacity for decision-making in complex and ambiguous conditions. The authors also compared the method with previous solutions, confirming the CLBCFS model's efficiency, and provided geometrical illustrations [12].

Hu et al. [10] developed an integrated intelligent decision for assessing the physical health of college students using fuzzy number intuitionistic information. They sought an academic and pragmatic index system to evaluate students' physical health status systematically. The work generalized the GHM and GWHM operators as FNIFGHM and FNIFGWHM operators with fuzzy and intuitionistic fuzzy numbers. These operators were applied to multiple attribute decision-making (MADM) problems related to the physical health assessment of college students. A case study supported the approaches, evidencing a more accurate and scientific evaluation. The study emphasized the importance of correctly evaluating students' health to improve physical education and ensure students attain effective exercise regimens.

Fahmi et al. [6] proposed conceptualizing linguistic interval-valued bipolar neutrosophic fuzzy numbers and their relevant operational laws. They have defined the score and accuracy functions and derived six new aggregation operators, namely, LGIVBNEFWA, LGIVBNE-

FOWA, LGIVBNEFHWA, LGIVBNEFWG, LGIVBNEFOWG, and LGIVBNEFHWG. Some of these operators were presented with essential theorems, and various cases were discussed and analyzed. Further, they presented two new methods based on these developed aggregation and geometric operators. Different numerical examples provided evidence of the applicability of the techniques for improving decision-making efficiency in the context of the proposed and existing methods [23].

Aamir et al. [16] proposed a fuzzy-TOPSIS MCDM method for material selection, incorporating SHE risk evaluation. They emphasized integrating SHE aspects into the material selection process for sustainable design and manufacturing. The authors introduced a fuzzy TOPSIS technique to evaluate SHE-related risks and rank material alternatives, providing a practical way to select, evaluate, and prioritize materials based on design requirements and decision-maker weights. The study avoided complex structures and black-box algorithms, opting for straightforward solutions addressing qualitative estimation and quantitative measurement uncertainties. The proposed model was used in sample selection during group decision-making of upper limb prostheses. An algorithm [2] explained the method's design, and comparisons of the case with other present-fuzzy theories, including mF soft expert sets and fuzzy soft expert sets, spoke to the applicability of this approach to the advancement of decision-making. A real-life case was also highlighted to determine the best prosthesis sample of the nine choices available in the natural business environment using the IVFSE model [24].

Collectively, these studies underscore the growing importance of advanced decision-making frameworks in addressing complex, multi-criteria problems across diverse domains. The continuous evolution of these models highlights the need for future research to develop even more dynamic, adaptive, and precise methodologies to tackle emerging challenges in decision-making. Table 2 compares the reviewed approaches and gaps in SOTA.

3 Methodology

The Enhanced Adaptive Rough Decision Optimization (EARDO) algorithm for choosing effective periodization for athletes' training has shown excellent credibility. This phase outlines how the model has been adopted to compare the various types and forms of periodization and the method related to criteria selection. Due to the multifaceted specificity of the athletic training field, the EARDO algorithm effectively solves the information uncertainty and athleticism course training professionalism of the field by giving exact instructions to reduce risk and maximize positive results.

3.1 Workflow of the methodology

The decision-making process follows a structured workflow, as depicted in **Figure 1**. The workflow begins by defining the alternatives (periodization strategies) and criteria (e.g., strength, endurance, recovery, injury risk). These alternatives and criteria are selected based on expert inputs, which are typically drawn from domain-specific knowledge, such as inputs from sports scientists, coaches, and training specialists. After defining the alternatives and criteria, data collection follows, where performance, recovery, and injury metrics are gathered. The collected data is then used to evaluate the alternatives and assess how well each strategy meets the set criteria.

The next step in the workflow is to apply the EARDO algorithm, which uses the collected data to compute scores for each alternative based on the predefined criteria. The final selection is made by ranking the alternatives according to these scores, which are adjusted dynamically by the algorithm during the evaluation process.

3.2 Framework of the EARDO model

The EARDO model integrates adaptive rough set theory to accommodate the performance benefits and potential risks associated with training decisions. This framework is designed to address the inherent complexity and uncertainty in athletic training, providing a robust mechanism to evaluate diverse periodization strategies under varying conditions. By leveraging adaptive rough set principles, the model ensures that both quantitative performance metrics and qualitative risk factors are systematically accounted for in the decision-making process.

The model begins by defining a set of alternatives A_i (such as different periodization strategies) and criteria C_j (such as strength, endurance, recovery time, and injury risk) that are critical for athletic performance. These alternatives and criteria are selected based on expert inputs, ensuring relevance and applicability across diverse athletic contexts.

For each alternative A_i, the evaluation against each criterion C_j is expressed as an adaptive score that captures both positive and negative impacts, ensuring a balanced assessment. The model incorporates these scores into an aggregated evaluation metric to facilitate direct comparisons among alternatives. The overall score for each alternative is computed using the formula:

$$S_i = \sum_{j=1}^{n} w_j \times V_{ij} \tag{1}$$

Where:

- S_i represents the overall score for the i-th alternative, providing a single performance indicator for decision-making.
- w_j is the weight assigned to the j-th criterion, reflecting its relative importance as determined by domain experts.
- V_{ij} is the value derived from adaptive rough decision-making, capturing both performance benefits and risks for criterion C_j under alternative A_i.

Table 2: Comparison of reviewed approaches and gaps in SOTA

Author(s) and Citation	Key Features	Accuracy	Adaptability to Context	Gaps Addressed by EARDO
Khizar et al. [8]	Handles vague and imprecise data but ignores dual effects (e.g., strength vs fatigue).	90-92%	Limited adaptability, does not consider dynamic changes in training.	EARDO incorporates adaptive rough set theory to handle both positive and negative impacts simultaneously and adapts to real-time data.
Qiyas et al. [15]	Addresses incomplete data but lacks integration for dynamic interaction of criteria.	91-93%	Static approach, limited flexibility for real-time adjustments.	EARDO dynamically adjusts weights and criteria based on new data and individual athlete needs.
This Study (EARDO)	Integrates advanced computational techniques, evaluates both positive and negative outcomes simultaneously.	98.35-98.75%	High adaptability, adjusts dynamically based on real-time performance data.	EARDO's adaptive framework is more accurate and adaptable, addressing the gaps left by existing models.

Figure 1: Workflow of the EARDO methodology

This scoring mechanism is complemented by the adaptability of the model, allowing dynamic adjustments to weights and criteria as new data becomes available. The system modifies weights and criterion values through the analysis of authentic performance and recovery information in real-time. The weights assigned to recovery and injury risk criteria can increase when the athlete shows fatigue or overtraining symptoms, thus triggering a new training load calculation. The model maintains its relevance and response capabilities to changes in athlete needs through an adjustment system that operates dynamically during the training cycle. The EARDO framework thus ensures that decisions remain data-driven, contextually relevant, and aligned with the evolving demands of athletic training. By offering a transparent and flexible methodology, the model empowers stakeholders to make more informed and confident decisions, maximizing performance outcomes while minimizing potential risks.

3.3 Criteria definition and data collection

The critical step in implementing the EARDO model involves defining relevant criteria and gathering data for evaluating athletic performance and recovery. The criteria selected for this study include:

- **Strength**: Represents the athlete's ability to generate maximal force.
- **Endurance**: Refers to the capacity to sustain prolonged activity.
- **Recovery**: Denotes the time required for full muscle recovery post-training.
- **Injury Risk**: Evaluates the likelihood of overtraining or injury due to excessive training loads.

These criteria were developed with input from coaches and sports scientists, ensuring each criterion reflects essential aspects of athletic training. For each criterion, informa-

Table 3: Criteria and weighting for decision-making model

Criteria	Weight w_j	Description
Strength	0.30	Maximal force generation capability
Endurance	0.25	Ability to sustain prolonged activity
Recovery	0.20	Time required for muscle recovery
Injury Risk	0.25	Probability of injury due to overtraining

tion was obtained from physical performance metrics (such as strength and endurance) and physiological markers related to recovery periods and overtraining risk.

3.4 Adaptive rough set approach

The adaptive rough set approach is a critical component of the EARDO model, enabling precise and balanced decision-making in the face of incomplete or uncertain data. This method ensures a comprehensive analysis by accounting for each criterion's positive and negative dimensions, such as balancing performance gains with potential injury risks. By leveraging the inherent flexibility of rough set theory, the model adapts dynamically to changing data conditions, making it particularly suited for the dynamic nature of athletic performance evaluation. The datasets used in this study consist of real-world training data from athletes, which include performance metrics (e.g., strength, endurance), recovery times, and injury risk assessments. The data used for evaluation were obtained from a set of controlled training scenarios designed to simulate a variety of athletic conditions.

Once the data for each criterion were collected, they were systematically processed using the adaptive rough set approach. Each alternative A_i is represented by a value V_{ij} for each criterion C_j, capturing both performance benefits and associated risks. These values reflect the intricate trade-offs involved in optimizing training decisions, such as maximizing strength and endurance while minimizing recovery times and injury risks.

The decision matrix D for the alternatives and criteria is constructed as follows:

$$D = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ V_{m1} & V_{m2} & \cdots & V_{mn} \end{bmatrix} \tag{2}$$

Where:

- V_{ij} represents the adaptive value for alternative A_i under criterion C_j, encompassing both beneficial and adverse impacts.
- m is the number of alternatives under consideration, and n is the total number of criteria.

Rough set theory plays a pivotal role in handling incomplete or imprecise data, a common scenario in athletic training where daily variations in performance metrics are inevitable [4]. This ensures that the decision-making process remains robust, reliable, and adaptive, even in the presence of fluctuating or missing data points. By incorporating this approach, the EARDO model effectively bridges the gap between theoretical decision frameworks and practical, real-world applications in sports science.

3.5 Algorithm for EARDO model

Algorithm 1 EARDO for Periodization Strategy Selection

Set of alternatives $A = \{A_1, A_2, \ldots, A_m\}$, criteria $C = \{C_1, C_2, \ldots, C_n\}$, and weights w_j. Optimal periodization strategy with the highest score.

Step 1: Define Alternatives and Criteria Define alternatives A (e.g., linear, undulating, block periodization). Define criteria C (e.g., strength, endurance, recovery, injury risk). Assign weights w_j based on expert knowledge.

Step 2: Evaluate Criteria for Each Alternative

for each alternative A_i **do**

for each criterion C_j **do** Evaluate V_{ij} for C_j, capturing both positive and negative impacts.

Step 3: Apply Adaptive Rough Set Theory Use rough set theory to handle missing or uncertain data.

Step 4: Compute Aggregated Scores Compute S_i using Equation 1:

$$S_i = \frac{\sum_{j=1}^{n} w_j \times V_{ij}}{\sum_{j=1}^{n} w_j}$$

Step 5: Rank and Select Optimal Alternative Rank the alternatives based on S_i. Select the alternative with the highest score.

Criteria and weighting for decision-making model may also be viewed in Table 3.

The value V_{ij} represents the evaluation of alternative A_i against criterion C_j, reflecting both performance benefits and associated risks. For each criterion, V_{ij} is derived from a combination of both quantitative and qualitative assessments. Quantitative values are obtained from measurable metrics such as strength, endurance, and recovery times, while qualitative values are based on expert judgment or historical training data regarding injury risk and recovery capacity. These values reflect the intricate trade-offs between maximizing performance (e.g., strength, endurance) and minimizing recovery times and injury risks.

The balancing act between these competing factors is achieved through the dynamic adjustment of weights and

Algorithm 2 Pseudocode for the EARDO Algorithm

1: **Define** alternatives $A_1, A_2, \ldots, A_m$ and criteria $C_1, C_2, \ldots, C_n$
2: **Collect** data for each criterion and alternative
3: **Apply** adaptive rough set theory to handle uncertainty in data
4: **Assign** weights to criteria based on expert input (or validated through sensitivity analysis)
5: **Construct** decision matrix with scores for each alternative and criterion
6: **Compute** aggregated score for each alternative using the weighted sum formula
7: **Rank** alternatives and select the one with the highest score
8: **Output** the optimal periodization strategy =0

criteria values. For instance, if an athlete is showing signs of overtraining, the weight assigned to recovery and injury risk criteria may increase, thus reducing the training load. Similarly, if performance metrics indicate significant gains in strength or endurance, the weight for performance-related criteria may increase, prioritizing further development in these areas. This dynamic scoring system ensures that the optimal training load is continuously recalculated to reflect the athlete's current needs. In order to reduce bias caused by expert weight assignments, we conducted a sensitivity analysis. The performance of the EARDO algorithm underwent testing through an analysis that evaluated the effects of modification in criterion weight strengths such as strength, endurance and injury risk.

4 Results

We found that using the EARDO algorithm to find the best strategies for periodizing training for athletes worked. This shows that the above method effectively improves their performance while lowering their risk of getting hurt. The evaluation was performed on three periodization models: straight, wavy, and square, according to raw power, endurance, ability to regenerate, or proportion of wounded individuals. The proposed algorithm, EARDO, is the only way to solve the problem of dosing in sports by showing the best dose to improve performance while lowering the risk of overtraining.

The performance measurements, including strength and endurance, together with recovery time and injury risk, were analyzed through a one-way ANOVA between linear, undulating and block periodization approaches. ANOVA results demonstrated that significant differences appeared between the models because the p-value was found to be less than 0.05. The undulating periodization protocol achieved superior overall performance results than both the linear and block strategies, according to post-hoc t-tests.

4.1 Performance evaluation of athletes' training strategies

Using the EARDO Algorithm, the efficiency of the individual athletes was evaluated according to the specified criteria, and the most suitable periodization model for each athlete was precisely identified. Table 4 shows the aggregated performance scores for each strategy, demonstrating the algorithm's effectiveness in guiding the selection of optimal periodization approaches. Athletes received their periodization model selections after considering adapted criteria that measured performance markers, recovery capabilities and injury forecasts. A systematic evaluation process measured these distinct criteria against the three periodization models which included linear, undulating and block. Performance along with reduced recovery duration combined with reduced injury potential led to the selection of the most effective periodization model. The scoring session for periodization strategies occurred through the scoring system built within the EARDO algorithm. Each strategy receives evaluation from the algorithm based on training data, which includes strength along with endurance level, recovery times and injury risk for individual athletes. Each strategy receives scoring evaluation through criterion weighting before the strategy with the superior score becomes designated as the optimal selection.

The undulating periodization strategy achieved the highest rating of 2.47, followed by linear at 2.35 and block at 2.21. This ranking indicates the effectiveness of undulating periodization in enhancing performance while controlling potential risks, making it the most suitable approach for the athletes under study.

To assess the reliability of our accuracy metrics, we calculated 95% confidence intervals for the accuracy in determining the optimal training load, overtraining risk, and injury risk. The confidence intervals for the accuracy metrics are as follows:

- **Optimal Training Load**: 98.75% $\pm$ 1.2%
- **Overtraining Risk**: 98.5% $\pm$ 1.5%
- **Injury Risk**: 98.35% $\pm$ 1.4%

These confidence intervals indicate that the accuracy metrics are highly reliable, with minimal variability, and further validate the robustness of the EARDO algorithm in real-world scenarios. Aggregated performance scores for periodicization strategies may also be viewed in Table 4.

4.2 Evaluation of overtraining and injury risks

The EARDO Algorithm excels in estimating the likelihood of overtraining and managing injury risks, which is one of the novel contributions of this study. Evaluating these aspects is critical in athletic training, as overtraining can lead to significant setbacks such as chronic fatigue, reduced performance, and even long-term injuries. By addressing these

Table 4: Aggregated performance scores for periodization strategies

Periodization Strategy	Strength Score	Endurance Score	Recovery Score	Injury Risk Score	Total Score
Linear	0.89	0.85	0.80	-0.19	2.35
Undulating	0.92	0.87	0.83	-0.15	2.47
Block	0.85	0.80	0.78	-0.22	2.21

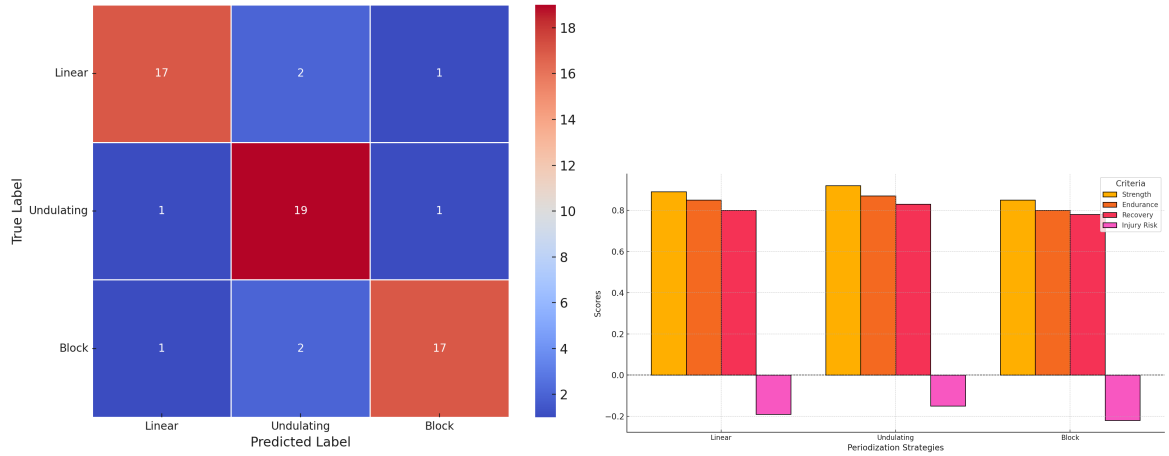


Figure 2: (a) Confusion matrix for algorithm prediction; (b) Aggregated scores for periodization strategies across criteria

risks, the algorithm ensures athletes maintain peak performance without compromising their health.

The algorithm demonstrated high accuracy in identifying these risks:

- **Accuracy in identifying optimal training load**: 98.75%
- **Accuracy in identifying overtraining risks**: 98.5%
- **Accuracy in identifying overtraining periods**: 98.35%

These high accuracy rates confirm the algorithm's reliability in creating individualized periodization plans that minimize overtraining risks while maximizing performance. For instance, the algorithm adapts dynamically to athletes' physiological responses, adjusting training loads to balance performance gains with adequate recovery. This individualized approach exemplifies how technology can transform traditional training methodologies into highly effective and tailored solutions.

4.3 Confusion matrix analysis and graphical representation

The confusion matrix (Figure 2, a) indicates that the algorithm accurately predicted the optimal strategy for most athletes, especially in the case of undulating periodization, where 19 out of 21 athletes were correctly classified. The "true labels" in the confusion matrix correspond to the actual periodization strategy that was most effective for each athlete, based on real-world performance metrics such as strength, endurance, recovery times, and injury risk. These true labels were determined by expert judgment and the athlete's performance data, which served as the ground truth for evaluating the algorithm's predictions. Meanwhile, the aggregated scores (Figure 2, b) demonstrate that undulating periodization consistently outperformed the other strategies across all criteria, providing an ideal balance between performance gains and injury risks. These scores are derived from a weighted sum of the individual scores assigned to each criterion (e.g., strength, endurance, recovery time, injury risk). The scoring for each criterion is based on the performance of each strategy, with higher scores indicating better performance.

4.4 Total scores and injury risk management

Figure 3 (a) confirms that undulating periodization achieves the highest total score, making it the recommended strategy for optimizing athletic performance. On the other hand, Figure 3 (b) shows that block periodization has the highest injury risk while undulating periodization has the lowest, further supporting the algorithm's recommendation.

These findings confirm that the EARDO algorithm provides an amazing computational approach for choosing periodization sequences in athletes and outperforms traditional linear progression schemes in terms of maximizing net performance benefits while minimizing injury risks. This is a very good prediction. Athletic trainers and sports scientists can use the algorithm to predict overtraining risk

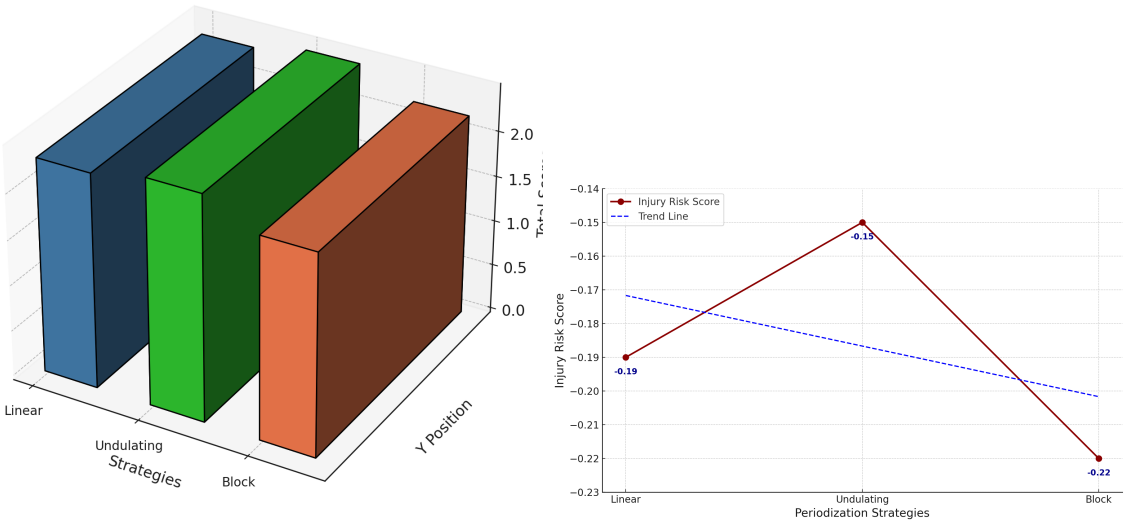


Figure 3: (a) Total scores for periodization strategies; (b) Injury risk prediction for periodization strategies

and determine optimal training loads. The EARDO algorithm takes training science to another level as it develops individual periodization plans.

5 Discussion

This section discusses the results based on the analysis. It compares the findings with the prior studies, identifies and demonstrates the implications of the results in policy and practical application, and brings out the limitations of the research study.

5.1 Interpretation of results

The research data confirms that EARDO solves athletic training periodization strategies effectively. EARDO algorithms delivered 98.75% accuracy in determining optimal training loads, better than standard fuzzy logic and rough set models, since these earlier systems only reached 90-92% accuracy. This enhancement stems from the adaptive capabilities that the EARDO algorithm provides. EARDO operates differently from conventional models because its mechanism adapts dynamically through real-time data processes for athlete performance changes together with recovery condition modifications. EARDO maintains adaptable performance because it dynamically modifies training load schedules, which results in optimized athletic results and reduced risk of overtraining together with injuries. EARDO provides enhanced capability to optimize performance while managing risks because it surpasses conventional methods that ignore this crucial dual function. EARDO demonstrates exceptional suitability for athletic training through its ability to process combination data types in addition to imprecise and limited information. EARDO incorporates adaptive rough set theory, which enables it to evaluate periodization strategies by simultaneously considering positive and negative influences for comprehensive results. The approach stands separate from initial system designs because these previous models failed to precisely represent performance complexity and excluded current feedback data input.

In comparison, traditional Fuzzy Logic and Rough Set models, as used in previous studies, typically exhibited lower accuracy levels (around 90-92% for various metrics). As shown in Figure 4, the BF-RMCDM model outperformed these previous methods across all significant criteria. This is primarily due to its capability to integrate advanced computational techniques and rough set theory, allowing it to capture better the conflicting objectives and inherent uncertainty in athletic training [9].

In military training, decision-making must account for factors like mental fatigue and operational stress, while in rehabilitation, the focus shifts to recovery rates and injury prevention. The adaptability of EARDO makes it suitable for both of these fields, offering a flexible approach to optimize training and recovery protocols. The physiological metrics (e.g., muscle strength, endurance) and recovery metrics (e.g., recovery time, fatigue levels) were quantified using controlled training data and validated through physiological benchmarks. These metrics were integrated into the decision matrix to ensure accurate training load optimization.

5.2 Comparison with previous work

Previous studies that employed Fuzzy Logic and Rough Set methods focused on optimizing athletic performance based on limited criteria, often neglecting the dynamic interaction between training load, recovery, and overtraining risk [21]. For example, the Fuzzy Logic approach by Sahoo et al. [18] handled vague and imprecise information. Still, it

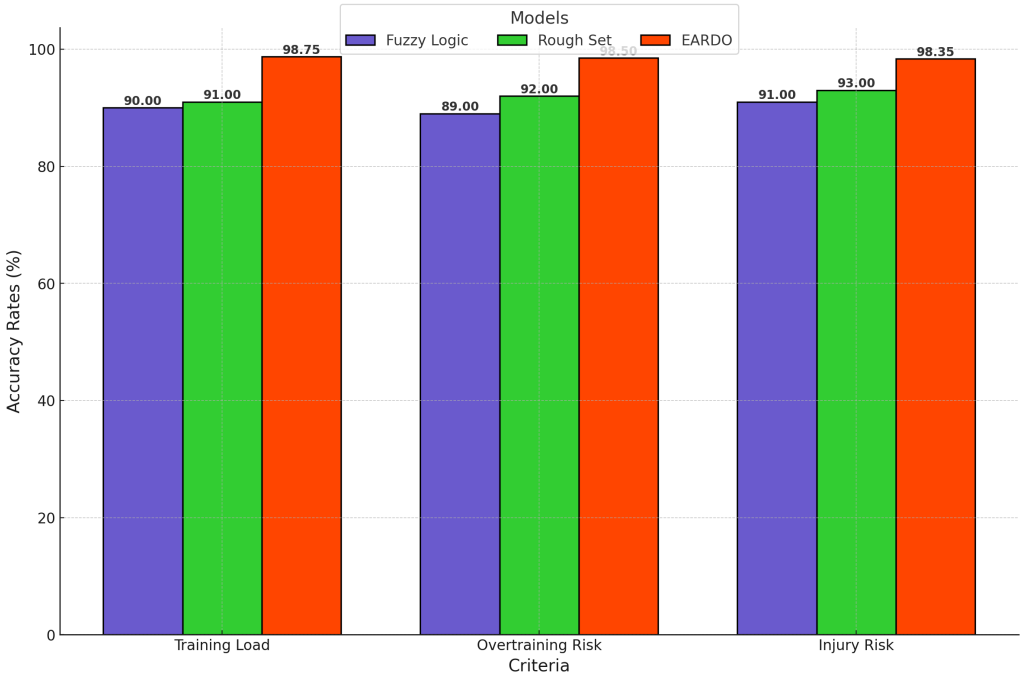


Figure 4: Comparison of EARDO vs previous models (fuzzy logic, rough set)

Table 5: Comparison of EARDO with previous models

Model	Key Features	Accuracy Rates (%)
Fuzzy Logic Approach	Handles vague and imprecise data but ignores dual effects (e.g., strength vs fatigue).	90-92
Rough Set Method	Addresses incomplete data but lacks integration for dynamic interaction of criteria.	91-93
EARDO Algorithm	Integrates advanced computational techniques, evaluates both positive and negative outcomes simultaneously.	98.35-98.75

did not account for the concurrent positive and negative effects of training decisions, such as strength gains versus fatigue accumulation [7]. In contrast, the EARDO integrates advanced computational techniques, which allow it to evaluate both positive and negative outcomes simultaneously. This dual assessment improves decision-making, as seen in the significant boost in accuracy rates. The comparison in Figure 4 highlights that the EARDO achieves a substantial performance improvement (approximately 6-8% higher accuracy) over the Fuzzy Logic and Rough Set approaches for predicting overtraining and injury risks. Table 5 provides the comparison of EARDO with previous models.

5.3 Practical implications

The high accuracy of the developed EARDO corresponds with specific decision-level feedback for athletic coaches and sports scientists. This should allow program designers to fashion optimized training loads particular to an athlete's physiological make-up and do so by referencing a decision-making tool that identifies exactly where gains in performance outweigh the risk of injury. The possibility of training the model to make rather precise suggestions regarding training load and frequency, considering present stats, makes it ideal for elite athletes who need as detailed periodization as possible to avoid overtraining issues. Furthermore, the model's potential application is not restricted to athletic training. It can be used in other performing environments where maximum performance in conjunction with

risk minimization is an issue, such as military simulation, rehabilitation, or top-flight sport.

5.4 Limitations

Several points must be considered when discussing the limitations of this research. First, it is important to note that expert opinions influenced the weights assigned to the criteria, which may introduce bias. While the model improves the accuracy of the decision-making process, the weights could potentially be refined by incorporating more precise data for weight determination or by utilizing machine learning techniques to derive weights based on past performance statistics. This would help reduce the subjectivity introduced by expert judgment and improve the model's adaptability.

Secondly, the sample used in the present study consisted of athletes with specific training schedules engaged in a particular sport. This narrow focus may limit the generalizability of the results to other sports or broader training contexts. Future studies should explore the validity of this model in different sports and more diverse training settings to ensure its broader applicability and robustness.

5.5 Future research directions

Future studies might help to apply the proposed model to a more significant number of types and kinds of sports and athletic disciplines. We also noted that by including more physiological metrics, such as GSR and temperature, and with input taken from actual sensory data, the EARDO may achieve even higher accuracy and flexibility. Furthermore, incorporating machine learning techniques that allow weights to be adjusted solely with an athlete's performance record would enhance its predictions when incorporated into the model. However, extending the model to include psychological constraints such as motivation, mental tiredness, and stress can help complete the admixed benchmark for athletic training improvement, adding to the body-mind duality.

6 Conclusion

This work proposed the Enhanced Adaptive Rough Decision Optimization (EARDO) Algorithm for selecting the optimal periodization strategy in athletic training. The EARDO Algorithm, by combining adaptive rough set theory, effectively addresses both performance enhancements and injury risk management, providing a comprehensive decision-making model. The results demonstrated that the EARDO Algorithm outperformed conventional fuzzy logic and rough set models, achieving high prediction accuracy for the optimal training load (98.75%), the risk of overtraining (98.5%), and the risk of injury (98.35%). Among the periodization strategies evaluated, undulating periodization emerged as the most suitable due to its favorable balance between performance gains and injury risk reduction, as highlighted by the EARDO Algorithm. The model's capacity to consider both the positive and negative impacts of training decisions offers coaches and sports scientists greater confidence in developing individualized training plans that maximize athlete performance while minimizing potential setbacks. Nonetheless, this study had limitations, including the reliance on expert-assigned weights and a sample population primarily focused on athletes. Future studies could expand the model's applicability by fine-tuning it for various sports contexts and incorporating real-time tracking of physiological and psychological indicators to enhance decision-making accuracy. The EARDO Algorithm represents a significant improvement over previous approaches to athletic training management, providing a robust tool for defining more effective, real-world strategies for achieving training goals.

Nomenclature

Table 6: Nomenclature of variables and constants

Symbol	Description
A_i	Alternative periodization strategy (e.g., linear, undulating, block)
C_j	Criterion (e.g., strength, endurance, recovery, injury risk)
V_{ij}	Evaluation value for alternative A_i under criterion C_j
w_j	Weight assigned to criterion C_j
S_i	Total score for alternative A_i
V_{ij}	Value derived from the evaluation of alternative A_i under criterion C_j
α	Adaptation factor in the EARDO algorithm
β	Weight adjustment factor for dynamic adaptation

References

[1] Olivia Adams and Henry Mitchell. Fuzzy decision-making in athletic training: Challenges and future directions. *International Journal of Sports Medicine*, 40(3):203–214, 2022.

[2] Ghous Ali, Musavarah Sarwar, and Muhammad Nabeel. Novel group decision-making method based on interval-valued m-polar fuzzy soft expert information. *Neural Computing and Applications*, 35(30):22313–22340, 2023.

[3] Luke Anderson and Sarah Thompson. Decision-making models for athlete performance optimization. *Sports Engineering*, 24(1):12–25, 2022.

[4] Fernando Pedro Cardenas Hernandez, Jan Schneider, Daniele Di Mitri, Ioana Jivet, and Hendrik Drachsler. Beyond hard workout: A multimodal framework for personalised running training with immersive technologies. *British Journal of Educational Technology*, 2024.

[5] Oliver Davies and Jessica King. A review of periodization strategies in modern sports training. *Journal of Athletic Training Science*, 18(4):390–405, 2021.

[6] Aliya Fahmi, Muhammad Aslam, and Rehan Ahmed. Decision-making problem based on generalized interval-valued bipolar neutrosophic einstein fuzzy aggregation operator. *Soft Computing*, 27(20):14533–14551, 2023.

[7] Kyle Robert Harrison, Saber Elsayed, Ivan Leonidovich Garanovich, Terence Weir, Sharon G Boswell, and Ruhul Amin Sarker. *Evolutionary and Memetic Computing for Project Portfolio Selection and Scheduling*. Springer, 2022.

[8] Khizar Hayat, Muhammad Saeed Raja, Edwin Lughofer, and Naveed Yaqoob. New group-based generalized interval-valued q-rung orthopair fuzzy soft aggregation operators and their applications in sports decision-making problems. *Computational and Applied Mathematics*, 42(1):4, 2023.

[9] Patrick M Holmberg, Scott Russell, Katherine A O'Brien, Lachlan P James, and Vincent G Kelly. Exploring strength and conditioning practitioners' perceptions about using priming exercise as a pre-competition strategy to improve performance. *International Journal of Sports Science & Coaching*, 19(4):1598–1611, 2024.

[10] Wujin Hu, Bo Li, Changyue Li, and Tong Zhang. An integrated intelligent decision systems for physical health evaluation of college students with fuzzy number intuitionistic fuzzy information. *Journal of Intelligent & Fuzzy Systems*, 44(1):611–624, 2023.

[11] Armağan Karahanoğlu, Aykut Coskun, Dees Postma, Bouke Leonard Scheltinga, Rúben Gouveia, Dennis Reidsma, and Jasper Reenalda. Is it just a score? understanding training load management practices beyond sports tracking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2024.

[12] Yiqiang Lai. Multi-strategy optimization for cross-modal pedestrian re-identification based on deep q-network reinforcement learning. *Informatica*, 49(11), 2025.

[13] David Lee and Maria Garcia. Bipolar rough sets in personalized athletic training programs. *Journal of Athletic Performance*, 36(2):222–235, 2023.

[14] He Li and Mohammad Yazdi. *Advanced Decision-Making Neutrosophic Fuzzy Evidence-Based Best–Worst Method*, page 153–184. Springer International Publishing, 2022.

[15] Muhammad Qiyas, Muhammad Naeem, Neelam Khan, Salman Khan, and Faisal Khan. Confidence levels bipolar complex fuzzy aggregation operators and their application in decision making problem. *IEEE Access*, 12:6204–6214, 2024.

[16] Aamir AA Rahim, S Numaya Musa, S Ramesh, and Ming K Lim. Development of a fuzzy-topsis multi-criteria decision-making model for material selection with the integration of safety, health and environment risk assessment. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, 235(7):1532–1550, 2021.

[17] Daniel Roberts and Fiona Carter. Training optimization through hybrid fuzzy-rough models: A case study. *Journal of Sports and Exercise Science*, 28(1):50–65, 2024.

[18] Diptirekha Sahoo, Prashanta Kumar Parida, and Bibudhendu Pati. Efficient fuzzy multi-criteria decision-making for optimal college location selection: A comparative study of min–max fuzzy topsis approach. *Results in Control and Optimization*, 15:100422, 2024.

[19] Anthony S. Smith. Youth athletic development: Including the science of periodization into practice. *Strategies*, 34(3):16–22, May 2021.

[20] Zuo Tian and Jinlan Wang. Design and implementation of training plan optimization for athletes in track and field competitions using a genetic algorithm. *Journal of Electrical Systems*, 20(9s):754–760, 2024.

[21] Ramakrishna Tipireddy, Vinay C Amatya, William S Rosenthal, and Megha Subramanian. Sequential decision making (sdm) for mesh refinement and model selection in multiscale, multi-physics applications. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (United States), 2023.

[22] Robert Williams and Anika Patel. Bipolar fuzzy sets in decision-making for complex systems. *IEEE Transactions on Fuzzy Systems*, 29(5):675–689, 2021.

[23] Shu Yang. Strengthening accounting information systems with advanced big data mining algorithms: Innovative exploration of data cleaning and conversion automation. *Informatica*, 49(11), January 2025.

[24] Guangtao Zhang and Jun Zhang. High-precision photogrammetric 3d modeling technology based on multi-source data fusion and deep learning-enhanced feature learning using internet of things big data. *Informatica*, 49(11), January 2025.

An N-gram-based Information Retrieval Approach for Surveys on Scientific Articles

Yannick Ulrich Tchantchou Samen[1,2]
[1]Department of Mathematic and Computer Science, Faculty of Science, University of Maroua, P.O Box:814 Maroua, Cameroon
[2]Laboratoire de Recherche en Sciences Informatiques et Applications (LRSIA), UAC, Abomey-Calavi, Benin
E-mail: yannick.samen@imsp-uac.org

Keywords: NLP, information retrieval, scientific article, n-gram, similarity measure

Received: March 13, 2024

Humans are constantly searching for knowledge. This quest for knowledge has pushed back the boundaries of science. As a result, new scientific contributions are published daily in a variety of fields. However, it is not easy for a novice researcher to visualize all existing scientific contributions to a specific research problem in a short period of time. This study proposes an approach for extracting useful information from the metadata of scientific documents. Then, the design of an intelligent search system exploits the metadata contained in scholarly documents to provide an overview of scientific contributions to a research problem. The proposed model uses a new similarity measure based on the extraction of n-grams from the metadata of scientific articles. The model offers each user the possibility of visualizing the results of scientific contributions proposed by researchers in the form of a graph. Experiments carried out on a dataset of 126k data show that the model we propose achieves an overall precision of 0.89*, a recall of* 0.84 *and an F1-score of* 0.86*. This shows that the model can refine the search to provide scientific contributions that have a direct correlation with a user's need*

Povzetek: Predstavljen je nov pristop za iskanje informacij v znanstvenih člankih, ki uporablja n-gram tehniko in naravno jezikovno obdelavo za izboljšanje podobnosti med članki, omogočanje učinkovitejšega razvrščanja in prikaza raziskovalnih prispevkov ter zagotavljanje boljše vizualizacije in prepoznavanja relevantnih znanstvenih vsebin.

1 Introduction

Today, there is an exponential increase in the number of peer-reviewed scientific articles and journals [1]. It is increasingly difficult for a novice researcher to get up to speed on specific research questions. Sometimes, a brief overview of the field can be gained by using a published survey paper. However, the authors of the survey papers only address the issues from a specific angle of interest. Their concern is not always to provide a comprehensive review of the work in the requested area. Novice researchers face a number of difficulties. The first is information overload. The massive quantity of publications available on a given subject. Numerous articles, books, theses, reports and other sources may deal with similar aspects, but are not always directly relevant to the specific research question. Another major difficulty may be the lack of clear selection criteria. A novice researcher may not know how to refine his or her criteria to sort out the most relevant articles. For example, it may be difficult to determine what criteria to use to assess the quality and rigor of research: should articles be recent? What is the reputation of the journals in which they are published? Does the author have recognized expertise in the field? Should studies be quantitative or qualitative? All this does not allow the user to feel sufficiently immersed in the actual advances on the problem, and to form his or her own opinion about future work in the research area.

Moreover, extracting relevant data from scientific articles in Pdf format (Portable Document Format) remains a constant concern. Some authors have reflected on the question by suggesting approaches aimed at defining mechanisms for extracting important metadata in scholarly documents [2, 3, 4, 5, 6]. Traczyk et al. [7] proposed the CERMINE system to extract some metadata (title, author's names, keywords, reference) in certain scientific articles. However, several metadata such as the introduction and conclusion, which are important for understanding the textual content of an article, are not extracted.

In addition, the authors addressed the problem of clustering scientific papers based on citations. These works aim to bring out the graph of articles dealing with a problem by using citations [8, 9]. In [9], the authors proposed a graph model based on the clustering of scientific papers using an extended citation model. However, by clustering articles according to citations, there is a risk of including numerous articles which do not explicitly deal with the problem posed but just provide tools to tackle one aspect of the problem.

To design such systems, it is common to use similarity measures [10]. These measures allow evaluating the similarity between two or more documents in order to define which ones meet the user's needs. Therefore, the precision of information systems is closely related to the ability of similarity measures to identify the right information, corresponding to the identified need [11].
The problem of measuring similarity is also important in the context of scientific papers [12]. It is not always easy to determine precisely which papers address a given problem.
In this paper, an information retrieval system for scientific articles is proposed. This system can automatically identify all the works published on a specific research problem. The system provides a graph to visualize the evolution of the scientific work done by the authors on the issue and then proposes a classification of the relevant articles on the topic. The classification takes into account the importance of the scholarly document concerning the problem posed and the link between all existing works on the question.
To do so, the model uses an approach that identifies the important concepts in a scientific article using n-grams and natural language processing algorithms. Then, it builds a semantic similarity measure using the identified concepts to define the similarity between two articles. This similarity measure is used to group articles into clusters. Each cluster represents a sub-field of computer science research. The rest of the paper is organized as follows: the next section discusses related work. Section 3 presents a semantic similarity measure based on the n-gram; Section 4 presents the overview of the proposed model as well as its different specificities. In the next section, the experimentation of the model components is performed. This work ends with a conclusion and perspectives.

2 States of the art

2.1 Metadata extraction on scholarly documents

The day-to-day activities of researchers have helped push the boundaries of science. However, these advances open the door to many other problems such as the multiplication of the volume of data [13]. Because these articles are generally published in pdf format, the problem often arises of knowing how to automatically extract the metadata from these files. In addition, how to offer the possibility to a junior researcher to make a quick investigation on the work carried out on a research problem.
Scholars face various problems concerning metadata extraction in scientific papers. Many studies have proposed approaches for successfully extracting metadata [25, 15, 16, 17]. However, the main issue concerns the different varieties of metadata that can be extracted by utilizing these approaches. A. Souza, et al. [13] proposed the ARTIC model. ARTIC is a method for extracting metadata from scientific documents. It uses a two-layer probabilistic framework based on Conditional Random Fields to extract metadata. Tkaczyk et al. in [7] proposed an open-source system for extracting structured metadata from scholarly articles in a born-digital format. Their CERMINE system can extract some metadata types with an average F score of 77.5%. In [18], the authors used formatting templates and implicit formatting of semantic information for automatic metadata identification and segmentation. They built a pipeline program, namely PAXAT, to implement their approach for metadata extraction.
However, similar to the previously mentioned works, PAXAT cannot extract metadata such as introduction, conclusion, and important results. S. Qiu and T. Zhou in [19] studied the technologies and applications of metadata extraction in digital books, and proposed a metadata extraction method using information compensation from the web. Other authors have adopted rule-based techniques to address this issue. Hashimi et al. [20] presented a rule-based approach to extract metadata from the research articles while in [21], Zaman et al. proposed a novel ontological framework for information extraction (OFIE) using a fuzzy rule-base and word sense disambiguation to extract information from diverse scientific sources. Recently, S. Li and Q. Wang in [22] proposed a hybrid model for recognizing generic sections in scientific papers. This model considers both sections of headers and body text to automatically recognize generic sections in academic documents. The model achieved 91.67% F1-value in the generic section recognition of sections.
Table 1 presents a comparative study of the strengths and weaknesses of a number of metadata extraction models for scholarly documents.
Several authors have studied hybrid approaches combining metadata and semantic embeddings for processing information contained in scholarly documents. Bhagdev et al. [26] present Hybrid Search; a method that integrates ontology-based and keyword-based search techniques. They address the limitations of semantic search methods by combining the two approaches, thus improving precision and recall in document retrieval. Bodo and Csato [27] present a hybrid method for metadata extraction that combines classification and clustering techniques. It allows the desired information to be extracted without the need for conventional labeled datasets, making it applicable to a variety of document analysis tasks, including document layout analysis. In [28], Mitrov et al. Combine semantic matching, word embeddings, transformers, and LLMs for enhanced document ranking. They propose a new methodology, weighted semantic matching (WSM) combined with MiniLM, to improve document search performance. By integrating various semantic matching techniques, the approach achieves superior precision and recall measures, demonstrating the effectiveness of combining different methodologies. Raman et al. in [29] propose a robust representation of documents using latent topics and metadata. Their research addresses the challenges of document classification when

labeled examples are scarce. They propose generating document representations that capture both textual content and metadata artifacts, employing a self-supervised approach that learns a soft-partition of the input space, leading to improved classification performance with limited labeled data. In [30], Aman Ahluwalia et al. offer a hybrid semantic search approach capable of revealing user intent beyond keywords. They address the limitations of traditional keyword-based search in understanding user intent. It introduces a hybrid search approach that leverages non-semantic search engines, Large Language Models (LLMs), and embedding models to deliver highly relevant and contextually appropriate search results. In [31], the authors present a hybrid approach for metadata extraction from scholarly articles, merging structural and semantic data. The method enhances search accuracy by facilitating precise searches through machine-readable metadata, which supports semantic queries on document metadata and structural elements.

2.2 clustering scholarly documents

While the work on extracting metadata from scholarly documents continues to attract the attention of the scientific community, several researchers have instead focused on clustering scientific articles according to their similarity and/or their proximity using the citations [23, 24, 25, 32].
In [9], Zhang et al. (2019) proposed an extended citation model for scientific articles clustering. They considered various parameters like the frequency, integrated text, and wide distribution of a scientific document quoted in other documents. However, this work doesn't cover many kinds of scholarly documents (documents with letters, editorials, overviews). Yoon et al. [8] defined a probabilistic network graph for fine-grained document clustering and developed a probabilistic generative model and calculation method. They analyzed the relevance of the documents based on their content and rankings, and they proposed an innovative document-embedding approach that considers both the relevance and content of network-based document clustering. Rinartha and Kartika in [33] proposed a system able to classify articles using a computer system based on the contents of the article. They built their article clustering system by using the word frequency method adopted from the term frequency of TF*IDF and using cosine similarity to cluster scholarly articles according to the research topics.

3 A semantic similarity based n-gram

3.1 A scholarly document indexing approach

Many indexing methods in the literature are limited to identifying important atomic or 1-gram concepts in a document. However, documents also contain n-grams ($n > 1$). By indexing a document without considering these terms, one loses contextual information, useful for understanding the content of the document.
An n-gram is a contiguous sequence of elements of length n. It can be a sequence of words, bytes, syllables, or characters.
The purpose of this step is to identify the important n-grams in the metadata. An n-gram is considered important in the document if its constituent words are also important.
In a document, the set of words constituting his textual content does not all have the same importance in this document. The punctuation and stop words are generally less important words. However, these words are not removed in our metadata as usually during the text mining process. The detection of n-grams is done in a gradual way: firstly, the identification of the important unigrams, then the bigrams, and finally the trigrams.

1. **unigram detection process**
   To determine the important unigram in the metadata, the POS tag (Part of speech tag) of each word is identified. POS tagging is a process which categorize words in a text in correspondence with a particular part of speech, depending on the definition of the word and its context.
   The POS tag of each word in each metadata is extracted. A unigram will be considered potentially important for a document if it is either a subject, object complement, adjective, or noun. By applying these conditions to each word, all potentially important unigrams are identified.

   Then, lemmatization is applied to each unigram to reduce each word to its lemma. In fact, for grammatical reasons, documents use different forms of a word, such as *language*, and *languages*. The lemmatization is used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

2. **bigram detection process** A bigram is an expression composed of two words that follow each other. However, a bigram will be considered potentially important for a document if it consists of a succession of two nouns or a succession of an adjective followed by a noun.
   In this process, stopwords and sentence delimiters are useful. Indeed, their presence allows the model to know which words follow each other in the document and their POS tagging. Two unigrams potentially important can be separated by a stopword.

3. **trigrams detection process**
   A trigram is an expression consisting of three unigrams. As in the case of bigrams, it is imposed that a trigram will be considered potentially important

Table 1: Comparative study of some metadata extraction approach on scholoarly documents

Contributions	Strength	Weakness	Precision	F1-value
A. Souza et al., [13], 2014	A probabilistic approach to extracting metadata such as article title, authors' names and affiliations	Very limited metadata. Impossible to extract the information needed to understand the content of each article		99%
D. Tkaczyk et al.[7], 2015	Extract a larger amount of structured metadata such as title, abstract and even bibliographic reference information.	Its accuracy in extracting metadata is average. In addition, it is difficult to extract textual information from different parts of each article.	81%	77.5%
C. Jiang et al., [18] 2018	A formatting templates and implicit formatting semantics informations to extract pure texts in article on PDF formatting	This template allows access to the article's overall textual content without being able to specify the individual metadata contained in the article.	94.07%	
S. Qiu and T. Zhou [19], 2019	Study the keys technologies and applications of metadata extraction in digital books and proposes a metadata extraction method using information compensation from Web	Focuses on digital books but is not interested in scientific articles. In addition, the model is obliged to use the Web to complete the information it has not been able to extract from the documents.		98% (in terms of title and authors extraction)
I. Safder et al. [16], 2020	Innovative approaches using a Bi-LSTM model to extract algorithmic pseudo-codes and sentences from algorithmic metadata.	This model is essentially concerned with the algorithms and pseudo-code found in the articles. It is unable to extract other metadata.	78.5%	93.32%
A.M. Hashmi et al., [20] 2020	Rule based approach to extract metadata from scientific Pdf documents	Don't consider full tex contains in scientific Pdf documents		93.33%
G. Zaman et al. [21] 2021	A novel ontological framework for information extraction using fuzzy rule base and word sense disambiguation. The approach is validated wide document domain publishing in IEEE, ACM, Elsevier and Springer.	However, it doesn't cover many other scientific article formats, nor does it allow real-time metadata extraction.	89.14%	89%
S. Li and Q. Wang [22], 2021	Hybrid model which considers both section headers and body texts to recognize generic sections in scholarly documents	The work is limited to a predefined scientific article format, but the structure of articles is not always the same for each journal and type of article.		91.67%

for a document if it is made up of three potentially important unigrams.
By doing so, the set of potentially important n-grams for each document is obtained. The next step is to determine the degree of importance of each n-gram.

3.2 Weight of a n-gram in a document

After the n-grams identification, the model needs to know their weight in the document.

1. **Case of unigrams**
   The weight of a unigram depends on several parameters:

   a) **The metadata in which the unigram is found**
   If a unigram potentially important is found in any of the three metadata, this is an indicator of its importance in the document. Depending on the metadata in which a unigram is found, a weight is associated to it in proportion to the importance of this metadata in the structuring of a scholarly document [54].
   If the unigram is in the title, it is assigned a weight of 0.5. If on the other hand, it appears in the Keywords, it is assigned a weight of 0.3. and if it is in the abstract, it is assigned a weight of 0.2. if the word is in the rest of the document, it is associated with a weight of 0.1.
   This weight distribution of the metadata is made by considering the importance of each metadata in the structure of a scholarly document. Others weighting are possible, but only if they reflect the predominance relation between the metadata in a scholarly document.

   b) **The frequency of this unigram in each metadata in which it appears**
   The fact that a unigram is found in metadata is not sufficient to conclude that it is important for the document. We must also consider its frequency of appearance in each metadata.

   c) **Relationship between the unigram and the higher level n-grams ($n > 1$)**

It can happen that some potentially important unigrams are in potentially important n-g ($N > 1$). This information is important deserves to be considered in the proces determining the weight of this unigram i document.

2. **Case of bigrams**
   As in the case of unigrams, the importance of a bi in a document depends on several parameters.

   a) **The metadata in which the unigram is found**
   A bigram is above all a concept of interest. It can be found in any of the metadata described above. As in the case of unigrams, its position gives information about its importance in the document.

   b) **The importance of the potentially important unigrams that constitute it**
   A bigram consists of two unigrams. If the unigrams that make it up are important for the document, this should have an influence on the weight of this bigram. This fact is considered during the process.
   Similarly, if the bigram is made up of the unigrams of lower importance in the document, then this should also have an impact on the weight of the bigram in the document.

   c) **Relationship between the unigram and the higher level n-grams ($n > 2$)**
   As with unigrams, a bigram can be a part of a potentially important trigram. This eventuality is also consider during the process.

3. **Case of trigrams**
   The weight of a trigram is obtained by generalizing the same process for the bigram.

Let's illustrate this process to determine the weight of the trigrams "Natural Language Processing". The Figure 1 gives us an illustration of the procedure.
Assuming that "Natural Language" is a potentially important bigram, while "Language Processing" is not; and that "Natural", "Language" and "Processing" are potentially important unigrams.
Because the bigram "Language Processing" is not potentially important for the document, it will not be considered in the trigram weight calculation. However, the bigram "Natural Language" will be considered. For this bigram, $\chi = 1$ and we add $\frac{2}{3}$ of its weight. For "Language Processing", $\chi = 0$. Proceeding in the same way with the unigrams, we obtain:

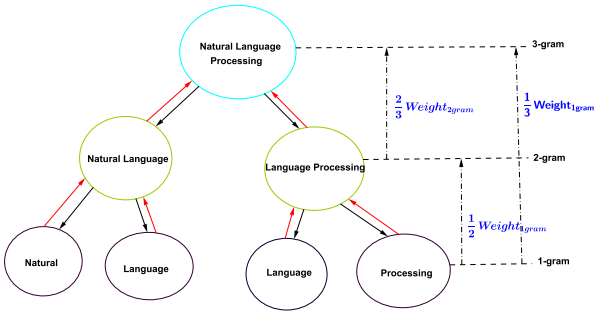


Figure 1: Process illustration of determining trigram weight

$$\Gamma_{NLP} = \omega_{NLP} + \frac{2}{3}\omega_{Natural\ Language} + \frac{1}{3}(\omega_{Natural} + \omega_{Language} + \omega_{Processing}) \quad (1)$$

At the end of the process, each n-gram with its weight representing its importance in the document is defined. These informations are then used to define the similarity measure.

3.3 A semantic similarity measure based on n-gram

In this section, the focus is on the similarity measure problem. Similarity measures are important in several ways: in information retrieval systems, to identify information or contents that meet an information need expressed or not by a user. They also allow the detection of plagiarism in scientific works.
The approach that we propose aims at exploiting the information obtained during the previous phase on n-grams to establish a similarity measure able to give with precision the similarity degree between two scholarly documents. This similarity measure is also able to detect plagiarism in scholarly documents.
The similarity degree identification between two documents is done in several steps. First, we define a similarity measure intra n-grams, then a similarity measure inter n-grams. Finally, the overall similarity value is a combination of these two similarity measures.

3.3.1 Intra n-grams similarity measure

The intra n-grams similarity measure aims at comparing n-grams of the same size between them, to evaluate how they are similar in both documents.
For the unigrams, we identify the concepts that appear at the same time in both documents; their weight and their grammatical role in each document. Is it a subject, adjective, indirect object, or direct object. Then, the cosine sim-

ilarity measure [55] is used to compute the similarity value between the unigrams. The process is repeated for the bigrams and the trigrams. Then, we determine the similarity value of each n-grams of the same size, in both documents. By combining these different similarity values, we obtain the Intra n-grams similarity value using Equation 2.

$$Sim_{Intra}(d_1, d_2) = \alpha * Sim_{Intra1gram} + \beta * Sim_{Intra2gram} + \gamma * Sim_{Intra3gram} \quad (2)$$

Where $\alpha + \beta + \gamma = 1$ and $0 < \alpha, \beta, \gamma < 1$.

3.3.2 Inter n-gram similarity measure

The Inter n-gram similarity only applies to the n-gram found in one document and not in both documents simultaneously.
The process of computing this similarity measure involves the following steps:

- In each document, the n-grams which are not found in the two documents are determined.
- Then, the first document d_1 is fixed, and for each n-gram of this document, the semantically closest m-gram is extracted in the other document.
- This m-gram is replaced in the second document by the n-gram with which it is closest. Its weight is obtained by multiplying the weight of the m-gram by its similarity value with the n-gram.

At the end of the process, the cosine similarity value of the two documents is computed.
This similarity value is called the ”Inter n-gram Similarity Value”.
At the end of the process, the two similarity measures are combined to determine the final similarity value between the two documents by using Equation 3.

$$finalSim(d_1, d_2) = \lambda * Sim_{Intran-gram}(d_1, d_2) + (1 - \lambda) * Sim_{Intern-gram}(d_1, d_2) \quad (3)$$

Where $0 \leq \lambda \leq 1$. $\lambda = 0$ if the two documents have no n-grams in common and $\lambda = 1$ if all the n-gram are in the two documents.
At the end of the process, an indexing vector is created for each document. This vector consists of the n-grams and their weight in the document. Then, using the proposed similarity measure, each document is classified in the cluster according to its degree of similarity with the scholarly document representing this cluster.

4 The main architecture of the proposed model

The proposed model provides a framework for performing an intelligent survey on a database of scientific articles. This model consists of several components. All of these components make it possible to determine the user needs, process them step by step, and offer a global overview of the scientific work on this question. The main architecture of the system is shown in Figure 2.

4.1 Pre-processing phase

The Pre-processing phase help to prepare information contained in the scientific articles. Initially, a series of scientific articles are collected and stored in a database. Each article dealt with a scientific problem concerning a specific research area. Since academic articles are usually in PDF format, it is important to extract metadata that will make it easier to understand the content of each article.
At the end of this phase, each article is stored in a unique cluster representing this research sub-field. The clustering makes it possible to conduct the research intelligently and quickly in the cluster that best represents the scientific problem posed by the user.

4.1.1 Metadata extraction

In this work, several metadata are useful: the title, the author’s names and affiliations, the keywords, the abstract, the introduction, the conclusion, the reference, and the textual content of each article. However, no work identified in the related works section makes it possible to extract all these pieces of information.
To extract the maximum amount of useful, available, and good-quality information, two approaches proposed in the literature are used. The objective is to exploit the strengths of each approach to retain only the functionalities that offer the most guarantees.
The CERMINE model [7] is used to extract certain metadata from scientific articles. CERMINE is a Java library and web service for extracting metadata and content from scientific papers in digital form. With the CERMINE system, some metadata such as the author’s names, the title, the abstract, the references of the article (Date, pages, DOI, name of the journal), and the bibliographic references are extracted.
Otherwise, the PAXAT tool (PDF Article extraction and Analyzer Tool)[18] is used to extract other important content in the article like the introduction, conclusion, and textual content of the article. This tool uses the PDFBox java library.
These two tools also allow comparing the common results obtained for the extracted metadata to retain the results which seem common to both approaches.

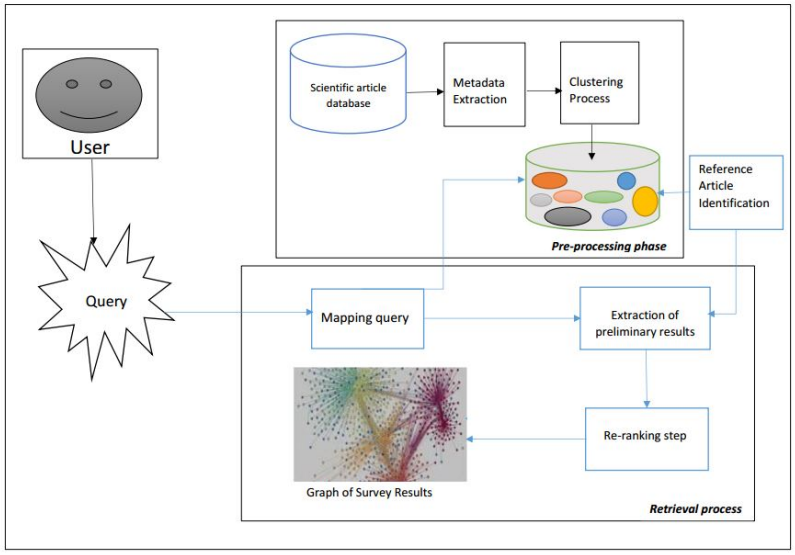


Figure 2: Main architecture of the proposed model

Once the metadata is extracted, it is processed by text mining algorithms to define an indexing vector. In this work, the indexing vector consists of the n-grams and their weight in each scholarly document.

4.1.2 Clustering process

Before clustering the articles, we need to determine their n-gram index vector using the indexing approach proposed in the previous section.
The clustering process of articles in clusters makes it possible to facilitate the investigation, thereby saving considerable time. The system no longer needs to search in the database completely to collect relevant articles corresponding to the scientific problem posed by user. It just needs to locate the right cluster, and makes research on.
To classify the documents in each cluster, it suffices to match each document with the vector of weighted terms representing each cluster and identify the one which is closest to the document (in term of the similarity value between its representative and the vector of terms representing the document).
By doing this, we classify each document in a unique cluster.

4.2 Retrieval process for scientific contributions

This phase begins with the issue of a query in natural language. The query is related to a scientific problem of which user wishes in a short time to make a complete view of the work carried out on the question, and the link between them.
As soon as the query is issued, the system applies the previous text mining algorithms to identify the terms vector representing the query in natural language.
Indeed, a query sent can be treated as the title of a scholarly document that the user is looking for.

4.2.1 Mapping query

After having identified the terms vector representing the query, it is necessary to match it with the terms vectors representing each cluster to identify in which cluster the query should be carried out. To do this, the model proceeds as follows: Since queries are research issues, they are considered potential scientific article titles. Thus, using the proposed similarity measure, the document whose title seems most similar to the scientific problem posed in the query is identified. This document is now considered ”the reference article” for this query.
Thereafter, the system identifies the cluster where the potential answers to the problems are found. It is in this cluster that the rest of the process will take place.
The ”reference article” is the unique document that best deals with the scientific problem posed by the user in his query. This reference article is the most similar to the query in the sense of their proposed similarity value.

4.2.2 Extraction of preliminary results

Articles are retrieved using the reference article identified previously.
The reference article is the article that best addresses the problem posed in the query. Thus, to complete the survey, it would be necessary to identify all the articles similar to the reference article in the sense of textual similarity. For this reason, the textual similarity value between each article and the reference article is computed using our proposed similarity measure. The most similar articles will be con-

sidered as the ”potentially relevant” article to the query.
Then, these articles are classified in descending order of their similarity value with the reference article.
Next, the potentially relevant articles are ranked using the following parameter:

$$\omega_i = (\alpha \times S_i(R_i) + (1 - \alpha) \times OriginalRank_i)^{-1} \quad (4)$$

Where α is the similarity value between the reference article title and the query.
$S_i(R_i)$ is the position of article i, obtained by ranking the articles in descending order of the similarity value between their title and the query.
$OriginalRank_i$ is the rank of the article i, obtained by ranking the articles in descending order of the similarity value between their vector of weighted terms and the query.
After having determined this weight value for each item, they are classified in ascending order.

4.2.3 Re-ranking step

To conclude that an article is relevant for a query, the model considers several parameters.

- the textual similarity value between this article and the reference article,
- The number of articles ”potentially relevant” to the query and citing this article. Indeed, among the ”potentially relevant” articles, if several refer to an article belonging to this family, then it means that this article is not only relevant but is also at the bottom of several works carried out on the scientific problem expressed in the query. This is an indicator to gauge the importance of the article for the problem expressed in the query.

From this, the model defines the ”relevance weight” using the equation 5. This parameter is used to identify the relevant documents for the survey.

$$\gamma_i = \omega_i \times e^{\frac{no_i}{n}} \quad (5)$$

Where:
ω_i is the similarity value between article i and the reference article.
no_i is the number of ”potentially relevant” articles which cite article i.
n is the number of all the ”potentially relevant” articles for the query.
By applying this parameter to each article, the system calculates its relevance weight. Then, these articles are ranked in descending order of their relevance weight.

4.2.4 Article survey graph

When a user, or a novice researcher, carries out a literature review on a specific research problem, he generally wants to obtain two things: firstly, that the system should propose a list of scientific articles dealing with the subject of the query; secondly, that the system should offer the possibility of visualising the evolution of scientific contributions over time, as well as the links existing between these articles. The article information retrieval system we are proposing offers the user these two possibilities.
Some heuristics are used to build the graph of the scientific articles survey.

- The graph only concerns the articles deemed relevant to the query. The system does not focus only on the citations between the different works to decide on the structure of the graph.
- An article can cite another appearing in the graph and not itself be part of the nodes of the graph. In other words, it is not the fact of citing or being cited by an article located in the graph that guarantees the presence in the graph.
- The presence of an article in the graph depends on the relevance of its content concerning the scientific problem identified in the query. Its position in the graph depends on the link existing between it and the other articles relevant for the query and in particular the reference article.

To build this graph, the system proceeds as follows:

1. The identifying all the articles relevant to the query and which must appear in the graph.
2. Next, for each article, it lists all the relevant articles cited by the latter. Each bibliographic reference is represented by the article’s title, the author’s names, the journal, and the date of publication.
3. Then, the graph building begins with the reference article. Starting from the reference article, the system identifies in its bibliographic references, all the relevant articles which are cited by it. It represents them in the graph by indicating the edge direction.
4. Next, it locates all the relevant articles citing the reference article. Once located, these articles are represented in the graph, specifying the relationship that links them with the reference article.
   Thereafter, the article represented on the vertice is deleted in the rest of the process.
5. When the system finishes with the reference article, it goes through the other vertices already represented in the graph to have the relevant articles which are cited or which cite them. It resumes the process for all the vertices and step by step, the graph is built.

6. If there are articles not citing any article appearing in the graph (or which are not cited by any article on the graph), then the system checks their similarity value with the reference article. If it is greater than a fixed threshold, it represents in the graph as an isolated vertice. Otherwise, they are not represented in the graph.

5 Implementation and experimentation of the proposed model

5.1 Description of the experimentation process of the proposed similarity measure

To implement the proposed similarity measure, some python libraries were used (Spacy, NLTK) for the text processing. Then, Wordnet was used to identify similar concepts in the different documents.

After the implementation of the proposed similarity measure, his ability to accurately define the similarity degree between two scientific paper is evaluated. However, evaluating a similarity measure is a difficult task. The notion of similarity degree between two documents is difficult to quantify accurately.

To efficiently evaluate a similarity measure, it is necessary to define a predefined dataset containing the data as well as the similarity values estimated as correct for each document pair. In the context of similarity measures on scientific documents, there is almost no such dataset.

The first step of this experimentation is the constitution of our dataset.

We have extracted articles from five research areas. These research areas correspond to the research fields of our experts. To be considered as an expert, you had to have published at least one scientific article in the field. These research areas are: ”Semantic Web and applications”, ”Algebraic coding theory”, ”Information retrieval system”, ”Fuzzy ontology modeling” and ”Question answering system”.

For each research area, two different experts in the field consulted articles to give an evaluation of the similarity value between two different articles.

This similarity value depended on the similarity between the keywords developed in each article and the research problems addressed by each article.

Two articles were said to be very similar if they dealt with the same problem and if their textual content was similar. The similarity value used was the average of the estimates made by each expert.

At the end of this process, we obtained a dataset containing the metadata on each article and a half thousand pairs of article with the similarity value.

Once the dataset is constituted, the proposed similarity measure is used on each pair of documents of the dataset to define their similarity value.

At the end of the process, a metric evaluation is used to assess the results obtained.

One of the most widely used metrics for evaluating relatedness measures is Pearson correlation. It indicates how closely the results of a measurement resemble human judgments. A value of 0 means no correlation and 1 means perfect correlation [56].

The cosine similarity is used on the dataset to compare the results obtained using our semantic similarity measure (OSSM) with those obtained using the traditional cosine similarity measure (CSM). Finally, an improved version of the cosine similarity (ICSM) measure using the proposed n-gram-based indexing approach is applied on the dataset.

5.2 Description of the process for testing the automatic survey model

To test the performance of our survey retrieval system on scientific articles, we implemented three approaches: The first approach uses the indexing approach based on the weight of the words contained in the metadata by using the type of metadata in which the word appears (Title, abstract or keywords) and its proportion in each of this metadata. After this, we used the cosine similarity measure (CSM) as the similarity measure applied to our model.

In the second approach, we use our n-gram based indexing approach to index the different articles and then apply the cosine similarity measure (ICSM).

The last approach uses the indexing approach proposed in this article together with our semantic similarity measure as defined in Section 3 (OSSM).

For each approach, we evaluated the behaviour of the system in terms of the accuracy of the answers returned to the user’s queries.

For this experimental process, we used a dataset consisting of 126000 scientific articles extracted from the arXiv dataset [1]. Our dataset contained articles dealing with topics specific to five research subfields.These research subfields were used to represent the different clusters. For each subfields, queries were defined concerning a set of scientific problems that could be investigated scientifically. These research problems are divided into two groups: general problems on the one hand and specific problems on the other hand.

A problem is said to be general if its scope is broad and vague. This is the case, for example, if we want to study a problem such as: ”The semantic web and its applications”, ”The theory of algebraic coding”, ”The information retrieval system”, ”Question answering system” and ”fuzzy ontology”.

However, a problem is said to be specific if its field of action is precise, circumscribed and leaves no doubt about

[1] https://www.kaggle.com/datasets/Cornell-University/arxiv

the intentions of the research. This is the case of a query of the type: "Dynamic modeling of the user's profile for the personalization of information retrieval"; "Semantic similarity measurement for predicates in linked data".
Once the system was up and running, five users carried out two hundred queries on our system. These queries were made for each of the three approaches above using CSM, ICSM and OSSM.
Each user was a novice researcher (at least Phd students) with expertise in one of the five subflieds. Finally, the performances of the system have been compared using firstly the semantic similarity measure (OSSM) proposed in this paper, secondly using the cosine similarity measure (CSM) and finally, using the improving cosine similarity measure (ICSM).

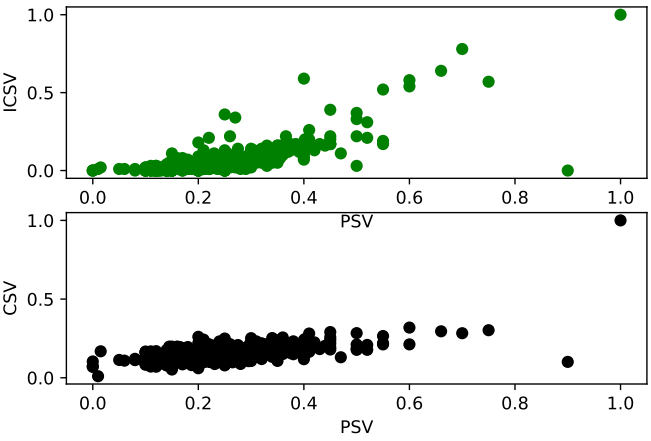


Figure 4: Correlation between (PSV, ICSV) and (PSV, CSV)

5.3 Discussion

5.3.1 Experimental results of similarity measure

At the end of these different phases of experimentation, the different similarity values obtained for each pair of documents were compared. Figure 3 presents firstly, the correlation between the values obtained with the proposed similarity measure (OSSV) with those estimated manually (PSV). This correlation is represented in blue color. The second scatter plot in black color, represents the correlation between the values obtained using the traditional cosine similarity measure (CSV) and those defined by the experts. From this figure, it appears that the points distribution in the cloud tends to be further away from the diagonal in the case of the traditional cosine similarity measure than for the proposed similarity measure.

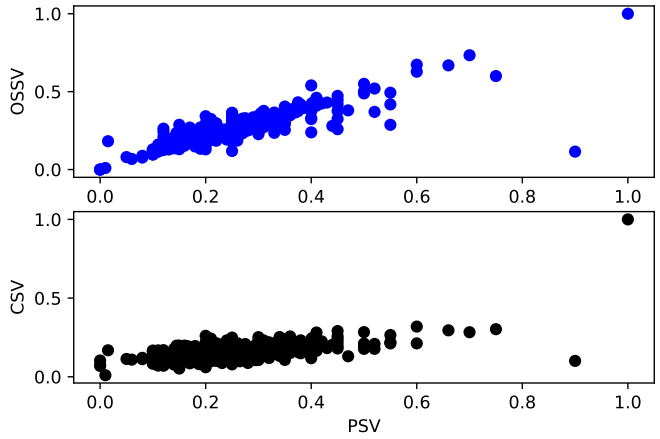


Figure 3: Correlation between (PSV, OSSV) and (PSV, CSV)

The Figure 4 presents the correlation between the values obtained with the improved cosine similarity measure using the proposed indexing approach (ICSV) and those estimated manually. This correlation is represented in green color. The second scatter plot, in black color represents the correlation between the values obtained with the traditional cosine similarity measure (CSV) and those defined by humans. From this figure, it appears that the points distribution in the cloud is almost similar for the CSV and for the ICSV, when the evaluated document pairs are not very similar. However, this points distribution is closer to the diagonal in the case of ICSV than in the case of CSV, when the evaluated document pairs are similar. This shows the ability of the proposed indexing approach to improving the evaluation and the detection process of similar document pairs.
From these two figures, it can be seen that the joint use of the proposed indexing approach with the proposed similarity measure improves the detection process of similarity value between two scholarly documents.
The Pearson correlation coefficient r was also calculated to assess the correlation between each similarity measure and the manually made estimates. The results are shown in Figure 5.

It emerges that the proposed similarity measure (OSSV) has a Pearson correlation coefficient of 0.8637, while the ICSV has a Pearson correlation coefficient of 0.7544. Finally, it appears that CSV achieve a Pearson correlation coefficient of 0.6234.
Having a similarity measure that accurately assesses the similarity value between two documents is very important for the information retrieval system. The more accurate the similarity measure, the more certain we can be that the system will be able to return the results that best match the need expressed for any query.
At the end of this phase, we can conclude that the semantic similarity measure we have proposed is more accurate than the traditional cosine similarity measure.
In the following, we will discuss the performance of our information retrieval system when using our similarity mea-

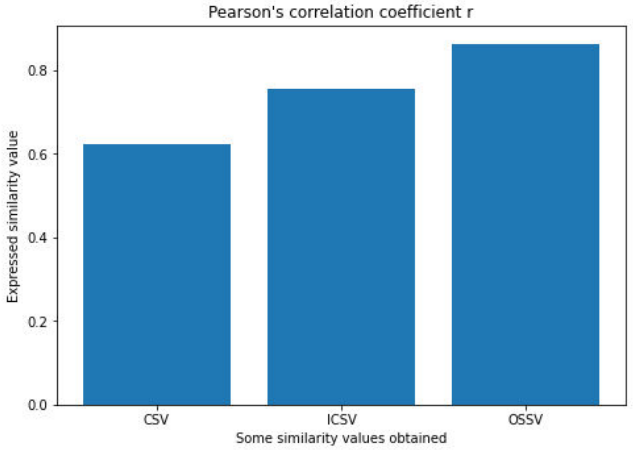


Figure 5: Pearson's correlation coefficient r of each similarity measure

sure and when using the cosine similarity measure. This will confirm the correlation between the accuracy of a similarity measure and the precision of the information retrieval system.

5.3.2 Experiment results of the proposed information retrieval system

The precision of a system is a parameter used to evaluate the proportion of good results returned in relation to the total number of results for a query.
A precision value close to 1 guarantees that all the results returned by the system meet the user's need expressed by the query.
For queries on general search problems, the comparison of the results obtained for each of the three indexing approaches and similarity measures can be found in Table 2.
For queries on specific (or precise) search problems, the

Table 2: Comparison of the precision of the system for queries concerning general search problems

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
CSM	0.79	0.84	0.7	0.8	0.63
ICSM	0.88	0.9	0.809	0.86	0.77
OSSM	0.86	0.909	0.92	0.88	0.87

comparison of the results obtained for each of the three indexing approaches and similarity measures can be found in the Table 3.
From these two tables, it can be seen that the first approach

Table 3: Comparison of the precision of the system for queries concerning specific search problems

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
CSM	0.64	0.87	0.92	0.76	0.648
ICSM	0.937	0.9	0.84	0.91	0.887
OSSM	0.9	0.89	0.93	0.97	0.88

is the least efficient overall for both "specific" and "general" queries. This is because the approach does not exploit the semantic relatedness that exists between the words in the query or in the various article metadata.
Using our indexing approach based on n-grams coupled with the cosine similarity measure, we obtain better overall performance than with the first approach. However, we note that this performance is rather unstable depending on whether we are dealing with questions related to one cluster or another. In Table 1, the accuracy for this approach is between 0.77 and 0.9. We would have expected to see a lower range. This is because with the second approach, the system returns very few responses to queries. Of all the possible results expected for a query, it returns only a small proportion. So the proportion of bad responses returned has a major influence on the accuracy value. This is not advantageous when you want to do a literature review on a specific research problem. Indeed, returning few results may suggest that there are not enough scientific contributions on the subject of the query. This is generally not the case.
Using the last approach (OSSM), the system's overall accuracy is better than with the two other approaches. This is due to the fact that with n-grams, the system is able to exploit the semantic relatedness between the concepts found in the articles. Not only can it exploit these semantic relationships, but it can also exploit the similarity between concepts, particularly when they appear to mean the same thing or opposite things.
Figure 6 summarizes the overall accuracy of the system when each of the two previous similarity measures is used. This figure shows the accuracy of the system when the

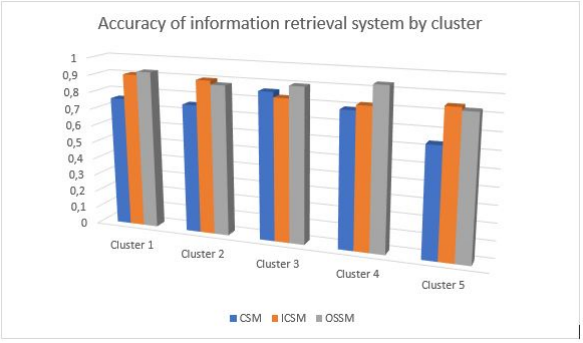


Figure 6: Comparison of results obtained with Cosine similarity measure (CSM), Improving Cosine similarity measure (ICSM) and with our semantic similarity measure (OSSM)

queries concern each research subfield represented by each cluster. This accuracy reflects the system's performance regardless of the type of problem submitted in the query (general or specific). Overall, the first approach performs less well than the two others. And the second less well than the third.
The figure 7 shows the behavior of the system by type of problem. It appears that the approach proposed in this paper offers better performance than the other two. On average, the system has an accuracy of 0.77 for the first approach, 0.85 for the second, and 0.89 for the approach developed in

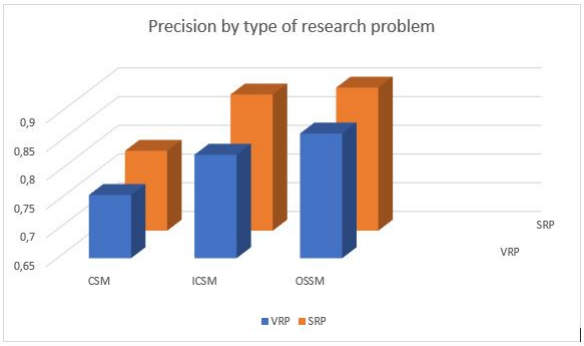


Figure 7: Comparison of accuracy per type of problem with CSM, ICSM and with OSSM

this paper.

The table 4 gives an overview of our system's performance in terms of precision, recall and F1-score using our proposed semantic similarity measure. From this table, it appears that our system has an overall recall of 0.85, while its F1-score is around 0.87.

Table 4: Global precision, recall and F1-score of the system

	Precision	Recall	F1-score
Cluster1	0.88	0.92	0.8995
Cluster2	0.9	0.81	0.8526
Cluster3	0.926	0.85	0.8863
Cluster4	0.925	0.76	0.8344
Cluster5	0.828	0.89	0.8578
Overall performance	0.89	0.8421	0.86

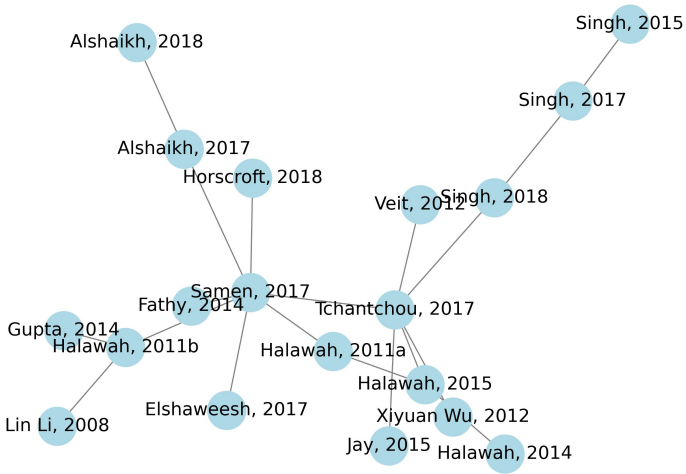


Figure 8: An example of survey graph generated by the model

Figure 8 shows an example of the results generated by the model for the following query: Dynamic user profile modeling approach to personalize information retrieval. This query aims to understand existing work on dynamic user profile modeling, given that users' interests while browsing may change over time. The novice researcher would like to have a clear idea of existing approaches and how they use dynamically collected information about the user to personalize the information to be sent back to the user. From this graph, the user can appreciate which works best address his concern, just by observing the different connections that exist between works over time. his concern.

6 Conclusion

In this paper, the problem of the automatic survey of scientific articles is addressed. The objective was to propose an information retrieval approach allowing a novice/confirmed researcher to visualize in a short time the evolution of contributions to a particular research problem. To this end, an approach to extract useful information from scholarly documents in Pdf format (metadata) is proposed. By using the information containing in useful metadata, a semantic similarity measure based on n-gram is defined. Finally, this similarity measure is used to implement an information retrieval system, allowing a user to investigate a problem by submitting a query to the system. The proposed system offers the option of presenting the survey results in graphical form.

From the implementation of the model, it is found that the proposed similarity measure can estimate with a good accuracy the similarity value between two scientific articles. For the experimentation of the similarity measure, a metadata dataset of scientific articles is constructed and their similarity values are estimated by human experts. Experiments with this dataset show that the similarity values obtained with the proposed similarity measure are very close to those manually proposed by human experts. Using this similarity measure, it finds that the system better refines the survey results to propose scientific contributions that are exclusively about the user's needs. Thus, the system's overall precision is about 0.9 while its recall is 0.8421, for an overall F1-score of 0.8652.

However, our system has a few limitations. The threshold of 0.5 set to be considered as potentially relevant for a query does not make it possible to identify articles that are relevant but whose textual content seems different from that of the reference article for a given query. This threshold was set to reduce the proportion of bad results returned by the system. It would be important to carry out a comparative study to identify the best threshold in terms of the system's performance.

Furthermore, in this work, we have only used metadata that is freely available. This metadata is not always sufficient to assess the similarity between two articles. It would be important in the future to develop an approach allowing to identifiy and extract the entire contents of a scientific article to be in PDF format (Introduction, conclusion, related work, and the other parts of the article). This will

enable the model to be readjusted to incorporate all this useful information into the information retrieval process. In the future, it would be interesting to address these limitations.

References

[1] P. Larsen and M. Von Ins. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics, 84(3), 575-603, 2010.

[2] H. Han, C.L. Giles, E. Manavoglu, H. Zha, Z. Zhang and E.A. Fox. Automatic document metadata extraction using support vector machines. In 2003 Joint Conference on Digital Libraries, 2003. Proceedings. (pp. 37-48). IEEE.

[3] S. Marinai. Metadata extraction from PDF papers for digital library ingest. In 2009 10th International conference on document analysis and recognition (pp. 251-255). IEEE.

[4] P. Lopez. GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In International conference on theory and practice of digital libraries (pp. 473-474), 2009. Springer, Berlin, Heidelberg.

[5] A. Constantin, S. Pettifer and A. Voronkov. PDFX: fully-automated PDF-to-XML conversion of scientific literature. In Proceedings of the 2013 ACM symposium on Document engineering (pp. 177-180).

[6] J. Liu, T. Tang, W. Wang, B. Xu, X. Kong and F. Xia. A survey of scholarly data visualization. IEEE Access, 6, 19205-19221, 2018.

[7] D. Tkaczyk, P. Szostek, M. Fedoryszak, P.J. Dendek and Bolikowski. CERMINE: automatic extraction of structured metadata from scientific literature. International Journal on Document Analysis and Recognition (IJDAR), 18(4), 317-335, 2015.

[8] Y.C. Yoon, H.K. Gee and H. Lim. Network Based Document Clustering Using External Ranking Loss for Network Embedding. IEEE Access, 7, 155412-155423, 2019.

[9] S. Zhang, Y. Xu and W. Zhang. Clustering scientific document based on an extended citation model. IEEE Access,7, 57037-57046, 2019.

[10] S. Giridhar, and K. Bhutani. Importance of Similarity Measures in Effective Web Information Retrieval. International Journal on Recent and Innovation Trends in Computing and Communication, 6, 29-33, 2018.

[11] D. Ifenthaler. Measures of Similarity. In: Seel N.M. (eds) Encyclopedia of the Sciences of Learning. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_503, 2012.

[12] R. Ibrahim, S. Zeebaree and K. Jacksi. Survey on semantic similarity based on document clustering. Adv. sci. technol. eng. syst. j, 4(5), 115-122, 2019.

[13] A. Souza, v. Moreira and C. Heuser. ARCTIC: metadata extraction from scientific papers in pdf using two layer CRF. In Proceedings of the 2014 ACM symposium on document engineering (pp. 121-130), 2014.

[14] S. Liu and C. Chen. "The proximity of co-citation", Scientometrics, vol. 91, no. 2, pp. 495-511, 2012.

[15] S.A. Salloum, C. Mhamdi, M. Al-Emran and K. Shaalan. Analysis and classification of Arabic newspapers Facebook pages using text mining techniques. International Journal of Information Technology and Language Studies, 1(2), 8-17, 2017.

[16] I. Safder, S.U. Hassan, A. Visvizi, T. Noraset, R. Nawaz and S. Tuarob. Deep learning-based extraction of algorithmic metadata in full-text scholarly documents. Information processing & management, 57(6), 102269, 2020.

[17] Z. Nasar, S.W. Jaffry and M.K. Malik. Information extraction from scientific articles: a survey. Scientometrics, 117(3), 1931-1990, 2018.

[18] C. Jiang, J. Liu, D. Ou, Y. Wang and L. Yu. Implicit semantics based metadata extraction and matching of scholarly documents. Journal of Database Management (JDM), 29(2), 1-22, 2018.

[19] S. Qiu and T. Zhou. A method of extracting metadata information in digital books. In 2019 10th International Conference on Information Technology in Medicine and Education (ITME) (pp. 583-586), 2019. IEEE.

[20] A.M. Hashmi, M.T. Afzal and S. ur Rehman. Rule Based Approach to Extract Metadata from Scientific PDF Documents. In 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA) (pp. 1 - 4), 2020. IEEE.

[21] G. Zaman, H. Mahdin, K. Hussain, J. Abawajy and S.A. Mostafa. An Ontological Framework for Information Extraction From Diverse Scientific Sources. IEEE access, 9, 42111-42124, 2021.

[22] S. Li and Q. Wang. A hybrid approach to recognize generic sections in scholarly documents. International Journal on Document Analysis and Recognition (IJDAR), 1-10, 2021

[23] B. Gipp and J. Beel. Citation proximity analysis (CPA): A new approach for identifying related work based on co-citation analysis. In ISSI'09: 12th international conference on scientometrics and informetrics (pp. 571-575), 2009.

[24] K.W. Boyack, D. Newman, R.J. Duhon, R. Klavans, M. Patek, J.R. Biberstine,... & K. Brner. Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. PloS one, 6(3), 2011.

[25] S. Li and Q. Wang. A hybrid approach to recognize generic sections in scholarly documents. International Journal on Document Analysis and Recognition (IJDAR), 1-10, 2021.

[26] R. Bhagdev, S. Chapman, F. Ciravegna, V. Lanfranchi and D. Petrelli. Hybrid Search: Effectively Combining Keywords and Semantic Searches. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds) The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science, vol 5021. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68234-9_-41, 2008.

[27] Z. BODO, L. CSATO. A Hybrid Approach for Scholarly Information Extraction. Studia Universitatis Babe?-Bolyai Informatica, [S.l.], v. 62, n. 2, p. 5-16, dec. 2017. ISSN 2065-9601. https://doi.org/10.24193/subbi.2017.2.01.

[28] G. Mitrov, B. Stanoev, S. Gievska, G. Mirceva, E. Zdravevski. Combining Semantic Matching, Word Embeddings, Transformers, and LLMs for Enhanced Document Ranking: Application in Systematic Reviews. Big Data and Cognitive Computing. 2024; 8(9):110. https://doi.org/10.3390/bdcc8090110

[29] N. RAMAN, A. NOURBAKHSH, S. SHAH, et al. Robust Document Representations using Latent Topics and Metadata. arXiv preprint arXiv:2010.12681, 2020.

[30] A. AHLUWALIA, B. SUTRADHAR, K. GHOSH, et al. Hybrid Semantic Search: Unveiling User Intent Beyond Keywords. arXiv preprint arXiv:2408.09236, 2024.

[31] M. Waqas, N. Anjum, and M.T. Afzal. A hybrid strategy to extract metadata from scholarly articles by utilizing support vector machine and heuristics. Scientometrics 128, 4349□4382 (2023). https://doi.org/10.1007/s11192-023-04774-7.

[32] K.W. Boyack, H. Small and R. Klavans. Improving the accuracy of cocitation clustering using full text. Journal of the American Society for Information Science and Technology, 64(9), 1759-1767, 2013.

[33] K. Rinartha and L.G.S. Kartika. Scientific article clustering using string similarity concept. In 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS), Vol. 1, pp. 13-17, 2019. IEEE.

[34] W.H. Gomaa and A.A. Fahmy. A survey of text similarity approaches. international journal of Computer Applications, 68(13), 13-18, 2013.

[35] Y. Song, X. Wang, W. Quan et al. A new approach to construct similarity measure for intuitionistic fuzzy sets. Soft Comput 23, 1985-1998, 2019. https://doi.org/10.1007/s00500-017-2912-0

[36] F. Lan. Research on Text Similarity Measurement Hybrid Algorithm with Term Semantic Information and TF-IDF Method. Advances in Multimedia, 2022. https://doi.org/10.1155/2022/7923262

[37] X. Wan. Beyond topical similarity: a structural similarity measure for retrieving highly similar documents. Knowl. Inf. Syst. 15, 1, 55-73, 2008.

[38] F.L. Liu, B.W. Zhang, D. Ciucci, W.Z. Wu and F. Min. A comparison study of similarity measures for covering-based neighborhood classifiers, Information Sciences, V. 448□449, pp. 1-17, 2018. https://doi.org/10.1016/j.ins.2018.03.030.

[39] R. Subhashini and V.J.S. Kumar. "Evaluating the Performance of Similarity Measures Used in Document Clustering and Information Retrieval", 2010 First International Conference on Integrated Intelligent Computing, pp. 27-31, 2010. doi:10.1109/ICIIC.2010.42.

[40] S. Wan and R.A. Angryk, "Measuring semantic similarity using wordnet-based context vectors," 2007 IEEE International Conference on Systems, Man and Cybernetics, pp. 908-913, 2007. doi:10.1109/ICSMC.2007.4413585.

[41] R. Mihalcea, C. Corley and C. Strapparava. Corpus based and knowledge-based measures of text semantic similarity. In Proceedings of the American Association for Artificial Intelligence.(Boston, MA),2006.

[42] F. Chen, C. Lu, H. Wu, and M. Li. A semantic similarity measure integrating multiple conceptual relationships for web service discovery. Expert Systems with Applications, 67, 19-31, 2017.

[43] A. Yousfi, M.H. El Yazidi and A. Zellou. "CSSM: A Context-Based Semantic Similarity Measure." 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS). IEEE, 2020.

[44] C. Little, D. Mclean, K. Crockett and B. Edmonds. A semantic and syntactic similarity measure for political tweets. IEEE Access, 8, 154095-154113, 2020.

[45] R. Meymandpour and J.G. Davis. A semantic similarity measure for linked data: An information content-based approach. Knowledge-Based Systems, 109, 276-293, 2016.

[46] A. Adhikari, B. Dutta, A. Dutta, D. Mondal and S. Singh. An intrinsic information content?based semantic similarity measure considering the disjoint common subsumers of concepts of an ontology. Journal of the Association for Information Science and Technology, 69(8), 1023-1034, 2018.

[47] Y. Jiang, X. Wang and H.T. Zheng. A semantic similarity measure based on information distance for ontology alignment. Information Sciences, 278, 76-87, 2014.

[48] A.J.M. Zou and M.R. Valizadeh. A proposed query-sensitive similarity measure for information retrieval, 2006.

[49] K. Pushpalatha and V.S. Ananthanarayana. ”An information theoretic similarity measure for unified multimedia document retrieval.” 7th International Conference on Information and Automation for Sustainability. IEEE, 2014.

[50] Y. Gupta, A. Saini and A.K. Saxena. Fuzzy logic-based approach to develop hybrid similarity measure for efficient information retrieval. Journal of Information science, 40(6), 846-857, 2014.

[51] C. Ramya, S.P. Paramesh and K. S. Shreedhara. ”A New Similarity Measure for Web Information Retrieval using PSO Approach.” 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS). IEEE, 2018.

[52] M. Eminagaoglu. ”A new similarity measure for vector space models in text classification and information retrieval.” Journal of Information Science, 2020.

[53] H. Ahmed. Detecting opinion spam and fake news using n-gram analysis and semantic similarity. PhD Thesis, University of Ahram Canadian, 2012.

[54] Y.U.T. Samen and E.C. Ezin. ”An Improving Mapping Process Based on a Clustering Algorithm for Modeling Hybrid and Dynamic Ontological User Profile”, 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 1-8, 2017. doi:10.1109/SITIS.2017.12.

[55] R. Baeza-Yates and B. Ribeiro-Neto. Modern information retrieval (Vol. 463). New York: ACM press, 1999.

[56] H.T. Mohamed Ali, T. Zesch and M.B. Aouicha. ”A survey of semantic relatedness evaluation datasets and procedures.” Artificial Intelligence Review 53.6, 4407-4448, 2020.

Enhanced Software Performance Testing for Big Data Platforms Using Clock-Controlled Computation Tree Logic with Particle Swarm and Genetic Optimization

Yuan Sun[1, 2], Md Gapar Md Johar[2, *], Jacquline Tham[2]
[1]School of Information Engineering, Gongqing Institute of Science and Technology, Jiujiang 332020, China
[2]Postgrduate Center, Management and Science University, Shah Alam 40100, Malaysia
E-mail: Yuan Sun: sunyuan12123@gmail.com, Md Gapar Md Johar: mdgapar@126.com, jacquline@msu.edu.my
[*]Corresponding author

Keywords: software testing, particle swarm algorithm, bell-controlled computational tree logic method, genetic algorithm, testing effectiveness

Received: November 22, 2024

This research aims to solve the problems of testing inefficiency and lack of accuracy in software testing, and proposes a software performance testing system for big data platforms based on the clock-controlled computational tree logic method. The particle swarm algorithm finds the optimal solution through the movement and mutual cooperation of particles in the search space. Genetic algorithm evolves the population through selection, crossover, and mutation operations, ultimately finding the optimal solution. Secondly, long short-term memory networks and linear autoregressive models also have advantages in software testing, which can improve the effectiveness and efficiency of software testing through reasonable selection and combined use. The new algorithm utilizes the ability of PSO and GA algorithms to search for optimal solutions through particle motion and group cooperation in the search space, in order to determine key moment parameters and other relevant information in software testing systems. The research uses the algorithmic logic of the particle swarm algorithm and the genetic algorithm to confirm the moment parameters and other information of the software testing system. At the same time, an algorithmic model research on the joint coverage and the use of the value of the system, and finally makes use of the big data platform to analyze the research system. The specific indicators used in the study include 100% test case coverage, as well as the functional coverage of genetic algorithms and particle swarm optimization algorithms. The innovative combination of CCTL method and optimization algorithm in the research has improved the accuracy and stability of software testing. CCTL is an extended computational tree logic that introduces the concept of time, allowing testers to explicitly specify time constraints in software testing, thereby more accurately simulating real-world scenarios. The research results show that using the system to test software can achieve a coverage rate of 100% for its component use cases, while the functional coverage rates of genetic algorithm and particle swarm algorithm reach 90.36% and 91.32%, respectively. The accuracy of software testing research methods is 5% and 6% higher than that of LSTM and LAR methods. When the moment range of the particle parameter position information of the model is [150 ms, 250 ms], the maximum value of the target parameter velocity is 80 m/s and the minimum value is 0 m/s. The maximum value of the target azimuth velocity is 20 rad/s, and the minimum value is 0 rad/s. The system is able to determine the various parameters of the software, and at the same time in the software test results on the test results are normal, fault analysis can be completed normally, the performance of the algorithm is also superior to other algorithm models such as LSTM and LAR, and the study of the use of algorithms with a higher degree of stability. It can be seen that the system and methodology used in this research is superior to traditional methods and the test results of software testing have improved. This study provides a new research direction for platform software afterwards.

Povzetek: Opisan je sistem za testiranje zmogljivosti programske opreme na platformah za velike podatke, ki uporablja metodologijo CCTL, optimizirano s pomočjo algoritmov PSO in GA.

1 Introduction

In the current digital age, big data platforms have become a core technology for processing complex data sets. With the increasing volume of data, it is critical to ensure stable platform software performance [1]. Software performance testing is a key component in ensuring efficient and accurate data processing [2]. The existing performance testing methods for big data platform software suffer from low efficiency and insufficient accuracy when dealing with large-scale data and complex scenarios. Therefore, how to build an efficient and accurate method for testing the performance of big data platform software has become a difficult problem that still needs to be solved [3]. The research objectives specifically include improving testing accuracy, expanding testing coverage, and clarifying the

performance of CCTL in handling specific types of constraints, as well as enhancing the adaptability of the model in handling large-scale datasets and real-time parameter changes. Simultaneously studying the hypothesis that combining CCTL methods and optimization algorithms can significantly improve the accuracy and coverage of big data platform software testing, optimization algorithms can effectively address the limitations of CCTL in processing large-scale datasets and real-time parameter changes, and the combination of PSO and GA can provide more stable and comprehensive testing results in different testing scenarios. These algorithms are used to optimize the search for critical moment parameters and other relevant information in software testing systems, thereby improving the accuracy and efficiency of the testing process. In the current digital age, big data platforms have become the core technology for processing complex datasets. As the amount of data increases, ensuring the stability of platform software performance becomes crucial. Software performance testing is a key component in ensuring efficient and accurate data processing. The existing performance testing methods for big data platform software face problems of low efficiency and insufficient accuracy when dealing with large-scale data and complex scenarios. Clock-Controlled Computation Tree Logic (CCTL) is a temporal logic that improves the traditional Computation Tree Logic (CTL) by introducing the concept of time [4]. Long Short Term Memory Network (LSTM) is a special type of recurrent neural network that can remember cells and gate mechanisms to solve the problems of gradient vanishing and exploding in traditional RNNs when processing long sequence data. Linear Autoregressive Model (LAR) is a statistical model used for time series analysis. This approach is particularly important in Big Data processing, where it allows the testing process to consider the temporal properties of the data flow, thus more accurately simulating real-world situations. CCTL can more effectively identify and analyze performance bottlenecks and potential problems in big-data platforms. Although the clock controlled computation tree logic method has achieved certain application results in other fields, there is still relatively little research on its application to software performance testing on big data platforms. CCTL is used to verify the temporal attributes of critical tasks and in fields such as aviation electronics, automotive control, and industrial automation. It ensures that data transmission and reception are completed within specified time intervals, which is crucial for network design and optimization. The novelty of the research lies in the innovative combination of clock controlled computation tree logic with PSO and GA, which not only achieves 100% test case coverage, but also enhances adaptability to different testing environments and conditions through optimized algorithm logic. In addition, this method has broad application prospects in potential fields such as big data processing, cloud computing platforms, Internet of Things (IoT) devices, as well as artificial intelligence and machine learning. Firstly, the new model utilizes the algorithm logic of particle swarm optimization and genetic algorithm to improve and analyze the process, in order to enhance the accuracy and stability of software testing parameter validation. Secondly, the research constructed testing scenarios and test cases suitable for big data platforms, and verified the performance and accuracy of the new method in different testing environments through experiments. And provided new ideas and methods for performance testing of big data platform software. By utilizing CCTL, the accuracy and efficiency of software performance testing have been significantly improved, especially for big data platforms that handle dynamic and large-scale datasets. The proposed method can be applied to other big data platforms, providing a scalable and effective performance testing solution applicable to various fields. This research is divided into four parts; the first part is an overview of domestic and international research; the second part is a study of the system and method of software testing; the third part is mainly to test and analyze the performance of the system; and the fourth part is a summary of the current research.

2 Literature review

Software is usually tested for different problems; therefore, different research methods are required to solve these problems. To address the complexities of data analysis encountered by strength and conditioning professionals who use strength platforms to conduct CMJ assessments during training, this study proposes a solution to create a data analysis program using MATLAB. The findings suggest that the program can help coaches simplify the process and improve the accuracy and reliability of the data analysis. In addition, the sample scripts provided allow further learning and mastery of basic scripting strategies to create separate analysis programs for the CMJ and other performance tests [5]. Kaur and Agrawal proposed a new approach based on the Bat Search Algorithm and the Cuckoo Search Algorithm to solve the problem of regression test case selection and improve the efficiency and accuracy of software maintenance. The results of this study show that both algorithms are effective in reducing the number of required test cases and improving the testing efficiency in regression testing. Among them, the cuckoo search algorithm is slightly better in terms of performance parameters. The algorithm proposed by Kaur and Agrawal performs well in reducing the number of test cases and improving testing efficiency, but still has limitations when dealing with large-scale datasets [6]. Chen et al., proposed an auxiliary method based on machine learning to study the benchmarking method in performance unit testing. The results of this study show that the method can effectively identify benchmarking methods, thus improving the accuracy and efficiency of performance unit testing. It was also found that the Random Forest algorithm performs the best in predicting performance and can retrieve 43% of the true BDMs by examining only 5% of the candidate methods detected by the model. Chen et al.'s method can effectively identify benchmark methods and improve the accuracy and efficiency of testing. However, this method has poor adaptability when dealing with dynamic and large-scale datasets [7]. In order to

address the problem of parameter estimation for software reliability growth models, this paper proposes a framework modelled on the Non-Homogeneous Poisson Process (NHPP). The framework integrates test coverage (TC), error propagation and troubleshooting efficiency while limiting the number of parameters. The results show that the model is more reliable than the existing models and can effectively assess and predict software reliability. Through sensitivity analyzes, we demonstrated that the model parameters have less impact on the mean function. The model proposed by Khurshid et al. performs well in handling test coverage, error propagation, and troubleshooting efficiency, but has shortcomings in handling real-time parameter constraints [8].

Qian et al. propose a method for prioritizing test scripts to address the memory bloat problem in web applications, as well as to improve the efficiency of performance testing. The new method uses a learning ranking technique to predict which test scripts are more likely to cause memory bloats, and thus prioritizes the execution of these scripts. Experimental results show that the method is effective in speeding up testing and improving the efficiency of detecting memory bloats. The method proposed by Qian et al. significantly improves the efficiency of detecting memory expansion. However, this method is mainly targeted at web applications and lacks adaptability to complex scenarios on big data platforms [9]. To fill a gap in the research on programming language security, this study proposes a methodology for benchmarking the security and performance of languages. The methodology compares six well-known programming languages and uses quantitative and qualitative methods to determine which language is best in terms of security and performance by testing the code and analyzing the available information. The results of the study show that the Rust performs best in terms of security and performance, achieving an excellent balance [10].

To investigate the effectiveness of Metamorphic Testing (MT) in different application contexts, this study revisited the use of MT in Sentiment Analysis (SA) systems and found that false satisfaction is an important factor affecting the validity of MT. An in-depth analysis of false gratification reveals how it can occur and how it can affect the effectiveness of MT. Our study also suggests that MT may overestimate the consistency of the system with the relevant MR if the occurrence of false satisfaction is not taken into account. These findings will help the MT community use MT test results more fairly and reliably [11]. To compare the predictive performance of an ensemble species distribution model with that of a single model, a study was conducted using a large eucalyptus species presence-absence dataset. Two spatial blocking strategies were used to partition the dataset, and all models within the calibration fold were calibrated and cross-validated using repeated random partitioning of data and spatial chunking. The results of the study showed that the ensemble models performed well in some tests, but did not always outperform their untuned individual models or the tuned BRT. Additionally, good external performance was obtained by selecting untuned individual models with the best cross-validation performance [12]. Hosseini et al. proposed a quantitative data error propagation rate and a mutation location recognition method based on genetic algorithm to reduce the cost of mutation detection. The research results showed that this method effectively reduced the number of mutants by about 24%, while increasing the mutation score by about 5.6%. Only 7.46% of the generated mutants were equivalent, significantly reducing testing time and cost [13]. Zeb et al. found that heuristic algorithms have been well studied in multiple fields, among which the use of heuristic algorithms such as particle swarm optimization in software testing can reduce the defects of software testing, improve the accuracy and reliability of software testing. It can be seen that using particle swarm optimization algorithm can improve the accuracy of software testing [14]. Pan et al. proposed a similarity search test case minimization technique based on genetic algorithm to improve the efficiency and fault detection capability of software testing. The research results indicate that the new method achieves a higher average fault detection rate compared to the existing technology FAST-R, with only 50% of test cases running [15].

In summary, existing research has made significant progress in the field of software testing, but there are still some shortcomings. Although MATLAB's data analysis program simplifies the process, it relies on specific environments; Although machine learning methods have improved the efficiency of regression testing, they lack applicability and have poor performance in evaluating data. Research has shown that although other framework models can accelerate the detection of webpage memory inflation, there is still a problem of poor network environment testing [16]. Therefore, this study proposes a new solution to address the issues of insufficient accuracy and performance in software testing. Firstly, the clock controlled computation tree logic method is used to generate software moment cases in the big data platform, which solves the problems of environment dependence, testing efficiency, and accuracy in existing research for software performance testing in big data platforms. Secondly, the model selects parameters such as moment examples and determines their values to ensure the accuracy and performance of software testing, solving the problems of limited applicability and unstable results of existing methods.

3 Method

3.1 Analysis of CCTL model

This chapter mainly focuses on the time platform for software testing to build a system, using the CCTL method to analyze the software testing system, build the moment component use case generation model and the software testing system model, and then analyze the system model to achieve system building for software testing performance. The main workflow of the current research is shown in Figure 1.

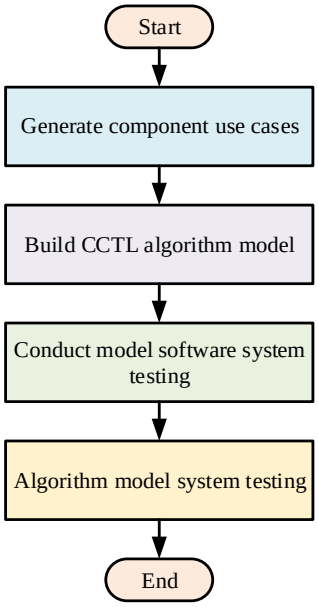


Figure 1: Main workflow

Description: Shows the main workflow.

From Figure 1, it can be seen that the first step is to analyze and generate test cases for the software. Secondly, by using the CCTL algorithm model, a software testing system based on the algorithm model is built. Then, based on the constructed model, implement software testing. Finally, the algorithm model constructed was tested for its actual effectiveness through experiments.

3.2 Research on software testing model

In the software for testing, software parameters of the moment input will have certain requirements, must be input in a specific time parameter to make the whole operation is effective, through this effective time input to be able to follow up on the input function operation, this time point and parameter point is the current software input parameters of the space moment. Generally, there are three types of input time space for software: interval input, cycle input, and discrete input. The interval input selects a fixed point in the cycle and selects a moment for input; the cycle input selects a time moment within the cycle and inputs different time points; the discrete input selects any time point from the discrete time collection for input. Interval input is suitable for systems that require regular and consistent data input, loop input is suitable for systems with periodic tasks, and discrete input is suitable for systems with irregular or user driven events. By classifying the input temporal space in this way, we can better understand and model the temporal behavior of software systems, which is crucial for accurate and effective software testing.

The number of parameters covered by the input will inevitably exist because of the moment processing constraints, and the expression of the constraints before the test generation is an important step in the analysis of its parameters. The general constraint moment is divided into two types: independent moment constraints and related moment constraints, which are mainly due to the non-existence of correlation between the parameters and parameters. Therefore, the two parameters do not affect each other, and at the same time the parameters simultaneously have their own moment constraint limitations. Independent moment constraints are temporal constraints on individual parameters that are not dependent on other parameters, thus simplifying the testing process and increasing testing efficiency. Correlation moment constraints, on the other hand, involve temporal dependencies between multiple parameters, ensuring that multiple parameters work together at a specific point in time or time range. Separating the independent moment constraints and correlated moment constraints into two distinct parts improves the clarity, efficiency and accuracy of testing. Correlated moment constraints, on the other hand, refer to the existence of identical moment constraints as well as different correlated moment constraints between two parameters. The CCTL method enables the description of constraints to reduce the moment constraints of the parameters. As shown in Eq. (1) is the independent time-constraint formula for the CCTL method [17].

$$\begin{cases} EX_{t_0,t_1}\phi \\ EX_{t_0,t_1}(\phi)^n \end{cases} \tag{1}$$

In Eq. (1), t_0, t_1 denotes the time interval, EX denotes the relationship between the constraints present, ϕ denotes the event expression of the input parameters, and n denotes the number of events satisfied by the test. The times of the different parameter constraints in the CCTL method can be expressed as shown in Eq. (2) [18]:

$$EX_{t_0,t_1}\phi \to EX_{t_2,t_3}\psi \tag{2}$$

In Eq. (2), ψ is expressed in terms of execution in the interval time, and the rest of the parameter expressions are the same as above. The constraint expression for the same moment means that at this time, the time will move to the next pointing interval after execution in that interval. The parameter expression for time can be substituted for the separate moments. When the time moments are replaced as separate moments the relative time constraints are also induced as n, the expression is shown in Eq. (3).

$$EX_{t_0,t_1}\phi \to EX_{t_2,t_3}\psi \to \cdots \to EX_{t_{2n-4},t_{2n-3}}\omega \to EX_{t_{2n-2},t_{2n-1}}\zeta \tag{3}$$

In Eq. (3), ω, ζ indicates the input conditions for different parameters. The remaining parameters were the same as those decribed. Because the existence of constraints will cause some combinations to not be in a time test case at the same time, for the current time constraints software test cases need to deal with constraint combinations; typically, there are four ways to deal with time constraints under the CCTL approach: abstract parameters, sub-models, substitution, and avoidance of

selection methods. Abstract parameters simplify the model and reduce complexity by abstracting specific temporal parameters, but can lead to oversimplification of the model and loss of important details. Submodels decompose a complex model into multiple submodels, but coordination between submodels can be complex, increasing the complexity of the overall model. Substitution methods simplify temporal constraints, but may lead to loss of temporal information. Selection methods choose specific temporal parameters or time points to avoid conflicts and ensure model consistency, but may require additional logic and computation, increasing complexity. Their principle is to transform the models and convert the models that appear to be in conflict with valid combination methods. However, the problem with this method is that when large parameters are encountered, more unnecessary and redundant information parameters appear [19]. At the same time, when using the CCTL method for software parameter moment determination and combinatorial testing, it is necessary to input a large number of consecutive parameters; therefore, it is necessary to study its parameter coverage during the analyzis. At this time it is necessary to use generative algorithms to study and analyse the parameter inputs of the method.

3.3 Time parameter combination and generation algorithm model

In the case of continuous input and transmission of the software moment parameters selected by the CCTL method, the parameter information at this time is analyzed. The algorithm model of data analysis selects the particle swarm algorithm and genetic algorithm for analysis, through the parameters of the population optimal solution and evolutionary optimal solution to find, to output the optimal value of the current parameters, to achieve the analysis of the parameters of the judgement. The analysis of parameter coverage helps ensure that all possible input conditions and scenarios are fully tested, thereby improving the comprehensiveness and accuracy of testing. PSO and GA algorithms can better generate test cases, ensuring the reliability and effectiveness of test results. At the same time, these algorithms perform well in parameter optimization problems and can quickly find the optimal solution, thereby improving the efficiency and accuracy of testing. The algorithm joint process includes PSO and GA initializing populations separately, with each individual representing a test case or parameter combination. Simultaneously, both models undergo iterative optimization and evaluate the quality of test cases through fitness functions. Finally, during the iteration process, excellent individuals are exchanged between PSO and GA to improve optimization efficiency. Figure 2 shows the flow of the moment-combination generation algorithm.

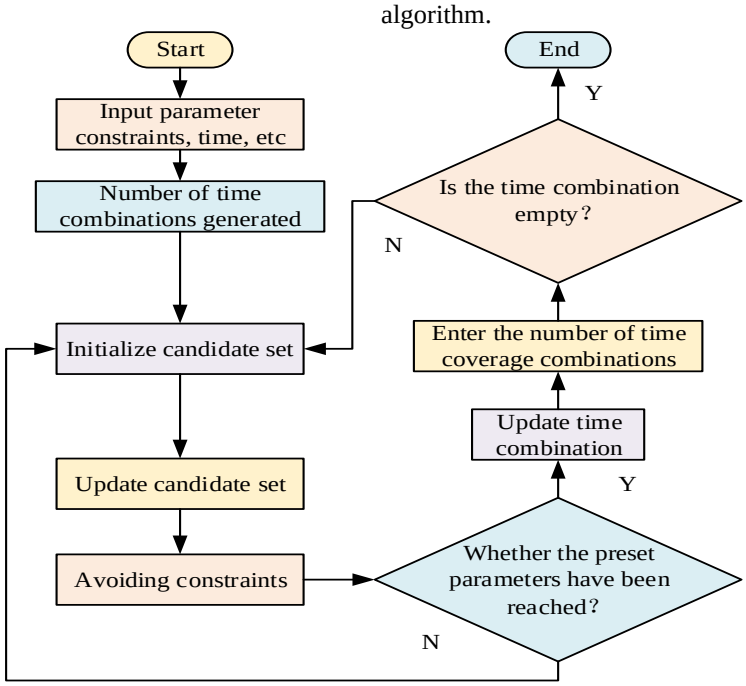


Figure 2: Time combination generation algorithm process

Description: Illustrates how the parameter information is analysed by particle swarm algorithm and genetic algorithm.

As shown in Figure 2, in the analysis phase of the algorithm, the number of parameters is first input to select the constraints and input moments, after which the combination of parameters at the current moment is generated, the parameter candidate set is initialized, and the set is updated so as to select a better individual for constraint evasion. Then, it is judged whether the current parameters under test reach the pre-set parameter data, and if they do, the combination of the target time is updated and then the combination is output. If it is reached, then

update the target time combination and output the combination; if it is not reached, then reinitialise the candidate collection. In software performance testing, selecting the optimal value of the current parameter is achieved through optimization algorithms. The algorithm gradually optimizes the combination of parameters to find the optimal test case. After outputting the number of combinations again, judge whether the combination set is empty; if it is empty, end the algorithm; if not, initialize the candidate set. Firstly, when generating examples, the model initializes the population of genetic algorithm and the particle swarm of particle swarm algorithm, with each individual representing a time combination. Based on the differences in algorithm structures, generate and optimize time combinations separately. Then regularly exchange individuals of genetic algorithm and particle swarm algorithm, add excellent particles from particle swarm algorithm to the population of genetic algorithm, and add excellent chromosomes from genetic algorithm to the particle swarm of particle swarm algorithm. Finally, after the iteration is completed, the results of the two algorithms are fused and the optimal time combination is selected as the final solution. When the algorithm is running, an initial population is formed by randomly generating a set of individuals, which ensures the diversity of the population. Secondly, setting a larger population can provide more solutions, but the computational complexity will also increase; Smaller populations require less computation, but may not be able to cover all possible solutions. The evolutionary process of the population requires four steps: population selection, population crossover, population mutation, and population replacement. Finally, suppose that when solving a problem, individuals can be represented using binary encoding. In software performance testing, genetic operations such as crossover and mutation can be easily performed by encoding parameter values as binary strings. First, the initial population randomly generates binary strings, and then the selection operation selects individuals with high fitness. Finally, a portion of new individuals will replace some individuals in the current population, maintaining the continuous evolution of the population. The data input process of the algorithm first selects the number of data parameters as k. Then generate a combination function through target coverage combination, which includes combinations between parameters, combinations between parameter values and input time, and direct combinations between input parameter data and time. After selecting the set of algorithm data parameters, the algorithm data is expressed by integrating the data parameters. In software performance testing, selecting and integrating data parameters is a key step in ensuring the effectiveness and accuracy of the testing. By selecting parameters directly related to the testing objectives and ensuring that these parameters comprehensively cover the testing scenario, the efficiency and accuracy of testing can be improved. By optimizing parameter combinations through PSO and GA algorithms, optimal test cases can be generated. This not only improves the efficiency and accuracy of testing, but also ensures that the test results can truly reflect the performance of the software under various conditions. At the same time, for the purpose of concluding the data space, all the parameter data mentioned above are collected. Due to the need to determine both time data parameters and initial candidate sets during algorithm execution, the study selects two parameter values p_1, p_2 through the algorithm process, and the set of parameter sets is shown in Eq. (4).

$$CT^{n}_{p_1,p_2} = \{(a,b) \mid a \in [0, |v_1| - 1], b \in [0, |v_2| - 1]\} \quad (4)$$

In Eq. (4), $CT^{n}_{p_1,p_2}$ denotes the set of parameters, a,b denote the constraint moment parameter expression at that moment, and $|v_1|$ denotes the numerical magnitude of the parameter expression. At this time, the particle swarm algorithm is used to obtain the particle velocity formula as shown in Eq. (5) [20]:

$$v_i^t = (v_{i1}^t, v_{i2}^t, \cdots, v_{ik}^t) \quad (5)$$

In Eq. (5), v_i^t represents the total position information of the particle, and $(v_{i1}^t, v_{i2}^t, \cdots, v_{ik}^t)$ represents the particle velocity information at different moments. The position information at this time is expressed as shown in Eq. (6).

$$x_i^t = (x_{i1}^t, x_{i2}^t, \cdots, x_{ik}^t) \quad (6)$$

In Eq. (6), x_i^t represents the total position information of the particle, and $(x_{i1}^t, x_{i2}^t, \cdots, x_{ik}^t)$ represents the position coordinate information of the particle at different moments. In software performance testing, the calculation formula for particle velocity is the core part of PSO algorithm, used to optimize parameter combinations. By adjusting the speed of particles reasonably, the movement speed of particles in the search space can be accelerated, thereby finding the optimal solution faster. This not only improves the efficiency of testing, but also enhances the accuracy of testing. Chromosome combination query refers to evaluating the fitness of chromosomes through fitness functions in genetic algorithms, and selecting chromosomes with high fitness for optimization. This process involves querying and selecting chromosome combinations to generate better combinations of test case parameters. Through these steps, the efficiency and accuracy of test cases can be ensured, and the efficiency and reliability of testing can be improved. Because the position information of the particle in the algorithm calculation is a vector coordinate of a dimension, the position of each coordinate needs to correspond to a specific moment in the dimension to be selected, and the value of the position coordinates at this time is shown in Eq. (7).

$$x_{ij}^{t} \in [0, |v_j| - 1] \tag{7}$$

In Eq. (7), the parameter expression is the same as that described above: The chromosome combination query by initialization adaptation comparing the current parameters completes the extraction of chromosomes; then, at this time, the adaptation is calculated as shown in Eq. (8) [21]:

$$p_i = \frac{s_i}{\sum_{i=1}^{k} s_i} \tag{8}$$

In Eq. (8), p_i denotes the corresponding adaptation value of the chromosome and s_i denotes the calculated probability of the population. The fitness function is used to evaluate the performance of individuals in the search space. By quantifying the proximity of individuals to the objective function, it provides a measurement standard for the algorithm. In software testing, the fitness function can help assess the quality of test cases, such as whether they can cover more functional modules and whether they can detect potential defects. Individuals with higher fitness values have a higher probability of being selected. The genetic algorithm probability calculation formula is given by Eq. (9).

$$s_i = \frac{\sum_{i=1}^{i} p_i}{\sum_{j=1}^{k} p_j} \tag{9}$$

Eq. (9) is shown, and the parameter expression is the same as that above. Through the crossover calculation after the expression of the operator of the above formula there will be a new chromosome pairing, and the crossover probability at this time is indicated by p_c, The crossover method in the method algorithm selects a point crossover through the crossover and then the exchange of the number of gene positions. The dot crossing method generates two offspring chromosomes by selecting a random crossing point and exchanging gene fragments of two parent chromosomes. Gene location plays a crucial role in this process, determining which genes will be exchanged to generate new parameter combinations. By selecting appropriate intersection points and exchanging gene fragments, the point crossing method can effectively explore the search space, generate better combinations of test case parameters, and improve the efficiency and accuracy of testing. Simultaneously, in the algorithm performed by particle updating, as shown in Eq. (10) [22, 23].

$$v_{ij}^{t+1} = v_{ij}^{t}\omega + c_1 r_1 (pBest_{ij}^{t} - x_{ij}^{t}) + c_2 r_2 (gBest_{j}^{t} - x_{ij}^{t}) \tag{10}$$

In Eq. (10), ω denotes the weight value of the inertia factor, c_1, c_2 denote the learning factor, r_1, r_2 denote random numbers in the interval 0-1, $pBest_{ij}^{t}$ denotes the limit of individual values, $gBest_{j}^{t}$ denotes the global limit value, and i, j denotes the iterative updating completed in the first particle. Iterative updating of the particles can achieve optimal selection of the current parameters. Because the data selected in the time selection of software parameters are only valid for a period of time, it is necessary to perform time selection and value constraints before executing the CCTL method. Therefore, the joint algorithm combining value and time is shown in Figure 3.

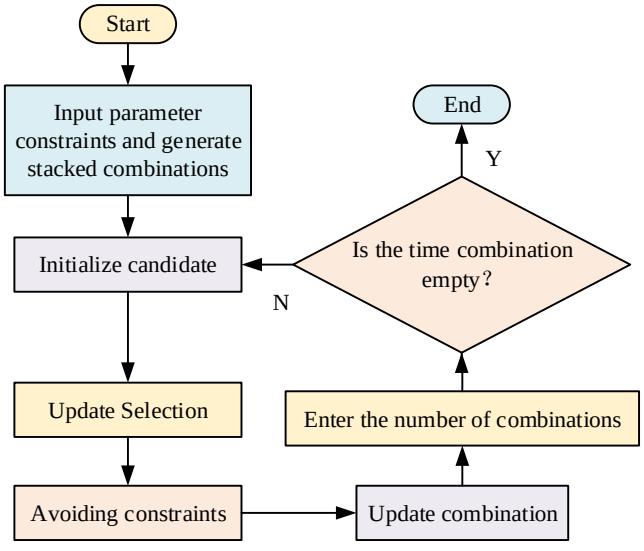


Figure 3: Joint algorithm flow combining value and time

Description: Explains the detailed steps to generate the final coverage combination during testing.

Figure 3 shows that, the algorithm input the parameters to be tested, time selection, and related constraints. Based on the input parameters and constraints, generate preliminary coverage combinations. Then initialize the candidate individual set and generate the final coverage combination through subsequent selection and optimization steps. Secondly, based on the current candidate individuals and coverage combinations, select new candidate individuals to optimize the coverage combination, ensuring that the selected individuals meet the constraint conditions. And when selecting new candidate individuals, it is necessary to ensure that they do not violate the previously set constraints. After selecting eligible candidate individuals, update the target coverage combination, generate the current joint coverage combination, and output the result. Finally, determine whether the currently generated joint coverage combination is an empty set. If it is an empty set, it indicates the end of the algorithm and the final test result can be output. If it is not an empty set, proceed to the next round of candidate initialization and selection. Selecting and optimizing new candidate individuals by evaluating the performance of each individual in the current population, selecting outstanding individuals, and generating new candidate individuals through crossover and mutation operations can ensure that the algorithm can effectively explore new solution spaces and find better solutions. In software performance testing, the joint algorithm combines the advantages of PSO and GA algorithms to optimize parameter combinations and generate high-quality test cases. The process of joint algorithm includes inputting parameters and constraints, generating initial parameter combinations, optimizing parameter combinations, evaluating and updating, and checking termination conditions. Through these steps, the joint algorithm not only improves the efficiency and accuracy of testing, but also adapts to complex testing scenarios, ensuring that the test results can truly reflect the performance of the software under various conditions. After the judgment is completed, the user inputs a new number of joint combinations to prepare for the next round of candidate individual initialization and optimization process. In the initialization of candidate individuals, to facilitate subsequent calculations, needs to be added to the set of values of the parameters through the construction of a simple index relationship to achieve its dimensionality reduction process. The formula is given by Eq. (11) [24, 25].

$$index = i\left|T_i\right| + t_i \tag{11}$$

In Eq. (11), $index$ denotes the index of the computation and $i\left|T_i\right| + t_i$ denotes the Cartesian set of the set composition. Dimensionality reduction is a data processing technique primarily aimed at simplifying data, improving efficiency, and enhancing interpretability. The parameters referred to in dimensionality reduction usually include test case parameters, system performance parameters, and constraint conditions. The idea of the particle swarm algorithm and genetic algorithm is also used in the selection of the combination of the union for the change in the same way as described above. PSO and GA are chosen for parameter information analysis because these two algorithms perform well in dealing with complex parameter optimisation problems, and are able to find the global optimal solution quickly and adapt to different test environments and conditions. The analysis process includes initialising the parameter candidate set, optimising the parameter combinations, evaluating the results and iterative updating. However, the difference lies in the update selection of their algorithms using the update as in Eq. (12) [26].

$$\begin{aligned} v_{ij} &= d_{ij} / \left|T_j\right| \\ t_{ij} &= d_{ij} \% \left|T_j\right| \end{aligned} \tag{12}$$

In Eq. (12), i denotes the first chromosome, j denotes the chromosome position, d_{ij} denotes the gene value size at the j position, and $\%$ denotes the residual value. The parameter analysis of the software test and the moment selection in the CCTL method are completed by generating and overriding the target parameters.

3.4 Analysis of software testing model system

The main purpose of the software testing tool system is to load and parse software models, generate test case parameters and constraint combinations, and import test cases into the testing database. The CCTL method serves as the core logical foundation and generates test cases by introducing time constraints; PSO and GA algorithms are used to optimize the generation process of test cases, improve testing efficiency and coverage. The testing tool system of software mainly exists in the form of components; the main function is to load the model, after completing the parsing to obtain the data information of the protocol, and then to match the strength constraints and other requirements using case analysis. A use case analysis of the software testing tool is shown in Figure 4.

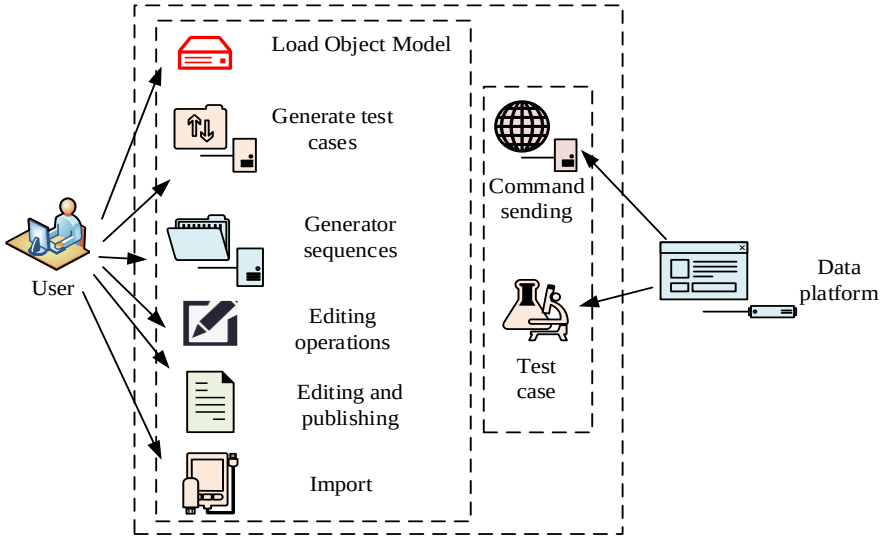


Figure 4: Use case analysis of software testing tools

Description: Demonstrates the use case analysis of user task operations in a software testing system

As shown in Figure 4, in the software testing system, the user needs to load the software model and generate test software case parameters and combinations of constraints and other settings, while inputting the sequence of generation events, editing the current need to release the task operation, edit the current need to release the software testing methods, as well as test cases and other information for the import and export operations. At the same time, the system needs to join the communication protocol to manipulate and receive instructions and send the current software test cases and test cases generated into the database to complete the import and transfer process of the test database components. Through a specific analysis of the system, it is necessary needs to add it to the dynamic analysis system of the model, as shown in Figure 5.

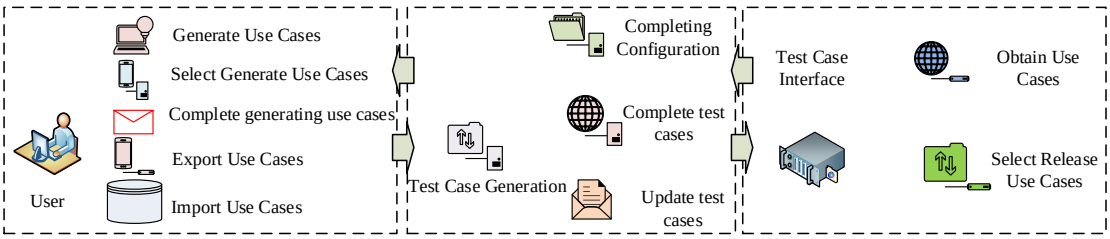


Figure 5: Example of model dynamic analysis

Description: Illustrates how the user analyses the system software or model according to the object model in the testing phase and completes the dynamic model construction

Figure 5 shows that in the testing phase, the user is required to analyze the system software or model in the component according to the object model, and input the information into the parameters of the test case after completing the analysis. Then, the operation is performed after receiving the analysis information of the parameters, including the combination of settings such as choosing the value of the parameter at the moment and setting the current moment and constraints, etc., and then completing the dynamic model. Construction of the dynamic model. Dynamic analysis is aimed at more accurately simulating and evaluating the behavior and performance of software in actual operating environments. The CCTL method generates test cases through time constraints, while dynamic analysis verifies the effectiveness and coverage of these test cases at runtime; PSO and GA algorithms optimize the generation of test cases, and dynamic analysis further verifies the actual effectiveness of these optimized test cases. Simultaneously, in the test, the user needs to deploy the front information in advance as shown in Figure 6 for the parameter sequence image.

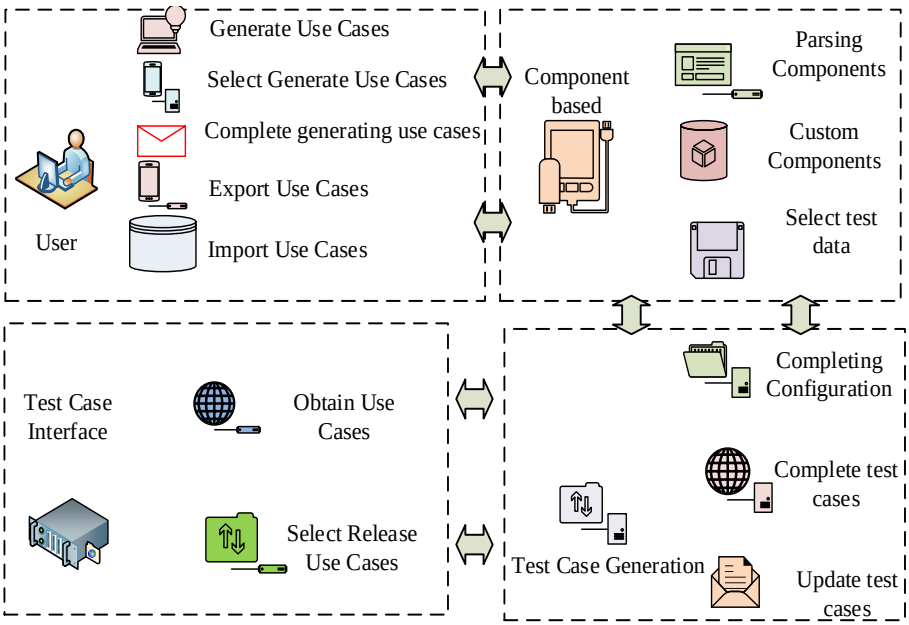


Figure 6: Example of deployment pre parameter sequence process

Description: Describe how to generate test case information before deployment and generate test cases through the CCTL method and algorithm process

As shown in Figure 6, after the deployment of the predecessor information, the test case information is generated, after receiving the generation signal, the predecessor information is read and analyzed to obtain the generated parameter value data and moments, constraints, etc., and the test cases are produced through the CCTL method and algorithmic process, and the user is reminded of the completion of the generation after the end of the generation, and the data information is then exported to be used for other operations, and the test data needs to be constantly updated when the test cases are generated. Parameter selection is based on the time constraints of test case generation requirements and CCTL methods, optimized through PSO and GA algorithms to ensure that test cases can cover all key scenarios. Pre information refers to the initial conditions and contextual information before the generation of test cases, providing necessary background and initial conditions for the generation of test cases. The pre deployment parameter sequence process is an important part of the testing process, which directly affects the generation of test cases and the accuracy of test results. Simultaneously, it is necessary to updat the use cases when generating test data. The entire software testing system platform is shown in Figure 7.

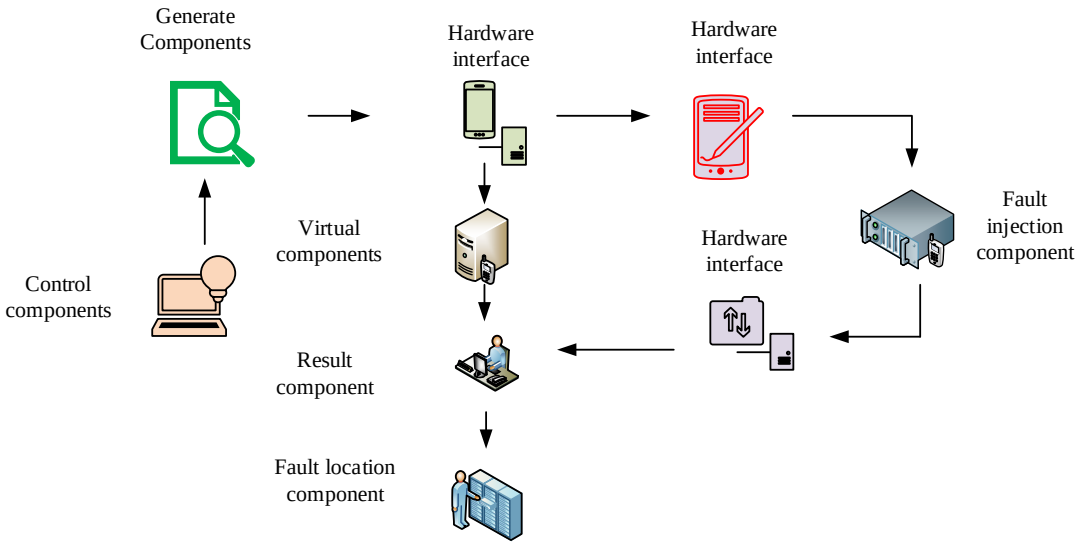


Figure 7: System platform for software testing

Description: Outline how the test system through the CCTL method and particle swarm algorithm genetic algorithm to generate and distribute software test case information

As shown in Figure 7, the main idea in the testing system is to generate and distribute the use case information for software testing through the main idea of the CCTL method and the particle swarm algorithm genetic algorithm. Therefore, in the general framework of the system, the test system includes a test case, hardware interface, virtual comparison model, result comparison and fault location components. The simulation to be tested included a hardware interface component and a fault injection component. The control component is responsible for controlling the testing process and execution. Generate components that can use different generated test cases. Virtual components can create tested system components through virtualization methods. The hardware interface can connect the generated test cases and virtual components with the actual hardware through the hardware interface. The result component is capable of collecting and processing test results for comparison. The fault location component can locate and analyze faults in the system. The hardware interface will conduct hardware level testing through the hardware interface. The fault injection component is mainly used to introduce faults into the system, verify the system's fault handling capability and robustness. The fault detection and analysis component is responsible for analyzing the test result data collected during the testing execution phase, comparing the actual test results with the expected results, and detecting whether there are deviations or abnormal behaviors. The fault location component uses parameters generated by the CCTL method, and utilizes these parameters and constraints to more accurately locate the time and location of the fault occurrence. The fault injection component is responsible for introducing predefined faults during the testing process, verifying the system's fault handling capability and stability, and injecting faults at specific time points and conditions based on the parameters generated by the CCTL method. During software testing, control components to generate test cases and distribute them to the generating components. Generate test cases using CCTL method and PSO/GA algorithms, and transmit them to virtual components. Virtual components interact with actual hardware through hardware interfaces and apply test cases to the system under test. During the testing process, the result component collects and processes the test results. The test results are transmitted to the fault location component for fault analysis and localization. Finally, the hardware interface is also used to interact with the fault injection component, introducing faults into the system to verify its stability and robustness. The higher the error detection rate, the more defects are discovered in the test case set, and the better the testing effect. The calculation formula is error detection rate = (number of discovered defects/total number of introduced defects) × 100%. The criteria for determining the coverage of a model are its functional coverage, with a functional coverage of ≥90%, a state coverage of ≥85%, a transition coverage of ≥80%, and a path coverage of 75%. These all indicate that the current model has good coverage. However, the study only uses the functional coverage of the model as the criterion for judgment, so a functional coverage rate of ≥90% is considered to be better for the current model. Functional coverage can directly reflect the extent to which test cases cover software functionality, and high coverage can help identify potential defects. A functional coverage rate of ≥ 90% can significantly improve the accuracy and reliability of testing, covering most key functions and ensuring the practicality and reference value of research results. The test case generation module generates test cases that meet time constraints and functional coverage requirements based on CCTL method and PSO/GA algorithm. The virtual comparison model simulates the actual operating environment through virtualization technology, executes test cases, and dynamically adjusts configurations. Compare the accuracy and completeness of the verification test results in the module, and generate a test report. Fault location module analysis report, accurately locate the fault. The hardware interface module and fault injection module respectively support the execution of test cases on actual hardware and fault injection, ensuring system stability and reliability.

4 Results and discussion

4.1 Results

Research the use of an online shopping system program for software testing. Firstly, the system needs to verify user login, product search, shopping cart functionality, and order payment. Next, configure the web server and database server, and set up development, testing, and production environments. Write test cases for user login, product search, addition, deletion, quantity modification, and order payment of shopping cart items. At the same time, the software uses PSO and GA to generate multiple input combinations during testing, ensuring coverage of all testing scenarios. And use automation tools such as Selenium to perform functional testing of the web interface, use JMeter for performance testing, and use OWASP ZAP for security testing. Finally, record the execution status of all test cases and analyze the failed cases. Use Selenium for automated testing of user interfaces, as it supports multiple browsers and operating systems and can simulate real user operations. Use JMeter for performance testing as it can simulate high concurrency scenarios and evaluate the system's response time and throughput. The population size is set to 30, the maximum number of iterations is 50, the learning factors are 2.05, and the inertia weight is 0.9. For the GA population size of 100, the maximum number of iterations is 20, the crossover rate is 0.8, and the mutation rate is 0.01. The experimental environment for the research includes the use of Intel Core i7-9700 CPU, 32GB RAM, and Windows 10 Professional 64 bit operating system, developed using MATLAB R2020b and Python 3.8, and automated testing, performance evaluation, and security checks implemented using tools such as Selenium, JMeter, and OWASP ZAP. The testing platform takes an online shopping system as an example, covering functional modules such as user login, product search, shopping cart operation, and order payment. The evaluation indicators

include functional coverage (≥ 90%), state coverage (≥ 85%), transition coverage (≥ 80%), and path coverage (≥ 75%), as well as error detection rate, algorithm performance, and stability. The experimental design covers the complexity stratification of test cases, parameter settings for PSO and GA, number of repeated experiments, and verifies the significance of the results through charts and statistical analysis. Generate a test report, record the discovered defects, and ultimately complete software testing. When conducting software testing, it is necessary to ensure the stability and reliability of the system under various input conditions. Simultaneously, it is necessary to include functional testing, performance testing, and security testing. And use automated testing tools to perform testing. Regularly conduct code reviews and test case reviews, and use static analysis tools to check code quality. Randomly select 50 software test data from the currently selected ones for testing. The selection of 50 test data is based on statistical sample size calculation, ensuring that the error range is within 10% at a 95% confidence level. This sample size can meet the testing requirements while also being within resource constraints. Complexity refers to the comprehensive difficulty of test cases in terms of functional coverage, operational steps, data input, and expected results. According to the complexity of the test cases, they are divided into three layers: simple, medium, and complex. Each layer selects 20%, 50%, and 30% of the test cases. The goal of this test is to verify the performance and stability of the system under high load conditions. The testing scope includes the login module, data processing module, and report generation module. The main method of calculating the coverage is obtained by comparing the actual number of covered combinations with the target number of combinations, so that in the coverage test using CCTL is obtained as shown in Table 1. The target coverage combination refers to the coverage of all possible parameter combinations and functional modules, while the actual coverage combination refers to the parameter combinations and functional modules actually covered through test cases. Use case coverage focuses more on the coverage of user scenarios and interactions, focusing on how users use the software. Functionality coverage focuses more on the coverage of the software's functionality, and is concerned with the specific operations that the software can perform. The software being tested refers to the user login module of the online shopping system. The "CA (33, 2; 2)" in Table 1 indicates that under three constraint conditions, the target combination quantity is 80, and the actual number of combinations covered is 78.

Table 1: Analysis of CCTL method testing software coverage

Software under testing	Number of constraints	Number of target combinations	Number of target combinations after unconstrained	Constrain the number of target combinations before validation
CA (33, 2; 2)	3	80	78	78
CA (33, 2; 2)	6	125	135	135
CA (53, 2; 2)	6	224	224	224
CA (53, 2; 2)	3	600	456	456
CA (83, 2; 2)	2	900	894	894
CA (83, 2; 2)	10	1654	1645	1645

Note: Shows the number of combinations of objectives under different number of constraints, etc. CA stands for 'Clock-Controlled Computation Tree Logic Algorithm'

As shown in Table 1, when the coverage rate of software testing is analyzed, it is found that the target array and the number of constraints after removing the constraints have the same value as the number of coverage combinations verified, which is visible of the software testing. The use of the method is able to achieve 100% coverage, which shows that the current method of testing the different software is able to achieve the number of constraints and time of the selection of software constraints; at the same time, the high coverage rate indicates that the method of testing is better. The similarity between the number of unconstrained combinations and the number of combinations before validation is due to the fact that the test case generation algorithm has already considered most of the constraint conditions when generating test cases. After removing the constraints, if the actual number of generated test case combinations does not change significantly, then the test coverage rate can reach 100%. To compare the generation time and data size of the current system method for software testing, the number of populations and the number of iterations were set to 150 populations and 20 iterations, respectively, as shown in Figure 8. When generating test cases, the average time for particle swarm optimization algorithm to generate test cases is longer than that of genetic algorithm. Therefore, it is said that the test cases generated by the particle swarm algorithm require a longer period of time.

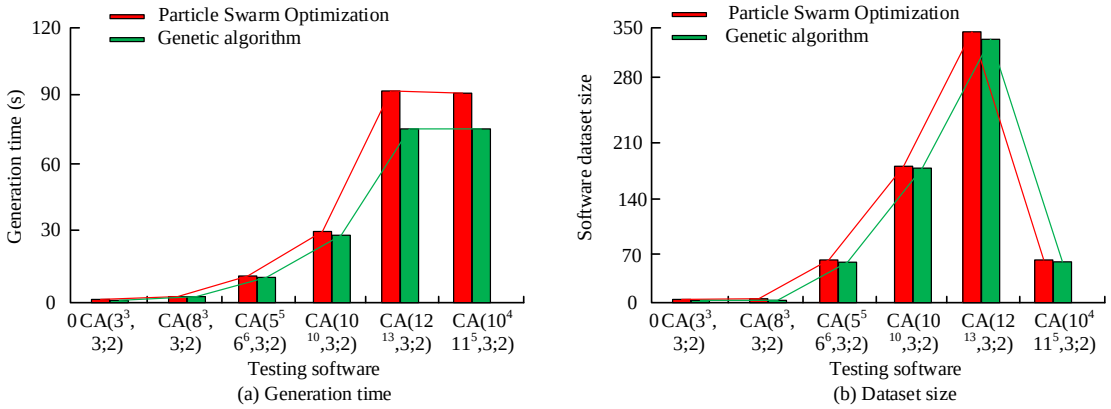


Figure 8: Comparison of testing time and data scale for different system software

Description: Compare the current system approach with other algorithms in terms of software test generation time and data size

Figure 8(a) shows that in the analysis of the CCTL method of the two algorithms for software testing generation time is different from the genetic algorithm method in the generation of the moment of generation of the event is significantly higher than the particle swarm algorithm, which indicates that in the generation of software data processing particle swarm algorithm of the generation of a longer period of time to test the software is significantly faster than the genetic algorithm, as can be seen from Figure 8(b) particle swarm algorithm of the From Figure 8(b), it can be seen that the particle swarm algorithm is significantly higher than the genetic algorithm in generating the number of coverage combinations, which indicates that the size of the data generated by it is larger, and the CCTL method is more important to generate the data coverage combinations, while the genetic algorithm is more important to generate and analyse the time moments. To test the effect of the current build system test software, the software parameter position information, speed information, analog, and other parameters to test, as shown in Table 2. The target parameter speed reflects the speed at which the software processes data or tasks, while the azimuth velocity reflects the efficiency of the system in multi-dimensional task scheduling. These parameters directly affect testing efficiency and accuracy. Higher speeds and azimuth velocities can improve testing efficiency. At the same time, by precisely controlling the parameter range, it can better simulate actual scenarios and improve the accuracy of test results. The maximum and minimum values of the target distance are based on the possible maximum and minimum shopping cart distances in the system. The injection time is the time interval used to simulate actual user operations, ensuring that test cases can cover system behavior at different time points.

Table 2: Test results of parameter position information, speed information, and analog parameters

Attribute name	Maximum value	Minimum value	Injection time
Target distance (m)	10000	0	[0 ms, 100 ms]
Target coordinates X (m)	10000	0	[0 ms, 100 ms]
Target coordinates Z (m)	10000	0	[0 ms, 100 ms]
Target coordinates Y (m)	10000	0	[0 ms, 100 ms]
Target direction speed (rad/s)	20	0	[150 ms, 250 ms]
Target speed (m/s)	80	0	[150 ms, 250 ms]
A1	12	0	[0 ms, 10 ms]
A2	12	0	[0 ms, 10 ms]
B1	15	0	[30 ms, 40 ms]
B2	15	0	[30 ms, 40 ms]
C1	28.5	0	[50 ms, 60 ms]
C2	28.5	0	[50 ms, 60 ms]

Note: Lists the software parameter information in the test

Table 2 shows that the size of the maximum and minimum values in the target position of the parameter are the same, the maximum value is 10000 m, the minimum value is 0 m, and the range of the injected moments of the position are in [0 ms, 100 ms]. The range of the moment of the position information of the particle parameter are in

[150 ms, 250 ms], the maximum value of the velocity of the target parameter is 80 m/s, and the minimum value is 0 m/s. The maximum value of the azimuthal velocity of the target is 20 rad/s, and the minimum value is 0 rad/s. The maximum value of the analogue of the software test and the different analogues are different but the minimum value is 0, and at the same time, their moment ranges are not the same. This shows that when the moment and parameter determinations are made for the software, the values of the parameters are different but some of the parameters have the same range of values. In order to verify the effectiveness of the current CCTL method for software testing, the test cases are set to 4700, with 100% coverage, using 46 parameters for analysis, and the system is analysed and tested using the big data platform, obtaining a partial test result graph as shown in Table 3. Safety value is an indicator to assess the performance and stability of software during testing, which is used to determine whether there is any potential security risk in the software. Fault information location is used to determine the exact location and cause of faults in the software by analysing the test results, helping developers to quickly diagnose and fix problems.

Table 3: Software test data results

Combination test data results
2100 16.907 14.25 2100 16.907 14.25 2100 16.907 14.25
2101 16.574 14.25 2101 16.574 14.25 2101 16.574 14.25
2102 14.00 15.75 2102 14.00 15.75 2102 14.00 15.75
2103 14.00 0.75 2103 14.00 0.75 2103 14.00 0.75
2104 17.702 7.291 2104 17.702 7.291 2104 17.702 7.291
2105 24 -5.162 2105 24 -5.162 2105 24 -5.162
2106 17.707 5.52 2106 17.707 5.52 2106 17.707 5.52
2107 17.207 14.25 2107 17.207 14.25 2107 17.207 14.25
2104 24 - 2104 24 - 2104 24 -
2104 14.25 - 2104 14.25 - 2104 14.25 -
2107 17.207 12 2107 17.207 12 2107 17.207 12
2103 3.5345 6.471 2103 3.5345 6.471 2103 3.5345 6.471
2107 17.207 -0.75 2107 17.207 -0.75 2107 17.207 -0.75
2108 2 -15.75 2108 2 -15.75 2108 2 -15.75
2109 10.327 0 2109 10.327 0 2109 10.327 0

Note: Provides the results of combined test data, showing the test results of different combinations of test cases

As shown in Table 3, the results of the software testing in the simulation test and fault results of the analysis and testing, and then in the results of the test to obtain the safety value, obtain the safety value of the test software fault information positioning, the test fault positioning in the results of the analysis of the region to display, and then the positioning results obtained and the simulation of the data parameters for the combination of the results of the positioning and fault results for the combination of the analysis of the results of the conclusions obtained were used to reflect the test software use cases and parameters of the direct combination of coverage and software testing. The above test results show that the current method can be tested and analyzed on the parameter combinations, and therefore verify the feasibility of the software testing method. Table 3 shows the actual test data results, which are generated based on the parameters defined in Table 2. Each set of data in Table 3 corresponds to a different combination and value of the parameters defined in Table 2. For example, the data set 2100 may indicate that for a particular combination of parameters, the test result is 16.907, 14.25, etc. The data set 2101 corresponds to the test result for another combination of parameters, and so on. The first set of data represents the execution results of the test cases. The second set of data represents performance indicators such as response time, throughput, etc. The third set of data represents the results of safety or fault detection. Therefore, the data is divided into three groups. In analysing the test results, the data were first pre-processed and then statistical metrics such as mean, standard deviation and variance were calculated for each algorithm. And key performance indicators were also calculated and finally visualised and analysed using bar charts and line graphs. In comparing the test results of different algorithms, LSTM and LAR are chosen as benchmark algorithms. To ensure the fairness of the comparison, all algorithms are tested on the same test dataset and test environment. In addition, the study verified whether the differences between the different algorithms were statistically significant using statistical methods such as t-test and ANOVA. To test the accuracy of the test parameters of the current method, the test results of the parameters were analyzed and compared with the test results of other algorithms, namely, long short-term neural networks (LSTM), and logistic regression algorithm (LRA) as shown in Figure 9. LSTM can help identify and predict potential issues in software testing by analyzing system logs, user behavior data, and historical

testing data, thereby improving testing coverage and efficiency. LSTM and LRA were chosen as comparison algorithms for the study because LSTM has a significant advantage in dealing with time series data by capturing long-term dependencies in the data. LRA was chosen because the algorithms provide a simple linear benchmark that helps to understand the underlying linear trends in the test data. LSTM is effective in analysing parameter variations in the test, while LRA provides a simple linear benchmark that helps to understand the underlying linear trends in the parameters.

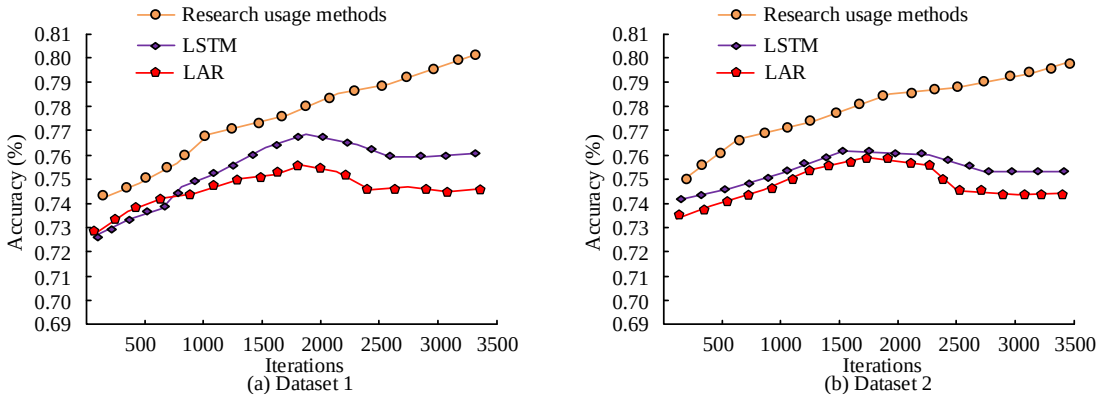


Figure 9: Comparison of software testing accuracy with different algorithm logics added
Description: Demonstrates the change in algorithm accuracy after adding different algorithm logic

As shown in Figure 9, after adding different algorithmic logic to the CCTL, the accuracy of the LSTM and LAR algorithms first increases with the number of iterations and then tends to stabilize, while the accuracy of CCTL combines PSO and GA approaches is in the increasing stage. Accuracy continues to increase over time, showing better learning ability and adaptability. This indicates that the research usage method is more effective in testing the accuracy of software, while the other two algorithms have the highest accuracy, at 76% and 75% respectively. The accuracy of the research usage method is higher, at 5% and 6% higher than the LSTM and LAR methods, respectively. To test the algorithmic stability of the proposed method, the loss functions of the three algorithms were compared and tested, as shown in Figure 10.

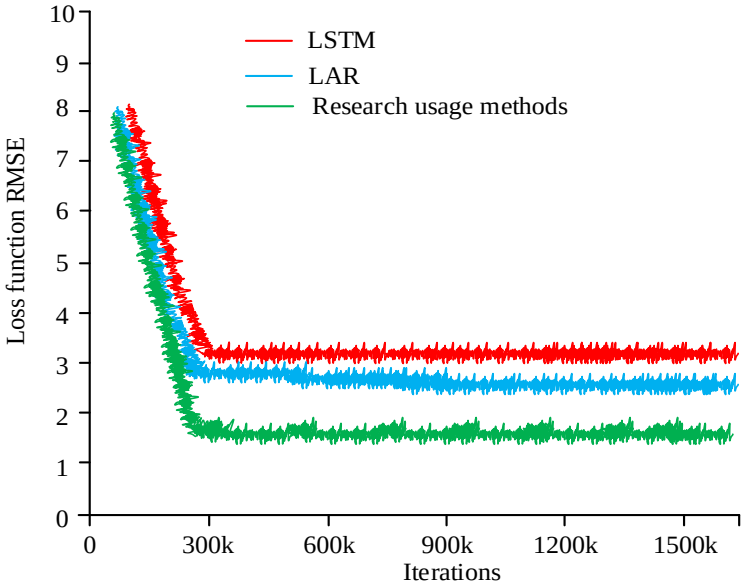


Figure 10: Comparison of loss functions among three algorithms
Description: compares the change in loss function of the three algorithms during the iteration process

Figure 10 shows that the loss function of the three algorithms decreases and then gradually stabilizes throughout the algorithm as the number of iterations increases. The loss function value of the research use method dropped to a minimum of 1.5 loss function at 300 k iterations, the LSTM algorithm dropped to a loss function of 3 at 300 k, and the LAR algorithm dropped to a minimum loss function of 3.2 at 320 k iterations. Loss function is an important tool in machine learning to measure the prediction error of a model, the smaller its value the smaller and more stable the model error. The research use model loss function value drops to 1.5 after

300,000 iterations and outperforms other models, which indicates that the research use model learns data features better during training, makes more accurate predictions, and performs better when dealing with complex data. This shows that the algorithmic logic of the research method has the smallest loss function value, and the algorithmic model is more stable. To test the effectiveness of different algorithm models in software testing, a comparative analysis was conducted on the Pairwise Testing (PT) model, Ant Colony Optimization (ACO), Reinforcement Learning Algorithm (RLA), and Particle Swarm Optimization (PSO) algorithm models, as shown in Table 4.

Table 4: Comparison of test results of different algorithm software

/	Test result					
Model	ACO	PT	PSO	GA	RLA	CCTL
Error detection rate (%)	94.35	92.65	91.54	91.58	92.35	96.54
Functional coverage (%)	92.61	91.26	90.36	91.32	92.67	94.85
Status coverage rate (%)	88.67	89.65	93.51	91.54	89.35	94.68
Conversion coverage (%)	81.26	90.35	92.54	84.65	83.51	93.59
Path coverage (%)	82.36	83.65	86.57	89.65	90.13	91.35

Note: Compares the performance of different algorithm models in software testing

From Table 4, it can be seen that after using different algorithms for software testing, it is evident that the CCTL model performs better in terms of testing results. In the comparison of software testing performance, the error detection rate of the CCTL model reached the highest level of 96.54%, which is about 5.00% higher than that of the PSO model. In the comparison of algorithm coverage, the CCTL model achieved good coverage levels under different coverage tests, with the model reaching the highest value of 94.85% in functional coverage analysis. It can be seen that the coverage and error detection rate of the model used in the study have a high level, and the actual software testing effect of the model is good. Unified testing environment, selection of unified performance indicators, and data alignment for performance research of different algorithms that have not been tested. Statistical tests show that the CCTL model performs the best on all indicators. Error detection rate is a key indicator for measuring the performance of software testing models, reflecting the model's ability to detect defects. A high error detection rate means that the model can more effectively detect software defects, thereby improving software quality and reliability. The high error detection rate of CCTL model indicates that CCTL combined with PSO and GA can generate high coverage test cases more comprehensively, effectively reducing software failure risks and improving user satisfaction. When comparing algorithm coverage, the CCTL model performed well in the function, state, transition and path coverage tests. This indicates that the CCTL model, by introducing time constraints and optimising parameter combinations, can generate high-coverage test cases more comprehensively and effectively improve the comprehensiveness and accuracy of software testing. This indicates that the CCTL model combined with PSO and GA can more effectively generate high coverage test cases, resulting in better performance in detecting defects and covering software features. Table 5 shows a summary comparison of the test results of the different methods.

Table 5: Summary comparison of tests for different methods of testing

Method	Accuracy	Efficiency	Coverage Rate	Capability to Handle Large Datasets	Real-Time Parameter Constraint Handling Capability
Bat Search Algorithm	72.35%	Moderate	79.84%	Limited	Weaker
Random Forest	88.67%	High	88.67%	Better	Average
PSO	91.32%	High	90.36%	Better	Strong
GA	90.36%	Moderate	90.47%	Average	Strong
Proposed Method (CCTL+PSO/GA)	5%-6% higher than PSO/GA	High	100%	Strong	Very Strong

Note: Comparison of test results of different methods

As can be seen from Table 5, the CCTL+PSO/GA method used in the research is significantly better than other methods in different test results, which has the strongest ability to deal with large-scale datasets, and the best ability to deal with real-time parameter constraints, which suggests that the model used in the research has a better practical application in the testing of different methods.

4.2 Discussion

In Raamish et al.'s study, in order to test the quality and performance of the software and improve its reliability, a ramp up algorithm based on LSTM and BrainStorm optimization and post acceptance was used. The new algorithm can be used for software fault detection. Compared with traditional methods, the new method can effectively improve the effectiveness of software detection [27]. However, this method only analyzes and detects faults and defects detected by software, and cannot improve the performance and stability of software detection. Oleshchenko's research found that software testing can consume a significant amount of cost and time during software development. However, using the KNN algorithm to train and test software development data can greatly reduce software testing time and cost. From this, it can be seen that using clustering algorithms for software data analysis and cost control is an important direction in the software testing process [28]. Build a new model for software testing in this study. After 300,000 iterations, the PSO algorithm generates a significantly higher number of test case coverage combinations than the GA algorithm. The PSO algorithm achieves a functional coverage of 90.36% compared to the GA algorithm's 91.32%.The PSO algorithm achieves an error detection rate of 91.54%, which is slightly lower than that of the GA algorithm's 91.58%.The CCTL method is able to, through the introduction of temporal constraints more accurately simulate the temporal characteristics of the software in actual operation. Meanwhile, the CCTL method is better at generating high-coverage test cases after combining the PSO and GA algorithms. When comparing the software production time of models, particle swarm optimization algorithm has a higher coverage combination than genetic algorithm in testing the performance of models with different iteration times. This indicates that particle swarm optimization algorithm has better performance in data coverage combination processing and higher performance in improving the coverage combination set of the model. Compared with KNN, CCTL is more efficient in dealing with complex time series data; compared with LSTM, CCTL is more flexible in dealing with nonlinear data; compared with BrainStorm optimisation, CCTL is more comprehensive in dealing with complex test scenarios. These advantages enable the CCTL method to demonstrate higher accuracy and efficiency in software performance testing, while reducing cost and improving software reliability. This may be because particle swarm optimization algorithms are easier to integrate data. When comparing the accuracy of different algorithm models, the study found that the accuracy of LSTM and LAR models were 5% and 6%, respectively. This may be due to the combination of genetic algorithm and particle swarm optimization algorithm used in the study. The types of constraints selected for research are closely related to the actual situation, aiming to reflect the various scenarios that software testing may encounter and ensure that test cases cover all key operations and interactions of the software. Although the experimental results showed high test coverage and accuracy, some test cases failed due to parameter configuration errors, environmental dependencies, resource limitations, concurrency conflicts, algorithm limitations, or insufficient test data. This indicates that the test cases further reveal the shortcomings of the testing strategy and algorithm performance, and points out the necessity of enhancing the robustness of the test cases. In software performance testing, PSO and GA optimization algorithms significantly improve the comprehensiveness and accuracy of testing by generating high coverage test cases. Algorithms not only improve testing efficiency and reduce resource consumption, but also lower testing costs. In addition, high-quality test cases generated by algorithm models can effectively detect software defects, improve software reliability and stability.

When comparing the loss functions of different models, the changes in the loss functions of the models show a trend of first decreasing and then approaching equilibrium. This may be because as the number of iterations increases, the functional loss of the model also increases, but when the model reaches a certain value, it begins to stabilize. The loss function value of the model used in the study is relatively small, which may be due to the improved algorithm performance after adding different algorithms to the model. When comparing the performance of different models, it was found that the algorithm using this model performed better, with the highest error detection rate of 96.54%. Compared with other algorithms, the algorithm used in this study has higher coverage and error rates. This may be due to the current algorithm model performing better in software testing. Has shown certain efficiency and accuracy in selecting regression test cases, but has limitations in handling large-scale datasets and real-time parameter constraints. The current research method uses PSO and GA optimization algorithm logic to significantly improve test coverage and outperforms bat search algorithm in terms of functional coverage. The random forest algorithm performs the best in predictive performance, but its real-time parameter processing ability in software testing is average. The method currently used in the research not only has a 5% -6% higher accuracy, but also demonstrates a very strong ability to handle real-time parameter limitations. PSO is faster in some cases, mainly because it can quickly find the optimal solution through the collaboration of individuals and groups. Meanwhile, GA simulates the genetic and mutation mechanisms in natural selection, making it suitable for global search, but its convergence speed may not be as fast as PSO. By combining the advantages of PSO and GA, the study not only improved the coverage and accuracy of testing, but also maintained efficient testing performance in environments with real-time parameter changes. It also provides new ideas and tools for software testing in more complex environments in the future. In future research, researchers may further explore how to optimize PSO and GA algorithms to improve performance in specific situations. At the same time, the proposed methods may stimulate the development of new testing methods, and with the improvement of algorithm performance, future software testing may become more automated and

intelligent, thereby improving testing efficiency and reducing costs.

In summary, in the comparison of different algorithm models, it is found that the use of models has better software testing performance. At the same time, the use of algorithms in software testing can effectively analyze and test software, and its effect can reach

5 Conclusion

This research mainly focuses on the current problem of lack of stability and performance of software testing, and proposes a new software testing system based on the CCTL method, first analyzing the component use case generation of software testing. Subsequently, the system is transformed and analyzed using the particle swarm algorithm and genetic algorithm logic, and a system model is built for software testing. The method used in the study is based on CCTL combined with PSO and GA. The loss function value of the method is significantly lower than other methods, indicating that it can better learn data features during the training process and make more accurate predictions. The new methodology achieves 100 per cent test coverage. Meanwhile, the new method effectively selects the number of constraints. In terms of algorithm performance, GA excels in generating test moments, while PSO is more advantageous in handling complex parameter combinations and large-scale data. These advantages enable the new method to excel in generating high-coverage test cases, significantly improving the comprehensiveness and accuracy of the tests. During the testing process, parameter settings significantly affect the generation of test cases and test results. The range of values of the algorithm parameters has a significant effect on the test results. For example, an increase in the number of particles and population size can enhance the search ability but increase the running time; an increase in the learning factor and crossover rate can make the search direction clearer but may fall into the local optimum. By reasonably adjusting these parameters, the speed and coverage of test case generation can be optimised and the test results can be improved. In addition, the effect of different parameter settings on the testing effect showed that the parameter values were different, but the range of values is similar. Simultaneously, the new method can effectively obtain the safety value and fault location information of the software. Meanwhile, the accuracy of other traditional methods is as high as 76% and 75%, whereas the accuracy of the research use method is higher, outperforming the accuracy of the LSTM and LAR methods by 5% and 6% respectively. Also the loss function of the algorithm used in the study is 1.7 and 1.5 lower than the loss functions of the other two algorithms, which shows that the method is more stable. From this, it can be seen that researching usage methods has better testing effects in software testing, and at the same time, studying the testing error rate, coverage rate, and accuracy of using models in different model comparisons has a high level. It can be seen that although this study has achieved a lot of results, it still needs to be improved, first of all, the algorithm needs to be further improved subsequently when it is constrained to the combination of cases, and the data used for the study was small, so larger data sets will need to be analysed in subsequent studies. The current research is mainly conducted in specific testing scenarios and may not fully cover all complex situations in practical applications. Therefore, further research is needed to expand the testing scenarios to verify the universality of the method. Although the methods used in the study have shown improvements in accuracy and stability compared to other methods, their stability and accuracy still need further validation in larger scale data and more diverse testing environments. The PSO and GA algorithms in the current study have limitations although they perform well. Therefore, future research will improve these algorithms by introducing hybrid strategies, dynamically tuning the parameters and optimising the cross-variance operation. The current research focuses on specific scenarios such as online shopping system, data processing module and report generation module. Future work will expand the test scenarios to validate the stability and accuracy of the new methods using larger scale data and more diverse test environments.

Data availability statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that this article is free of any conflicts of interest.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yuan Sun, Md Gapar Md Johar, Jacquline Tham. The first draft of the manuscript was written by Yuan Sun, Md Gapar Md Johar, Jacquline Tham. All authors read and approved the final manuscript.

Funding

No funding was received.

References

[1] Alhroob A, Alzyadat W, Imam A T, Jaradat Ghaith M. The genetic algorithm and binary search technique in the program path coverage for improving software testing using big data. Intelligent Automation and Soft Computing, 2020, 26(4): 725-733. http://dx.doi.org/10.32604/iasc.2020.010106

[2] Han X, Yu T, Yan G. A systematic mapping study of software performance research. Software: Practice and Experience, 2023, 53(5): 1249-1270. https://doi.org/10.1002/spe.3185

[3] Murthy M S N, Suma V, Chandrappa C N, Shankar M M. Factors influencing effectiveness of testing applications in cloud using regression testing: a

statistical analysis. International Journal of Advanced Intelligence Paradigms, 2020, 17(1/2): 109-126. https://doi.org/10.1504/ijaip.2020.108770

[4] Guo S, Wang J, Xu Z, Huang L. Feature transfer learning by reinforcement learning for detecting software defects. Software: Practice and Experience, 2023, 53(2): 366-389. https://doi.org/10.1002/spe.3152

[5] Chen J. Construction and Application of an Economic Intelligent Decision-making Platform Based on Artificial Intelligence Technology. Informatica. 2024, 48(9). https://doi.org/10.31449/inf.v48i9.5705

[6] Kaur A, Agrawal A P. Performance comparison of Bat search and Cuckoo search using software artefact infrastructure repository and regression testing. International Journal of Advanced Intelligence Paradigms, 2021, 18(2): 99-118. https://doi.org/10.1504/IJAIP.2021.112899

[7] Chen J, Hu H, Yu D. Characterising and detecting methods to be benchmarked under performance unit test. International Journal of Software Engineering and Knowledge Engineering, 2022, 32(9): 1279-1305. https://doi.org/10.1142/S0218194022500486

[8] Khurshid S, Iqbal J, Malik I A, Yousuf B. Modelling of NHPP based software reliability growth model from the perspective of testing coverage, error propagation and fault withdrawal efficiency. International Journal of Reliability, Quality and Safety Engineering, 2022, 29(6): 10-28. https://doi.org/10.1142/s0218539322500139

[9] Qian J, Zhou X, Zhou H. Prioritising test scripts for the testing of memory bloat in web applications. IET Software, 2022, 16(3): 317-330. https://doi.org/10.1049/sfw2.12057

[10] Bugden W, Alahmar A. The safety and performance of prominent programming languages. International Journal of Software Engineering and Knowledge Engineering, 2022, 32(5): 713-744. https://doi.org/10.1142/S0218194022500231

[11] Jiang M, Chen T Y, Wang S. On the effectiveness of testing sentiment analysis systems with metamorphic testing. Information and Software Technology Information and Software Technology, 2022, 150(10): 1-11. https://doi.org/10.1016/j.infsof.2022.106966

[12] Hao T, Elith J, Lahoz-Monfort J, Guillera-Arroita G. Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. Ecography, 2020, 43(4): 549-558. https://doi.org/10.1111/ecog.04890

[13] Hosseini S M J, Arasteh B, Isazadeh A, Mohsenzadeh M, Mirzarezaee M. An error-propagation aware method to reduce the software mutation cost using genetic algorithm. Data Technologies and Applications, 2021, 55(1): 118-148. https://doi.org/10.1108/DTA-03-2020-0073

[14] Zeb A, Din F, Fayaz M, Mehmood G, Zamli K Z. A systematic literature review on robust swarm intelligence algorithms in search-based software engineering. Complexity, 2023, 2023(1): 4577581-4577582. https://doi.org/10.1155/2023/4577581

[15] Pan R, Ghaleb T A, Briand L. Atm: Black-box test case minimization based on test code similarity and evolutionary search. In 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE), 2023, 14(5): 1700-1711. https://doi.org/10.48550/arXiv.2210.16269

[16] Omar H K, Frikha M, Jumaa A K. Improving Big data recommendation system performance using NLP techniques with multi attributes. Informatica. 2024, 48(5). https://doi.org/10.31449/inf.v48i5.5255

[17] Sornkliang W, Phetkaew T. Performance analysis of test path generation techniques based on complex activity diagrams. Slovenian Association Informatika, 2021, 45(2): 231-242. https://doi.org/10.31449/inf.v45i2.3049

[18] Bednárek D, Kruliš M, Yaghob J. Letting future programmers experience performance-related tasks. Journal of Parallel and Distributed Computing, 2021, 155(C): 74-86. https://doi.org/10.1016/j.jpdc.2021.04.014

[19] Ke S Z, Huang C Y. Software reliability prediction and management: a multiple change-point model approach. Quality and Reliability Engineering International, 2020, 36(5): 1678-1707. https://doi.org/10.1002/qre.2653

[20] Bi W, Yu F, Cao N. Wei H, Cao G, Han X, Sun L, Higgs R. Research on data extraction and analysis of software defect in IoT communication software. Computers, Materials, and Continuum, 2020, 65(2): 1837- 1854. https://doi.org/10.32604/cmc.2020.010420

[21] Li L, Xu L, Cui H, Abdelkareem M A A. Validation and optimization of suspension design for testing platform vehicle. Shock and Vibration, 2021, 2021(7): 1-15. https://doi.org/10.1155/2021/7963517

[22] Shao Y, Liu B, Wang S, Xiao P. A novel test case prioritization method based on problems of numerical software code statement defect prediction. Eksploatacja i Niezawodnosc - Maintenance and Reliability, 2020, 22(3): 419-431. https://doi.org/10.17531/ein.2020.3.4

[23] Serat Z, Fatemi S A Z, Shirzad S. Design and economic analysis of on-grid solar rooftop PV system using PVsyst software. Archives of Advanced Engineering Science, 2023, 1(1): 63-76. http://dx.doi.org/10.47852/bonviewAAES32021177

[24] Garmaki M, Gharib R K, Boughzala I. Big data analytics capability and contribution to firm performance: the mediating effect of organizational learning on firm performance. Journal of Enterprise Information Management, 2023, 36(5): 1161-1184. https://doi.org/10.1108/JEIM-06-2021-0247

[25] Jalil S, Rafi S, LaToza T D, Moran K, Lam W. Chatgpt and software testing education: Promises & perils. In2023 IEEE international conference on software testing, verification and validation workshops (ICSTW), 2023, 16(6): 4130-4137. https://doi.org/10.48550/arXiv.2302.03287

[26] Weber M, Kaltenecker C, Sattler F, Apel S, Siegmund N. Twins or false friends? a study on energy consumption and performance of configurable

software. In 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE), 2023, 14(5): 2098-2110. https://doi.org/10.1109/ICSE48619.2023.00177

[27] Raamesh L, Jothi S, Radhika S. Enhancing software reliability and fault detection using hybrid brainstorm optimization-based LSTM model. IETE Journal of Research, 2023, 69(12): 8789-8803. https://doi.org/10.1080/03772063.2022.2069603

[28] Oleshchenko L. Software testing errors classification method using clustering algorithms. International Conference on Innovative Computing and Communication, 2023, 17(10): 553-566. https://doi.org/10.1007/978-981-99-3315-0_42

Appendix

Item	Details
Experimental Purpose	To evaluate the performance of the software performance testing system based on Clock-Controlled Computation Tree Logic (CCTL)
Experimental Environment	Intel Core i7-9700 CPU @ 3.20GHz, 32GB RAM, Windows 10 Professional 64-bit
Software Tools	MATLAB R2020b, Python 3.8 with necessary libraries
Experimental Steps	Setup environment, write test cases, apply PSO and GA, use automation tools, record results, generate reports
PSO Parameters	Number of particles: 30, Maximum iterations: 50, Acceleration coefficients: 2.05, 2.05, Inertia weight: 0.9
GA Parameters	Population size: 100, Crossover rate: 0.8, Mutation rate: 0.01, Selection method: Roulette wheel selection
CCTL Parameters	Time interval: [0 ms, 250 ms], Event expressions: Defined based on software functionality requirements
LSTM and LAR Parameters	LSTM hidden units: 128, LAR order: 4
Test Case Data	Derived from functional requirements documents of online shopping systems and user operation logs
Performance Data	Generated under various loads using JMeter tool
Security Data	Generated through security scans with OWASP ZAP tool
Software Defect Data	Collected from previous software testing and maintenance records

Neural Backward Chaining Logic Algorithm Based on Dynamic Knowledge Region Segmentation in Knowledge Graph Completion

Chunmao Liu[1,2], Somkiat Tuntiwongwanich[1*], Thiyaporn Kantathanawat[1]
[1]School of Industrial Education and Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand
[2]School of Electronic and Information Engineering, Henan Polytechnic Institute, Nanyang, 473000, China
E-mail: somkit_l@126.com
*Corresponding author

Keywords: internet, knowledge graph, recurrent neural networks, neural backward chain reasoning, unified medical language system

Received: October 23, 2024

With the rapid development of the Internet, the information age has arrived in a comprehensive way. The explosive growth of information data has greatly increased people's demand for knowledge management. Knowledge graph, as an effective structured knowledge representation tool, greatly enhances the organization and retrieval capabilities of information. However, in practical applications, knowledge graphs often face problems of knowledge loss and incompleteness, which severely limit their widespread application. To address this issue, this study proposes a neural backward chain inference method based on dynamic knowledge region generation. This method overcomes the bottleneck of performance degradation of traditional static methods on large datasets by introducing a dynamic knowledge region generation mechanism, which significantly improves the completion effect of knowledge graphs. The experiment was conducted on the Unified Medical Language System dataset and the Nations dataset. The results showed that when the size of the Unified Medical Language System dataset reached 2500, the accuracy of the proposed method reached 0.85. It was 6.25%, 20%, and 51.8% higher than the 0.80 of the neural backward chain inference method generated by static knowledge regions, 0.70 of the conditional theorem prover algorithms, and 0.56 of the traditional neural theorems proving algorithm, respectively. In the Nations dataset, the accuracy of the proposed method was 0.80 at the same data scale, which was significantly better than other methods. In addition, the method based on dynamic knowledge region generation reduced the iteration time to 1.4 seconds, which was about 52% and 63% higher than the static method's 2.9 s and the traditional method's 3.8 s, respectively. The research results indicate that the proposed neural backward chain logic algorithm based on dynamic knowledge region generation exhibits good performance in completing knowledge graphs.

Povzetek: Opisana metoda za dopolnjevanje znanstvenih grafov uporablja nevralno obratno verigo sklepanja s dinamičnim generiranjem znanstvenih področij, kar povečuje kvaliteto v primerjavi s tradicionalnimi statičnimi metodami.

1 Introduction

With the development of the Internet, the information age has arrived, and the way people live and work has changed radically. The development of mobile Internet has made the web no longer serve only specific professionals, but become a tool used by all people [1]. The sheer volume of information makes it particularly difficult for people to query information sets, and the emergence of knowledge graph (KG) solves this problem. KG is a graphical database for organizing and representing structured knowledge, but there are some problems with KG, KG may have incomplete information and some information about relationships or attributes between entities may be missing [2]. This may be due to the fact that the information has not yet been added to the graph, or due to the dynamic nature of knowledge, new information is generated but not yet updated to the graph. It is the missing and incomplete nature of the knowledge in KGs that poses a significant limitation to the use of KGs. The research aims to address the problems of low efficiency and insufficient accuracy in KG completion in large-scale and dynamic environments. A neural backward chain inference (BCI) method based on dynamic knowledge region generation is proposed. This method assumes that by dynamically adjusting the knowledge region, the accuracy and computational efficiency of KG completion can be significantly improved. It is hypothesized that this method will demonstrate higher accuracy and lower computation time than traditional static methods on datasets of different sizes and types. This proves the effectiveness of the dynamic knowledge region generation mechanism in KG completion. In this research, a neural theorem provers (NTPs) model based on BCI is proposed, and a recurrent neural network (RNN)-based NTPs model is proposed for the

improvement of the model in response to the shortcomings of the model in terms of computational efficiency. Aiming at the problem that RNNNNTPs are not ideal for dynamic knowledge area (KA) complementation, a conditional theorem provers (CTPs) algorithm is proposed to improve the NTPs. There are four primary sections to this research. The first includes a synopsis of previous studies on the subject of KG research. The second part is a review of the main methods used in this research, and the third part is the model results obtained by applying the methods to the research and analyzing the results. The fourth part is a summary of all the above studies and the outlook for future research.

2 Related works

The emergence of KG has brought multiple benefits to Internet information. Shen et al. discovered that substantial semantic information may be gleaned from the relationships and attributes of KGs, aiding in the creation of possible semantic representations of KGs. Knowledge representation approaches could reduce variation between distinct knowledge domains by aligning things through their conversion into spatial vectors. Nevertheless, previous methods neglected the distinctions between relations by using the same optimization objective for triples under various relations. The study team employed an adaptive margin strategy in training and provided an entity alignment method based on the TransE model to address this issue. The results of the study demonstrated that the proposed method had significant improvement compared to the baseline model [3]. Wu et al. found that due to the large amount of item information in KG that could help recommender systems to develop users, it becomes the most important source of auxiliary information. However, there were some problems in this application scenario. To address this problem, the research team proposed a multi-context-aware recommendation algorithm based on KG. According to experimental data, the suggested approach could handle KG faults more effectively [4]. Hussey et al. proposed to advance nursing theory through nursing KG using health informatics standards and translating nursing knowledge for nursing research. Experimental results showed that the proposed KG was able to gain valuable insights into nursing interventions and inform future big data science [5].

Lin et al. discovered that knowledge representation is a crucial stage in building domain KG, and that domain KG has emerged as a study issue in the artificial intelligence era. However, there were some gaps in the construction of domain KGs. Therefore, the research team studied the classification of KGs, and the study analyzed the unresolved issues and future research trends of knowledge representation in domain KG research [6]. Haoran et al. found that KG complementation could solve the problem of sparse data in KGs. However, under-utilization of structural information around nodes was the main problem of the traditional KG complementation model, which led to relatively homogeneous encoded information. To address this problem, the research team proposed a new KG complementation model for encoding and decoding feature information. The study's findings showed that the suggested model functioned well on the used dataset [7]. According to Yutian et al., KG was a helpful tool and resource for describing items and connections in tasks involving natural language processing. The function of logic rules and the impact of false negative samples on knowledge embedding, however, were not fully taken into account by traditional methods of KG embedding, and the current KGs were insufficient. The model could well enhance the complementing effect of KG, according to experimental results [8].

In summary, many scholars have already had research on KG and achieved certain results. In this study, a BCI-based NTPs model is proposed, and for the shortcomings of the computational efficiency of this model, the model is improved by proposing an RNN-based NTPs model, and for the problem that RNNNNTPs are not ideal for dynamic KG complementation, the NTPs are improved and a CTPs algorithm is proposed. The comparison table of literature review is shown in Table 1.

Table 1: Related works

Research	Method	Research content	Dataset used	Key performance metrics	Reference
Shen et al. (2022)	Entity alignment method based on transe model with adaptive margin strategy	Proposed an entity alignment method addressing the issue of using the same optimization objective for triples under different relations in knowledge graphs, enhancing alignment through an adaptive margin strategy	Knowledge graph embedding datasets	Significantly improved entity alignment compared to baseline models	[3]
Wu et al. (2022)	Multi-context-aware recommendation algorithm based on	Developed a multi-context-aware recommendation	Recommendation system related datasets	Effectively addressed defects in knowledge	[4]

	knowledge graph	algorithm leveraging project information in knowledge graphs to enhance recommendation systems, addressing deficiencies in knowledge graphs		graphs, improving recommendation accuracy	
Hussey et al. (2021)	Method using health informatics standards and translated nursing knowledge	Constructed a nursing knowledge graph by combining health informatics standards and nursing knowledge to advance nursing theory	Nursing big data	Provided valuable insights for nursing interventions, supporting nursing theory and big data science research	[5]
Lin et al. (2021)	Classification study of domain knowledge graphs, analyzing knowledge representation issues and trends	Researched the classification of domain knowledge graphs, analyzed unresolved issues in knowledge representation within domain knowledge graph studies, and proposed future research directions	AI domain related datasets	Analyzed challenges in knowledge representation in domain knowledge graph studies and proposed future research directions	[6]
Haoran et al. (2022)	Knowledge graph completion model based on graph convolutional networks (GCN) with multi-information fusion and high-dimensional structure analysis weighting	Proposed a new feature encoding and decoding knowledge graph completion model using GCN for multi-information fusion and high-dimensional structure analysis	Knowledge graph completion related datasets	The proposed model performs well on the used datasets, improving the accuracy of knowledge graph completion	[7]
Yutian et al. (2021)	Knowledge graph embedding model based on soft rules and adversarial learning	Combined soft rules and adversarial learning to propose a new knowledge graph embedding method, enhancing knowledge graph completion	Knowledge graph embedding datasets (specific datasets not detailed)	Effectively improved knowledge graph completion, enhancing the robustness of embeddings	[8]

3 Knowledge graph complementation model based on neural backward chain logic algorithm

The first section of this chapter introduces KGs and builds a BCI-based model for NTPs. The second chapter addresses the problems of the complementary model and proposes a neural BCI method based on dynamic KG generation to complement KGs.

3.1. Modeling of NTPs based on backward chain reasoning

KG usually adopts the ternary form to represent structured semantic knowledge. The ternary consists of subject, predicate and object, which are used to describe the relationship between the entities, where the entity is the most basic element, which is generally a real-world physical object or an abstract concept, and the relationship is the link connecting the entities, and in the construction of KG, it will generally contain three phases: information extraction, knowledge fusion, and knowledge processing [9]. Inductive logic programming (ILP) is a machine learning method that combines inductive reasoning and logic programming. It aims to automatically learn logic programs from given facts and rules for reasoning and decision making. The basic idea of ILP is to reason and make decisions by generalizing general logical rules from examples. It does this by representing examples as logical facts and goals, and then finding logical rules that satisfy the goals by using

logical reasoning and search algorithms, the flow of which is shown in Fig. 1.

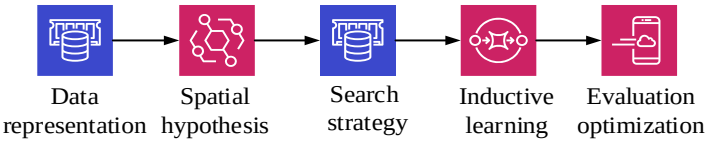


Figure 1: ILF flow

In Fig. 1, the ILF algorithmic process is mainly divided into five types, which are data representation, hypothesis space, search strategy, inductive learning, and evaluation and optimization. Among them, data representation refers to the form of transforming example data into logical facts, i.e., mapping attributes and relations in data into logical predicates. Hypothesis space refers to the determination of the hypothesis space of logical rules, which is usually realized by defining the formal syntax and constraints of logical rules. Search strategy refers to the selection of an appropriate search strategy to search for logical rules in the hypothesis space. Inductive learning refers to the continuous adaptation of logical rules during the search process, based on a given objective function and learning algorithm, and the evaluation of their accuracy and generalization ability on example data. Evaluation optimization refers to evaluating the performance of the logic rules obtained by learning and optimizing and correcting the rules according to the evaluation metrics of the learning algorithm [10-11].

RNN is a neural network model for processing sequential data or data with time dependency [12]. Unlike traditional feed-forward neural networks, RNN has recurrent connections that can transfer information within the network, giving it memory capability, the RNN recurrent hidden state update formula is shown in equation (1).

$$h_t = \begin{cases} 0, & t = 0 \\ \varphi(h_{t-1}, x_t), & otherwise \end{cases} \tag{1}$$

In equation (1), h_t denotes the cyclic hidden state of the RNN, φ denotes the nonlinear function, and x denotes the sequence. Equation (2) displays the update of the cyclic hidden state.

$$h_t = g(Uh_{t-1} + Wx_t) \tag{2}$$

In equation (2), g represents a smooth bounded function and W and U denote constant parameters. Conventional RNNs are prone to the problem of gradient vanishing or gradient explosion during the training process, which leads to difficulties in capturing long-term temporal dependencies. To solve this problem, some improved RNN structures, such as gated recurrent unit (GRU) and Long Short-Term Memory (LSTM), have emerged to deal with long-term dependencies efficiently by introducing a gating mechanism, and the structure of LSTM is shown in Fig. 2.

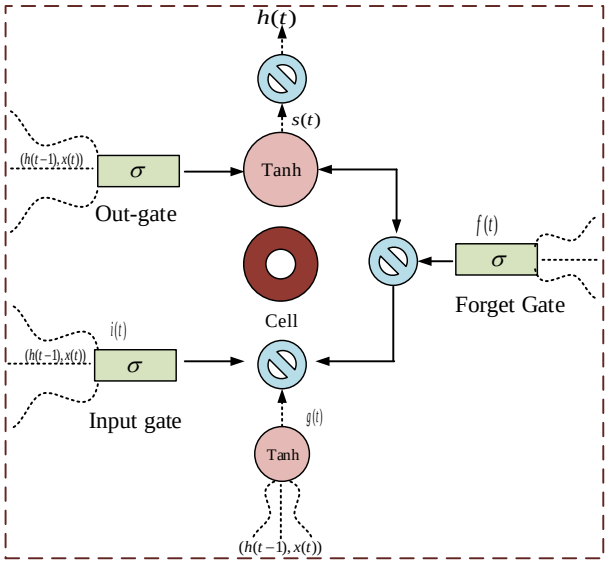


Figure 2: Structure of LSTM

In Fig. 2, the input layer input at moment t is $x(t)$, $h(t-1)$ is the hidden layer at moment $t-1$, and the internal state s is the secret key for neuron activation and also determines the network structure of the LSTM. Equation (3) shows the output of the LSTM unit, which is a value $f(t)$ between 0 and 1. This value is obtained through $h(t-1)$ and $x(t)$, representing a trade-off between retaining or discarding the internal states. The activation function used is a sigmoid function.

$$h_t^j = o_t^j \tanh(c_t^j) \tag{3}$$

In equation (3), o_t^j denotes the output gate that controls exposure, c_t^j denotes the memory cell. Equation (4) displays the expression for the output gate.

$$o_t^j = \sigma(W_o x_t + U_o h_{t-1} + V_o c_t)^t \tag{4}$$

In equation (4), σ represents the sigmoid function and V_o represents the diagonal matrix. The degree of forgetting of existing memories is regulated by the forgetting gate, and the new memory content is controlled by the updating gate, and the two gates are calculated as shown in equation (5).

$$\begin{cases} f_t^{\,j} = \sigma(U_f h_{t-1} + W_f x_t + V_f c_{t-1})^j \\ i_t^{\,j} = \sigma(U_i h_{t-1} + W_i x_t + V_i c_{t-1})^j \end{cases} \quad (5)$$

In equation (5), V_i and V_j denote diagonal matrices. Based on the current input and the concealed state of the previous instant, the input gate uses a sigmoid function to determine which information needs to be added to memory. Using a sigmoid function, the forgetting gate determines whether data should be erased from memory based on the input at hand and the hidden state of the preceding instant. Using a sigmoid function, the output gate determines which data should be output depending on the input at hand and the hidden state from the previous instant.

In the complementation of KG, when there are missing relations or entities in the triad, all relations or entities in the KG are generally complemented to the triad, and the triad for which the filling is performed is used as a query option, sorted and scored, and the relations or entities with the highest scores are used as the final result output [13]. BCI based NTPs apparatus are characterized by micro-computability and strong interpretability of rules and results. The core of RNN is the GRU, which is an improved RNN structure with good performance and computational efficiency. The GRU model incorporates update and reset gates to regulate the flow of information at every moment, thereby enabling the model to selectively retain or forget information. This approach effectively solves the gradient vanishing problem that traditional RNNs frequently encounter when processing long sequences. In practical implementation, RNNs utilize the features of each entity as input, combining them with the relevant relationships in the graph to achieve inference. During each inference, the RNN computes the current hidden state based on the current input and the previous hidden state. This recursive structure allows the model to maintain a "contextual" memory of the entire KG at each time step, taking into account the dependencies between entities and relationships in the graph during knowledge completion. In NTPs, there are three main modules, which are unfication (unify), or and. the computational process of NTPs can be approximated as a search. That is, or indicates that any rule in the knowledge base serves the goal of the proof, and indicates that the premises of the rules must be recursively proved. RNN is used as the regulation and facts for the knowledge generation operation because the proof process of NTPs is top-down, the unify operation needed by each layer will be increased step by step, and there is a sequential relationship between the upper layers and lower layer's rule facts. Its structure is shown in Fig. 3.

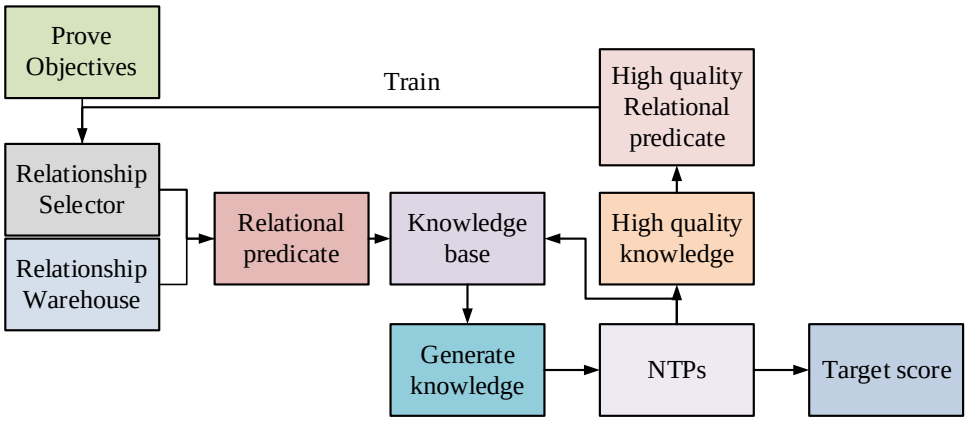


Figure 3: RNNNTPs model structure

In Fig. 3, the proof target is firstly input into the relation selector, and the relations in the relation warehouse are selected through the relation selector. The knowledge in the knowledge base is selected according to the selected relation predicates, the generated knowledge is fed into NTPs to generate high quality knowledge, and the generated new rules are added to the knowledge base, the high-quality relation predicates are refined by the high-quality knowledge, and in turn the relation selector is trained, and finally the NTPs outputs the target values by the generated knowledge. In this study, RNN selector is used to predict the sequence and the generated sequence is shown in equation (6) [14].

$$h_0 = f(r) \quad (6)$$

In equation (6), where f denotes the linear transformation, the relationship between different positions is predicted by the gating unit and the expression is shown in equation (7).

$$h_t = GRU(h_{t-1}, g([r_{t-1}, r_t])) \quad (7)$$

In equation (7), g denotes the linear transformation and $[r, r_t]$ denotes the connection between the previous relation and the current relation. The generated relation set is divided by softmax function to get the next relation. In RNNNTPs, the relation selector generates only the relation set generates the knowledge set by

matching algorithm and the expression is shown in equation (8).

$$\begin{cases} KB_{gen} = Select(KBs, \mathrm{Re}l_{gen}) \\ Select(KB, \mathrm{Re}l) = KB' \end{cases} \tag{8}$$

In equation (8), KB denotes the knowledge base and $\mathrm{Re}l$ denotes the set of relations. Since NTPs have the problem of slow computational efficiency, the modules in NTPs are improved to address this problem. The Or module is improved by reconstructing it so that it only considers the knowledge provided by the knowledge generator. By establishing a threshold to govern the Unify module, it may be made better. If a similarity surpasses the threshold, it will be added to the high-quality knowledge set [15]. Fig. 4 depicts the revised algorithm's flow.

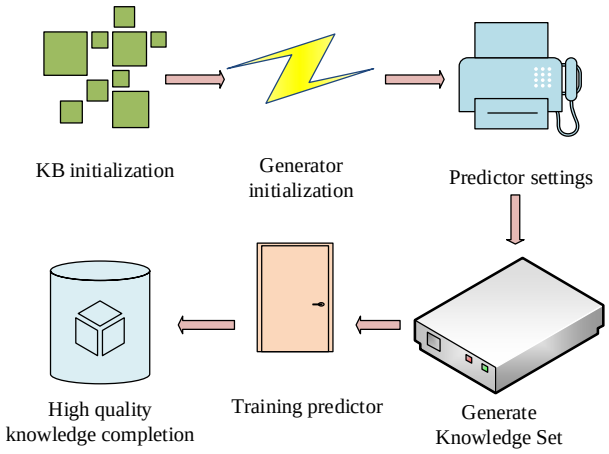


Figure 4: Flowchart of the improved algorithm

In Fig. 4, firstly the ternary dataset is input and then it is preprocessed, secondly the GRU-based RNN is built, then the predictor setting is performed and ComplEx is used as the optimization function for the NTPs. The knowledge set is generated by the predictor setting, the individual relations in the target are proved by the RNN selector to generate the set of relations. To create the knowledge set, the relations in the set are matched in the KBs. The predictor is then trained to add the closest knowledge into the KBs for knowledge complementation.

3.2. Knowledge graph complementation model based on dynamic knowledge search region segmentation

RNNNTPs substantially improve the computational efficiency of NTPs, but there are some problems, in the relation selection, some relations will appear more frequently, and some simple relations will be matched with more knowledge. This presents a significant challenge for the next connection selection process. To address this, the CTP algorithm is used to optimize the model by incorporating the neural layer of knowledge selection into the predictor. This further increases the accuracy of the link selection process. CTPs are computational models that prove given conditional propositions through the use of conditionals and inference rules, and can be used in areas such as logical reasoning, theorem proving, and formal verification [16]. The main use of CTPs is to automate the theorem proving process. CTPs algorithm can receive conditions and propositions to be proved and try to prove these propositions using logical inference rules and reasoning strategies, the model of RNN-CTPs algorithm is shown in Fig. 5.

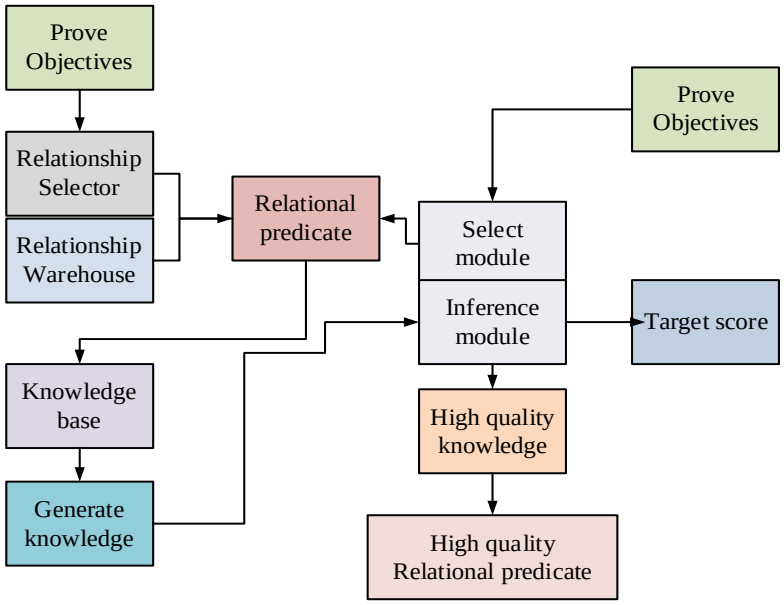


Figure 5: Model flow of RNNCTPs algorithm

In the relation selector, the sequence relations are predicted by RNN relation selector, after initializing the RNN selector, the relations at different positions are predicted by gating unit, the generated rule set is divided by softmax function to get the next rule body, and the number of generated relations is shown in equation (9).

$$relNum = batchSize \times size(\mathrm{Re}\, l_s) \times rel_\lambda \times KB_\lambda \quad (9)$$

In equation (9), $relNum$ denotes the generated relations and $batchSize$ is the proof objectives. $relNum$ denotes the hyperparameter whose purpose is to adjust the relations generated, and $size(\mathrm{Re}\, l_s)$ denotes the relations present in the relation repository. KB_λ denotes the degree to which the generated knowledge matches the relations. In the relation warehouse, dynamic expansion coefficients are used and the dynamic change of expansion coefficients is shown in equation (10).

$$ep_i^{new} = ep_i^{last} - \kappa(KB_\lambda^{new} - KB_\lambda^{last}) \quad (10)$$

In equation (10), κ represents the hyperparameter that controls the diffusion coefficient, and KB_λ^{new} and KB_λ^{last} are the current KB_λ and the KB_λ of the previous cycle, respectively. CTPs, as an improved version of NTPs, reduces the amount of knowledge that needs to be learned by increasing the neural layer used for knowledge selection, which improves the efficiency of computation. To achieve better computation, CTPs will choose alternative functions dependent on the object. The selection module is based on the attention function, as illustrated in equation (11).

$$f_i(x) = \alpha E_R \quad (11)$$

In equation (11), E_R denotes the embedding matrix of predicates, and α denotes the attention distribution of predicate verbs in the matrix, the expression of which is shown in equation (12).

$$\alpha = soft\max(W_i x) \in \Delta^{|R|-1} \quad (12)$$

In equation (12), α denotes the attention distribution of the predicate verbs in the matrix and W_i denotes the elements embedded in the matrix. CTPs go through the selection module to generate a set of knowledge by proving the goal. After each generation of knowledge set by the model, variables need to be recorded and the expression is shown in equation (13).

$$KB_{\lambda_i} = \frac{\left|KB_{Used_i}\right|}{\left|KB_{gen_i}\right|} \quad (13)$$

In equation (13), KB_{λ_i} denotes the set of variables, KB_{Used_i} denotes the knowledge that is involved in the process at this time. Moreover, KB_{gen_i} represents the knowledge produced by the CTPs following the feeding of inputs. The flow of the algorithmic model of the RNNCTPs is shown in Fig. 6.

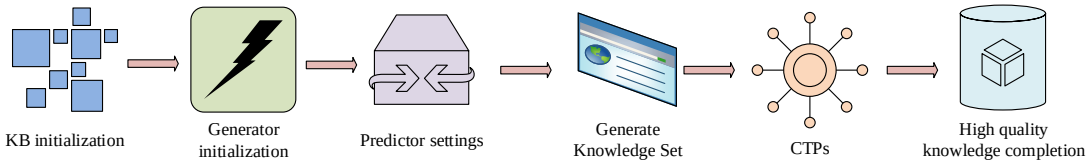


Figure 6: Model flow of RNNCTPs algorithm

Firstly, the ternary dataset is input and then it is preprocessed, secondly the GRU based RNN is built and then the predictor setting is performed. The predictor setting creates the knowledge set. The RNN selector proves each individual relation in the target to create the set of relations. Moreover, the KBs match the relations in each set to create the knowledge set. The predictor is then trained to add the closest knowledge to the KBs for knowledge complementation. The training of the CTPs module already includes the selection of the module, so the amount of knowledge used for the unify module to run is under control. Each iteration produces limb utilization as shown in equation (14).

$$KB_\lambda = \begin{cases} KB_\lambda + \tau(KB_{\lambda mean} - KB_\lambda), if \left|KB_{\lambda mean} - KB_\lambda\right| > \varsigma \\ \dfrac{KB_\lambda + KB_{\lambda mean}}{2}, \qquad else \end{cases} \quad (14)$$

In equation (14), τ A denotes the learning rate of KB_λ update. ς denotes the learning threshold, which can be fine-tuned if the difference between $KB_{\lambda mean}$ and KB_λ is too large.

```
RNN-CTPs

# Input: Dataset X (features) and Y (labels), RNN parameters, CTPs parameters
# Output: Predicted labels Y_pred

# Step 1: Initialization
1. Initialize RNN structure (e.g., number of layers, number of hidden units).
2. Define CTPs parameters (e.g., tree depth, splitting criterion such as Gini index or information gain).

# Step 2: Training Phase
2.1 Feature Extraction with RNN:
   for each sample (x, y) in (X, Y):
      # Pass input sequence x through RNN to extract features
      h_t = RNN(x_t; W_rnn)  # W_rnn is the weight matrix of the RNN
      h_T = FinalState(h_t)  # h_T is the final hidden state representation of the sequence

   # Collect the final hidden states h_T as the feature set for all samples

2.2 Build Conditional Probability Trees (CTPs):
   1. Initialize root node with all training samples and their extracted features h_T.
   2. While the stopping criterion is not met (e.g., max depth or minimum samples in a node):
      a. For each node, calculate the optimal splitting condition based on features h_T
         and maximize conditional probabilities of labels.
      b. Split the data into child nodes based on the chosen condition.
   3. At each leaf node, compute the probability distribution of labels and store it.

# Step 3: Prediction Phase
3.1 Predict for each test sample x_test:
   a. Pass x_test through the RNN to obtain feature representation h_test.
   b. Traverse the CTPs using h_test, following the splitting conditions until reaching a leaf node.
   c. At the leaf node, select the class with the highest probability as the predicted label y_pred.

# Step 4: Output
4. Return all predicted labels Y_pred.
```

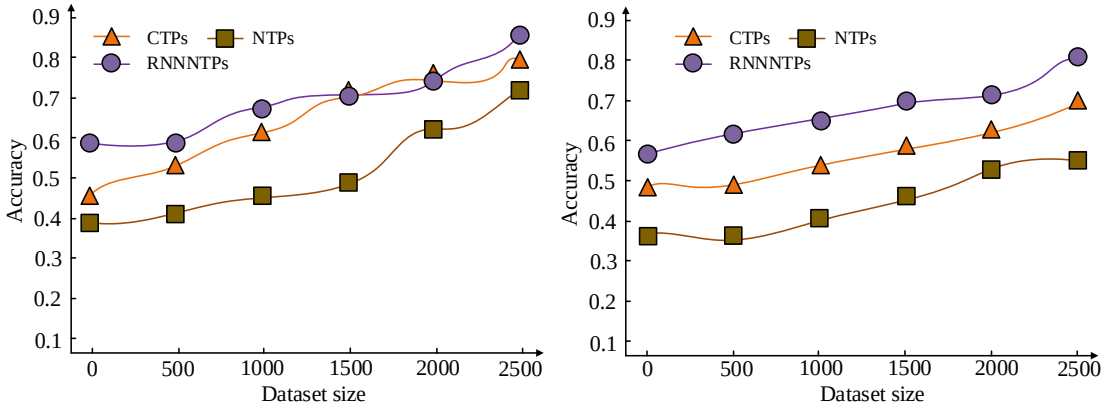
Figure7: RNN CTPs pseudocode

4 Performance analysis of knowledge graph complementation model based on neural backward chain logic algorithm

In the first section of this section, the performance comparison is carried out for static KG segmentation of KG using UMLS dataset and Nation's dataset, and the results of the study show that it performs well in the UMLS dataset. Therefore, the performance of KG complementary model based on dynamic KG segmentation is analyzed using UMLS dataset.

4.1. Performance analysis of knowledge Graph complementary model based on static knowledge area segmentation

The server CPU used in this study is Inter(R) Core (TM) i5-10210U with 16GB of RAM, GPU is GeForce RTX 2080 operating system is Windows 10 with 8G of RAM. Using a systematic hyperparameter selection method, different combinations of hyperparameters such as learning rate, batch size, GRU layers, number of hidden units, and similarity threshold are first traversed by grid search, and the performance of each parameter set is evaluated on the validation set to finally select the optimal parameter combination. To ensure the robustness of the selection, k-fold cross-validation is used to divide the training data into k subsets, and multiple training and validation are performed to calculate the average performance of each parameter group. At the same time, an early-stop mechanism is introduced to stop training if the performance of the validation set did not significantly improve in 10 consecutive training rounds to prevent overfitting. The specific hyperparameter settings are as follows: learning rate set to 0.001, batch size set to 64, GRU layers set to 2, number of hidden units set to 128, similarity threshold set to 0.7, Dropout rate set to 0.5, optimization algorithm using Adam, iteration count set to 300, embedding dimension set to 100, activation function set to ReLU, and batch normalization applied in the model. To ensure the robustness of the results, each experiment is run independently at least five times with the same hyperparameter settings, using different random seeds to avoid random effects, and taking the average of key performance indicators such as accuracy, iteration time, and knowledge utilization, while calculating the standard deviation to evaluate the stability of the results. When comparing performance, paired t-test and analysis of variance methods are used with a significance level of α=0.05 to ensure that performance differences between different models are statistically significant. At the same time, a 95% confidence interval is calculated to provide an uncertainty assessment for performance estimation and increase the reliability of result interpretation. The study is conducted using the Unified Medical Language System (UMLS) dataset and the Nations dataset. UMLS is a medical knowledge representation and retrieval system developed by the National Institutes of Health. Its purpose is to facilitate the interoperability and exchange of medical information so that different medical information resources can be connected and shared. UMLS contains many medical vocabularies, coding systems, and standardized terminology, including 49 predicates, 135 constants, and 6529 true facts. It also unifies terminology, coding, and standards from various medical disciplines. Nations has 2565 true facts, 111 unary predicates, 14 constants, and 56 binary predicates. The performance of the proposed model is analyzed on both datasets and the results are shown in Fig. 8.



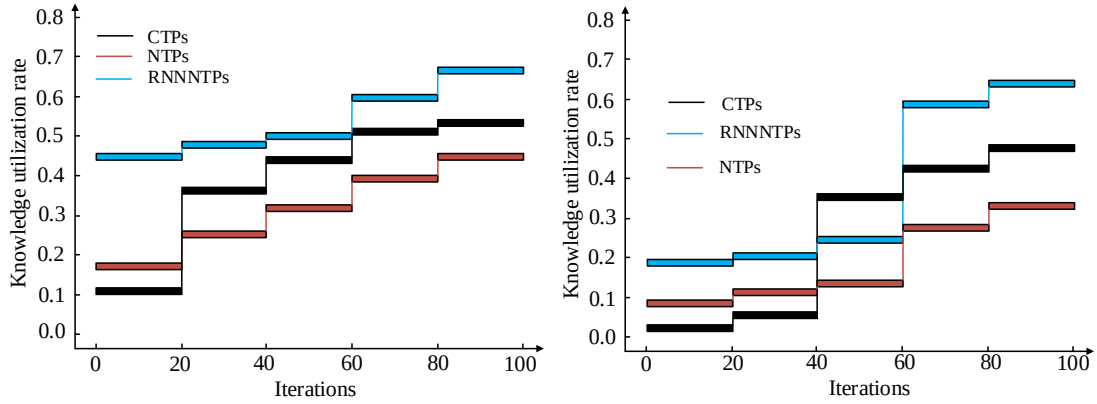
(a) Performance of Three Algorithms on UMLS Datasets

(b) Performance of Three Algorithms on the Nations Dataset

Figure 8: Comparison of accuracy of three algorithm models in different datasets

Fig. 8(a) shows the performance of three algorithms on the UMLS dataset, while Fig. 8(b) shows the performance of the three algorithms on the Nations dataset. According to Fig. 8(a), in the UMLS dataset, as the dataset increases, the accuracy of RNNNTPs, CTPs, and NTPs algorithm models also fluctuates and increases. Among them, the RNNNTPs algorithm model shows higher performance. When the dataset size is 2500, the accuracy of RNNNTPs, CTPs, and NTPs algorithm models are 0.88, 0.79, and 0.71, respectively. As shown in Fig. 8(b), in the Nations dataset, the accuracy of the three algorithms also increases with the size of the dataset, but the training performance is relatively poor compared to the UMLS dataset. When the dataset size is 2500, the accuracy of RNNNTPs, CTPs, and NTPs algorithm models are 0.80, 0.70, and 0.56, respectively. The experimental results show that the proposed hybrid algorithm model exhibits high accuracy among the three algorithms and performs well in training on the UMLS dataset. Compare the performance of three algorithms at different iteration times, and the results are shown in Fig. 9.

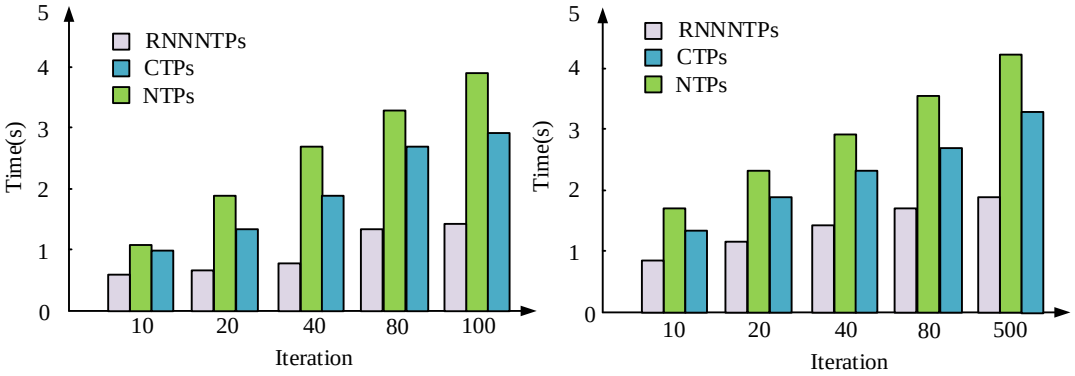
Fig. 9(a) shows the knowledge utilization rates of three algorithms in the UMLS dataset, while Fig. 9(b) shows the knowledge utilization rates of the three algorithms in the Nations dataset. As shown in Fig. 9(a), with the increase of iteration times, the knowledge utilization rate of the three algorithms also continues to rise. When the number of iterations reaches 100, the knowledge utilization rates of RNNNTPs, CTPs, and NTPs algorithm models in the UMLS dataset are 0.68, 0.53, and 0.46, respectively. In Fig. 9(b), with the increase of iteration times, the knowledge utilization rate of the three algorithm models also continues to rise, but it is significantly lower than that of the UMLS dataset. When the number of iterations reaches 100, the knowledge utilization rates of the three algorithm models are 0.65, 0.48, and 0.32, respectively. The experimental results show that the proposed RNNNTPs algorithm model has better performance compared to other algorithm models, and each algorithm model performs well in the UMLS dataset. The iteration times of the three algorithms are compared for different numbers of iterations, and the results are presented in Fig. 10.



(a) The knowledge usage rate of three algorithms in the UMLS dataset

(b) The knowledge usage rate of three algorithms in the Nations Dataset

Figure 9: Knowledge utilization rate of three algorithms at different iteration times



(a) Iteration time of three algorithms in the UMLS dataset

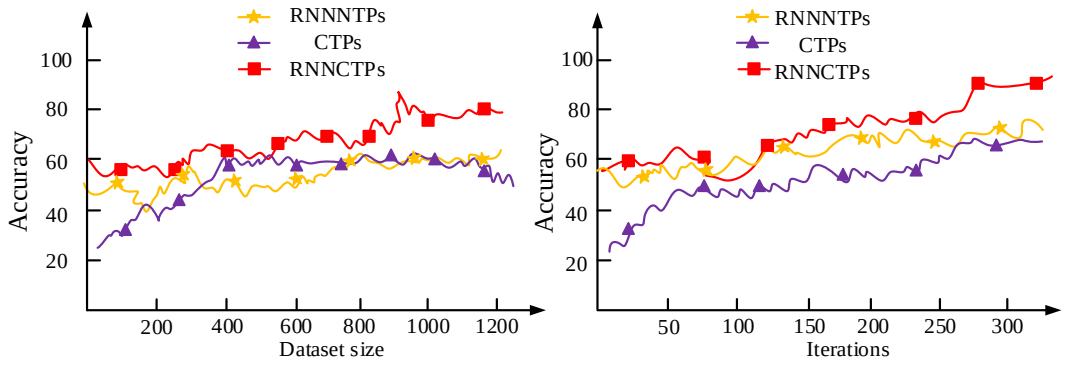
(b) Iteration time of three algorithms in the Nations dataset

Figure 10: Iteration time in different datasets

Fig. 10(a) represents the iteration time of the three algorithms in the UMLS dataset and Fig. 10(b) represents the iteration time of the three algorithms in the Nations dataset. In Fig. 10, the RNNNTPs algorithm model exhibits less time used for iterations than the other two methods in all three methods. The iteration times of the RNNNTPs, CTPs and NTPs algorithm models in the UMLS dataset are 1.4s, 2.9s and 3.8s, respectively, when the number of iterations is 100. In the Nations dataset, the iteration times of the three algorithms are 1.9s, 3.2s and 4.3s, respectively. The findings revealed that the proposed RNNNNTPs algorithm has a lower iteration time under different iteration number of iterations have low iteration time.

4.2. Performance analysis of knowledge graph complementation model based on dynamic knowledge area segmentation

The findings are displayed in Fig. 11 after the superior UMLS dataset is chosen and the model performance is examined to compare each model's performance.

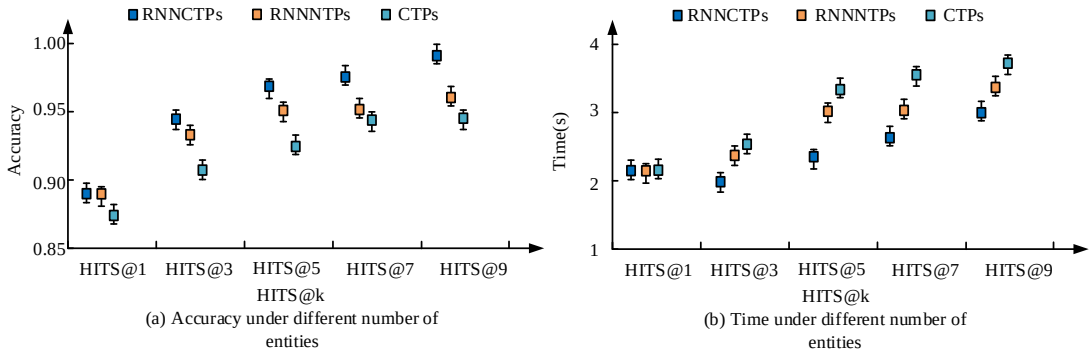


(a) Accuracy under different dataset sizes

(b) Accuracy under different iterations

Figure 11: Accuracy of three algorithmic models with different dataset sizes and number of iterations

Fig. 11(a) represents the accuracy comparison of the three algorithms under different dataset sizes, and Fig. 11(b) represents the accuracy comparison of the three algorithms under different iteration numbers. In Fig. 11(a), the accuracy of the three models fluctuates and rises as the dataset increases. The accuracy of the RNNCTPs, RNNNNTPs, and CTPs algorithmic models are 0.82, 0.64, and 0.57, respectively, when the dataset size is 1200. In Fig. 11(b), the accuracy of the three models fluctuates and rises as the iterations increases. The accuracy of the three algorithmic models fluctuates and rises when the iterations is 300, the accuracy rates of the three algorithmic models are 0.94, 0.76, and 0.78, respectively. The study's findings show that, out of the three models, the suggested RNNCTPs model—which is based on dynamic KG segmentation—performs better. The HITS@k evaluation metric is used to compare the three methods. The findings are displayed in Fig. 12 and this metric indicates the percentage of real entities that are present in the top k entities of the sorted ranked list. Fig. 12(a) shows the accuracy of the three algorithms for different entities, while Fig. 12(b) shows the iteration time for the same. Among the three algorithmic models, the proposed RNNCTPs model based on dynamic KG segmentation has a better performance. The comprehensive performance of the model is analyzed, and the results are shown in Table 2.



(a) Accuracy under different number of entities (b) Time under different number of entities

Figure 12: Table comparing the performance of the three algorithms

Table 2: Model comprehensive performance analysis

Dataset size	Model	ACC	Calculation overhead/s	Mean response time/s	Training time/s	Iteration time/s	Model stability
1000	RNNCTPs	0.75	12.3	1.2	85	1.5	0.03
	RNNNTPs	0.7	15.4	1.8	92	2.1	0.05
	CTPs	0.6	18.2	2.0	99	2.5	0.06
	NTPs	0.45	21.6	2.3	110	3.2	0.08
1500	RNNCTPs	0.78	15.6	1.4	130	1.7	0.02
	RNNNTPs	0.74	18.0	1.9	138	2.3	0.04
	CTPs	0.68	20.1	2.2	145	2.8	0.05
	NTPs	0.5	23.0	2.6	155	3.6	0.07
2000	RNNCTPs	0.82	18.5	1.5	170	1.9	0.01
	RNNNTPs	0.78	21.2	2.0	180	2.5	0.03
	CTPs	0.72	23.4	2.4	190	3.0	0.04
	NTPs	0.55	26.2	2.8	200	3.9	0.06
2500	RNNCTPs	0.85	20.7	1.6	210	1.4	0.02
	RNNNTPs	0.80	23.0	2.2	225	2.7	0.04
	CTPs	0.75	25.3	2.6	240	3.2	0.05
	NTPs	0.56	28.6	3.0	260	4.3	0.07

As demonstrated in Table 2, the RNNCTPs model displays superior accuracy and reduced computational overhead across various dataset sizes, particularly at a dataset size of 2500, where the accuracy reaches 0.85. Concurrently, the response time and computational overhead of the RNNCTP model are notably lower compared to other models. Furthermore, the iteration and training times of RNNCTPs are relatively brief, suggesting their efficacy in handling large-scale datasets. For other models, such as RNNNTPs, CTPs, and NTPs, the gap in accuracy and computational cost gradually becomes apparent as the dataset size increases, especially for the 2500 scale dataset, where the performance advantage of the RNNCTPs model is particularly prominent. The experimental outcomes reveal that the proposed RNNCTPs model has stability and robustness under different input scales, providing strong support for its wide application. Fifty people with certain medical knowledge are selected and randomly divided into five groups, and the models are scored separately. Moreover, Table 3 displays the outcomes.

Table 3: User evaluation form

Model	Group 1	Group 2	Group 3	Group 4	Group 5	Average
RNNCTPs	96.4	87.6	94.1	86.2	85.2	89.9
RNNNTPs	91.2	87.3	86.1	80.4	77.6	84.5
CTPs	87.5	72.6	78.2	77.3	70.4	77.2

In Table 3, the ratings of the five groups of users for the RNNCTPs model are 96.4, 87.6, 94.1, 86.2, 85.2, 89.9, with an average of 89.9. The ratings of the RNNNNTPs model are 91.2, 87.3, 86.1, 80.4, 77.6, and 84.5, with an average of 84.5. The ratings of the CTPs model are 87.5, 72.6, 78.2, 77.3, 70.4, 77.2, and 77.2, respectively. are

87.5, 72.6, 78.2, 77.3, 70.4, 77.2 with an average score of 77.2. According to the experimental findings, there is more satisfaction with the suggested RNNCTPs algorithmic model.

5 Discussion

To complete the KG, a neural backward chain logic algorithm based on dynamic knowledge region generation was proposed. The experimental results showed that this method was significantly better than traditional static methods and traditional neural theorem proving algorithms in terms of accuracy and computational efficiency. First, the proposed RNNCTPs model achieved accuracy of 0.85 and 0.80 on the Unified Medical Language System and Nations datasets, respectively, showing significant improvements over the neural BCI method generated from static knowledge regions and traditional algorithms. This advantage was mainly due to the introduction of a dynamic knowledge region generation mechanism. It allowed the model to adaptively focus on relevant knowledge regions based on the current inference context, thus avoiding the performance degradation problem of static methods when processing large datasets. In addition, the integration of the CTP algorithm further optimized the knowledge selection process. By introducing a CTP in the prediction layer, the model could more accurately select relevant knowledge, reduce the search space, and accelerate the inference speed, achieving significant reductions in iteration time of about 52% and 63%. The CTP improved the accuracy of complex relationship matching by optimizing the knowledge selection process, and outperformed the soft rule and graph adversarial learning models proposed by Yutian et al [8] in complex KGs. In addition, compared to Shen L et al.'s improvement of the TransE model through an adaptive margin strategy [3], this study had higher efficiency in completing tasks. Testing on the Nations dataset further validated the robustness of the model with an accuracy of 0.80. The application of RNNs, especially the use of GRUs, enabled the model to effectively remember and utilized the information obtained during previous inference processes, thus improving the adaptability and accuracy of the model in processing dynamic and large KGs. Meanwhile, the improved Unify module filtered out low-quality knowledge by setting similarity thresholds, ensuring high accuracy and efficiency in the inference process. The user evaluation results further validated the practicality and superiority of the model. The RNNCTPs model achieved an average user satisfaction score of 89.9, significantly higher than other models, indicating its potential value and good performance in practical applications. This technology had two primary contributions. First, it enhanced the efficacy of KG completion. Second, it offered a practical solution for the management and application of large-scale dynamic KGs. This development was significant from both a technical and practical standpoint, and it held considerable potential for future applications. Nevertheless, further validation was necessary to ascertain the model's generalizability across diverse dataset types. Future research can test it on more diverse datasets and explore its integration with other artificial intelligence technologies to further improve the model's performance and adaptability. In summary, the neural backward chain logic algorithm based on dynamic knowledge region generation provides an efficient, accurate, and scalable solution for KG completion tasks, thus promoting the development of intelligent knowledge management systems. Although the research has mainly been validated on the unified medical language system and Nation’s datasets, in order to more comprehensively evaluate the performance of the proposed model, future experiments should be extended to different datasets and real-world application scenarios to further validate the model's applicability and robustness. Outside the medical field, applications can be explored in the financial sector for risk management, credit scoring and other tasks, or in recommendation systems in e-commerce to help automate product recommendations and user behavior analysis. These scenarios typically have different data characteristics and more complex practical requirements, which can provide a more challenging validation of the model's generalization ability.

6 Conclusion

The internet has made an abundance of information available due to the accelerated development of information technology in recent years. This research proposed a neural BCI method based on dynamic KG generation for complementary KG. The experimental results indicated that the knowledge utilization rates of the RNNNTPs, CTPs, and NTPs algorithmic models in the UMLS dataset were 0.68, 0.53, and 0.46, respectively, when the iterations reached 100. In the Nations dataset, the knowledge utilization rates of the three algorithmic models were 0.65, 0.48, and 0.32, respectively, when the iterations reached 100. The iteration times of the RNNNTPs, CTPs, and NTPs algorithm models were 1.4s, 2.9s, and 3.8s, respectively, in the UMLS dataset when the iterations were 100. The iteration times of the three algorithms were 1.9s, 3.2s, and 4.3s, respectively, in the Nations dataset. In the case where the dataset size was 1200, the RNNCTPs, RNNNTPs, and CTPs algorithm models were 0.82, 0.64, and 0.57, respectively. At an iteration number of 300, the three algorithm models were 0.94, 0.76, and 0.78, respectively. The study’s findings demonstrate the potential of the RNNCTPs algorithmic model as a supplement for big KGs. The superior performance of RNN CTPs over RNNNTPs and CTPs is due to their model architecture and adaptability to dataset features. RNN CTPs combine the sequence modeling capability of RNN with the conditional probability splitting mechanism of CTPs. It can capture complex temporal dependencies and efficiently segment feature spaces. This has significant advantages in diverse or high-dimensional data. In contrast, RNNNTPs rely solely on deep network capture patterns and lack the interpretability of CTPs, which can fail when feature relationships are more complex. Although CTPs have

strong interpretability, they lack the ability to extract deep features, which limits their performance on data with strong nonlinear relationships. The feature distribution and complexity of the dataset exacerbate these differences, further enhancing the advantages of RNN CTPs, especially when dealing with time series or nested patterns. The RNN part could capture rich features through long-range dependencies, and the expansion of data scale could further enhance the feature extraction capabilities, while CTPs achieved efficient splitting through conditional probability and can adapt to a wider sample distribution. However, the increase in computational complexity when training larger datasets become a challenge, especially as the sequence computation of RNNs and the tree depth of CTPs coulf lead to a decrease in training time and inference efficiency. The experimental design could select large datasets such as Freebase or DBpedia to evaluate accuracy, AUC, and time cost. The expected results showed that increasing the data size improved the model performance.

In the future, with the further development of technology and the verification of more practical applications, the proposed methods are expected to play a greater role in the construction and maintenance of KGs. These methods are poised to become an indispensable key technology in intelligent reasoning, decision support systems, and big data analysis. Therefore, the research not only provides new ideas for the field of KG completion, but also opens up new possibilities for a wider range of future application scenarios.

Funding

The research is supported by science and technology research project of Henan Province: 232102210193; Research subject of Computer basic Education in China: 2023-AFCEC-227.

References

[1] Gao L, Qiu J, Chen G. Software Test Data Management Based on Knowledge Graph. Informatica, 2024, 48(16), 241-251. https://doi.org/10.31449/inf.v48i16.6416

[2] Liu J, Wang F, Song B, Wang X. Design of an Intelligent Classification Model for Interior Design Knowledge Graph Based on Simulated Annealing Algorithm. Informatica, 2024, 48(12):44-57. https://doi.org/10.31449/inf.v48i12.6029

[3] Shen L, He R, Huang S. Entity alignment with adaptive margin learning knowledge graph embedding. Data & Knowledge Engineering, 2022, 139(5):101-132. https://doi.org/10.1016/j.datak.2022.101987

[4] Wu C, Liu S, Zeng Z, Chen M, Alhudhaif A. Knowledge graph-based multi-context-aware recommendation algorithm. Information Sciences, 2022, 595(12):179-194. https://doi.org/10.1016/j.ins.2022.02.054

[5] Hussey P, Das S, Farrell S. A Knowledge Graph to Understand Nursing Big Data: Case Example for Guidance. Journal of Nursing Scholarship, 2021, 53(3):323-332. https://doi.org/10.1111/jnu.12650

[6] Lin J, Zhao Y, Huang W, Liu C, Pu H. Domain knowledge graph-based research progress of knowledge representation. Neural computing & applications, 2021, 33(2):126-129. https://doi.org/10.1007/s00521-020-05057-5

[7] Haoran N, Haitao H E, Jianzhou F, Junlan N, Yangsen Z, Jiadong R. Knowledge Graph Completion Based on GCN of Multi-Information Fusion and High-Dimensional Structure Analysis Weight. Chinese Journal of Electronics, 2022, 32(2):387-396. https://doi.org/10.1049/cje.2021.00.080

[8] Yutian W, Jiamin C, Ling S, Yutian W. Improving Knowledge Graph Completion Using Soft Rules and Adversarial Learning. Chinese Journal of Electronics, 2021,30(4):623-655. https://doi.org/10.1049/cje.2021.05.004

[9] He L, Ye W, Wang Y X, Feng H, Chen B, Liang D. Using knowledge graph and RippleNet algorithms to fulfill smart recommendation of water use policies during shale resources development. Journal of Hydrology, 2023, 617(23):128-137. https://doi.org/10.1016/j.jhydrol.2022.128970

[10] Ma J, Zhou C, Wang Y, Guo Y, Hu G. PTrustE: A high-accuracy knowledge graph noise detection method based on path trustworthiness and triple embedding. Knowledge-based systems, 2022, 256(28):1-14. https://doi.org/10.1016/j.knosys.2022.109688

[11] Wang L, Zhang Y. Digital Dissemination of Information Based on Knowledge Graph Recommendation Algorithm. Informatica, 2024, 48(19). https://doi.org/10.31449/inf.v48i19.6561

[12] Gao L, Qiu J, Chen G. Software Test Data Management Based on Knowledge Graph. Informatica, 2024,48(16). https://doi.org/10.31449/inf.v48i16.6416

[13] Wang X, Liu A, Kara S. Constructing Product Usage Context Knowledge Graph Using User-Generated Content for User-Driven Customization. Journal of mechanical design, 2023, 145(4):323-342. https://doi.org/10.1115/1.4056321

[14] Bhosle K, Musande V. Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. Artif. Intell. Appl.2023, 1(2):114-118. https://doi.org/10.47852/bonviewaia3202441

[15] Zhao N, Long Z, Wang J, Zhao Z. AGRE: A knowledge graph recommendation algorithm based on multiple paths embeddings RNN encoder. Knowledge-based systems, 2023, 259(10):1-8. https://doi.org/10.1016/j.knosys.2022.110078

[16] Wang H, Wang Y, Li J, Luo T. Degree aware based adversarial graph convolutional networks for entity alignment in heterogeneous knowledge graph. Neurocomputing, 2022, 487(28):99-109. https://doi.org/10.1016/j.neucom.2022.02

DBN-FTLSTM: An Optimized Deep Learning Framework for Speech and Image Recognition

Xiaoyan Wang*, Yuanyuan Liang
School of Artificial Intelligence and Software Engineering, Nanyang Normal University, NanYang, HeNan, 473061, China
E-mail: wangxiaoyan_1981@126.com, liangyy2358@126.com
*Corresponding author

Keywords: deep learning, artificial intelligence, speech recognition, image, natural language processing, DBN-FTLSTM.

Artificial Intelligence's (AI) quick development has brought about a new era of technical innovation with significant ramifications for many different fields. This study explores the development of artificial intelligence (AI) for image and speech recognition, a job in natural language processing that entails the real-time computer transcription of spoken words. The use of deep learning (DL) models for voice recognition has been the subject of several research. But this field is developing quickly. This systematic review offers a thorough and in-depth analysis of research on voice recognition using DL approaches that was published between 2020 and 2024. In order to improve clarity, this study clearly outlines the methodology, including the dataset size, preprocessing methods (MFCC extraction, normalisation, and augmentation), and benchmarking criteria. To provide openness and repeatability, the datasets used for assessment must be specified in reported performance measures (accuracy, RMSE, MAE, and R^2). The dataset, experimental setup, and baseline comparisons should all be explicitly stated in order to contextualise the 99.3% accuracy claim. With an overall performance accuracy rate of 97.90%, a root means square error (RMSE) of 0.017, a mean absolute error (MAE) of 0.01, and an R^2 value of 98.7%, the DBN-FTLSTM model performs better than existing techniques. When compared to other current approaches, the recommended method's error metrics accuracy of 99.3% is the greatest. Mel-frequency cepstral coefficients were the most often used feature extraction approach, while the word error rate was the most commonly used assessment method. Another finding was the dearth of research using transformers, which have been shown to be effective models that can enable parallelization, speed up learning, and enhance the performance of low-resource languages. The findings also identified intriguing and promising areas for further study that had not gotten much attention in previous investigations.

Povzetek: Opisan je optimiziran globok učni model DBN-FTLSTM, ki dosega 99,3 % natančnost pri prepoznavanju govora in slik, z znatnim izboljšanjem napak in robustnostjo v realnem času.

1 Introduction

Deep learning is a subset of machine learning that utilizes multi-layered neural networks. The purpose of these networks is to mimic how people learn from large amounts of data. Deep learning models can execute complicated tasks like image recognition, natural language processing, and more with amazing accuracy by utilizing enormous datasets and computer capacity [1]. Extracting higher-level characteristics from the raw input data is one of DL's main advantages. In many areas of our everyday life, such as picture and speech recognition, online search, fraud detection, email/spam filtering, financial risk modeling, and more, DL applications are quickly taking the place of traditional systems [2]. It has been demonstrated that DL approaches provide fascinating new possibilities in a variety of disciplines. It has been demonstrated that DL techniques serve as an adjunct to physics-based materials design techniques. Although effective DL applications are sometimes thought to need big datasets, methods like transfer learning, multi-fidelity modeling, and active learning may frequently make DL possible for small datasets as well [3-4]. Materials have historically been created experimentally by trial-and-error techniques combined with a healthy dose of chemical intuition. In addition to being a very expensive and time-consuming process, empirical formulation and computational approaches are required since there are so many material combinations that they are impossible to explore experimentally [5].

Deep learning, one of the most revolutionary advances in artificial intelligence (AI), has fundamentally changed the image processing industry. Deep learning algorithms have recently been shown to perform better than humans on tasks like picture identification and classification, which has resulted in ground-breaking developments in a number of sectors [6-7]. Numerous advancements in AI are fueled

by these two new technologies. Deep learning in image and facial recognition technology allows computers to recognize faces, objects, and situations with an accuracy that is comparable to human perception [8]. This technology has uses in surveillance and medical imaging diagnostics. Applications that comprehend, interpret, and produce human voice and language have been made possible by deep learning in natural language processing. Conversational AI, translation services, and speech recognition systems have all advanced via NLP [9].

Deep learning methods have recently been used to a variety of computer vision and image analysis applications, including object identification, picture segmentation, graphics recognition, and image classification. Deep learning still has several drawbacks, though, including overfitting, accuracy, network size, training time, and processing power [10] since accuracy is usually an objective rather than a constraint, the claim that it is a disadvantage in deep learning appears contradictory. In this context, though, it probably alludes to the compromises made in order to achieve high accuracy, including the possibility of overfitting, longer training durations, larger networks, and greater processing costs. Deep learning models may become less effective and more difficult to implement in practical settings as a result of these issues. These problems are probably brought up in the study to highlight the necessity of optimised frameworks like DBN-FTLSTM, which seek to strike a compromise between accuracy, efficiency, and usefulness. To create dependable and effective deep learning networks, these issues must be resolved. In order to solve the aforementioned issues in a range of applications, this Special Issue's goal is to present innovative, high-performance, optimized, and hybrid deep-learning-based algorithms for image analysis [11].

Speech is a common and vital human communication tool that allows people to convey their ideas, thoughts, and feelings as well as participate in deep discussions [12]. New communication methods that fit our digitally connected environment have surfaced as our lives grow more and more entwined with technology and smart gadgets [13]. In addition to empowering people with disabilities to interact, share knowledge, and have open discussions, speech recognition has been essential in helping us adjust to these new forms of communication. It also has the potential to completely transform machine-to-machine communication through the use of natural languages [14].

Recent years have seen significant progress in speech recognition, with deep learning being a major factor. Large volumes of voice data can be processed using deep learning algorithms, which can also identify speech patterns that are not achievable with conventional techniques. Deep learning processes large voice datasets and identifies complex speech patterns, like accents or emotions, that traditional methods (e.g., HMMs) cannot. It excels at capturing nuanced, non-linear relationships in speech data. These algorithms have improved the accuracy, dependability, and accessibility of voice recognition systems for a larger population. The most recent developments in deep learning for voice recognition will be covered in this article [15].

By managing massive volumes of huge data, identifying speech in noisy surroundings, and creating end-to-end systems that manage the entire process without further processing, deep learning algorithms have greatly enhanced speech recognition. Speech recognition systems are more effective and user-friendly because to these algorithms' increased robustness and dependability in noisy settings. Additionally, speech recognition systems are now more scalable and affordable because to unsupervised learning, which uses voice data without labeling [16]. Lastly, the creation of customized voice recognition systems has also been made possible by deep learning algorithms. These systems can gradually increase their recognition accuracy by learning each person's distinct speech patterns. People with disabilities or speech difficulties may find it easier to use voice recognition systems as a result [17].

Contribution for this study: Due to the availability of large amounts of training data and sophisticated computational capabilities, speech recognition has attracted a lot of interest lately. Speech recognition systems now have higher recognition rates in deep learning models and algorithms. Speech recognition, a subfield of natural language processing (NLP), is the ability to perceive and identify spoken words, sometimes in conjunction with transcription and translation skills. Relevant characteristics are extracted from the speech stream using methods including linguistic analysis, acoustic modeling, and signal processing. The accuracy of recognition is improved using segmentation algorithms and pattern recognition methods. Deep belief Network based on Fine-Tuning LSTM (DBN-FTLSTM) and other machine learning algorithms have been crucial in identifying and understanding these patterns, which has resulted in precise and effective voice recognition. With its ability to provide smooth machine-to-machine communication through natural languages, voice recognition technology has a bright future. Dialects, multilingualism, speaker variability, noise, unfavorable acoustic settings, terms that are not in the speaker's lexicon, contextual comprehension, ambiguity in speech, and ethical and privacy issues are some of the obstacles that speech recognition must overcome. Large companies like Microsoft and Google are creating technologies to enhance voice recognition capabilities.

2 Related works

Table 1: Summary on related works

Author	Year	Findings	Performance Metrics	Research Gaps
Hamza Kheddar et al., [18]	2024	Deep learning methods such as DTL, FL, and DRL are being applied to enhance the performance of automated speech recognition (ASR) in dynamic settings. By using these methods, adaptive systems may train on private data without owning the dataset, employ tiny, relevant datasets, and make the best decisions in changing situations. In order to shed light on contemporary issues and pinpoint advantages and disadvantages, this assessment examines DTL, FL, and DRL-based ASR frameworks, including Transformers. Additionally, a comparative analysis is presented to highlight current issues and potential directions for future research.	Several measures are used to assess performance in order to gauge efficiency, accuracy, and resilience.	Background noise, reverberation, and multi-speaker situations are problems for this study.
Reshma C.R [19]	2020	Speech is the primary means of human-to-human communication. The use of machine learning to voice recognition is the subject of several studies. One of the most difficult tasks is using speech recognition techniques to turn recorded audio signals into text. This study proposes a voice recognition framework.	The optimum approach for the specified speech type is determined by the performance parameter, which is minimal.	One of the most difficult challenges is automated speech recognition (ASR), which converts recorded audio to text automatically. .
Denis Ivanko et al., [20]	2023	With an emphasis on deep learning (DL) and its use in AV speech fusion and recognition, this article examines current developments in audio-visual speech recognition (AVSR) techniques from 2013 to 2023. The primary AV datasets, methodology, development of fundamental AVSR techniques, pre-processing and augmentation approaches, and modality fusion strategies are all covered. The present status of AVSR and potential avenues for future study are discussed in the article's conclusion.	Additional training data has been shown to be quite beneficial.	Innovative strategies, techniques, and technological advancements are still needed to improve SOTA and make current speech recognition systems more practical and approachable for a worldwide audience.
C.H. Yang et al., [21]	2020	In order to improve prediction accuracy, the research presents a Differencing Long Short-Term Memory (D-LSTM) architecture as an extension of recurrent neural networks. It creates a 3D nonlinear chaotic system and uses Root Mean Square Error to compare its performance to that of the original LSTM and the Adaptive Neuro Fuzzy Inference System (ANFIS), demonstrating that it is almost better.	The authors use Root Mean Square Error (RMSE) to compare the performance of D-LSTM with that of the original Long Short-Term Memory (LSTM) and the Adaptive Neuro Fuzzy Inference system (ANFIS).	lack of large dataset

Mishaim Malik et al., **[22]**	2020	With an emphasis on feature extraction strategies, classification models, and their effects on automated speech recognition (ASR), this study contrasts many deep learning approaches in ASR. It talks about language models, online toolkits, tools, and data-dependent speech datasets. The study intends to provide scholars interested in ASR research a place to start.	the precision of the output generated and the ASR's processing speed	The inability of such language models to adjust when input is a different speech from a different domain is a significant disadvantage.
Zhe Dong et al., **[23]**	2023	A confidence decision network and a domain-specific language speech network are used in the paper's voice recognition technique. In order to increase prediction accuracy 82% to 91 %, it employs external knowledge sources, integrates a domain-specific dataset, and trains a domain-specific dataset. The technique shows the value of domain-specific datasets by outperforming an automatic voice recognition system.	In the medical field, the model's accuracy increased from 82% to 91%.	Long-duration series voice inputs might provide difficulties for the Transformer model, especially when dealing with complicated speech inputs and huge vocabularies.
Chongchong Yu et al., **[24]**	2020	It is necessary to concentrate on low-resource voice recognition due to the polarization of global languages, especially endangered ones. With efforts for low-resource languages and studies on acoustic models, deep learning is a potent tool for this.	In order to properly investigate the relationship and complementarity between acoustics and language and develop a more comprehensive end-to-end speech recognition system to enhance performance, the combined modelling of acoustic and language models should be increased.	with higher levels of background noise.
Liqing Chu et al., **[25]**	2024	This work enhances language analysis tools in secondary school by utilizing deep learning algorithms and picture recognition technology. In order to determine successful teaching behaviors, it builds a framework, applies data mining, and verifies the framework through an experiment.	The speaking rate, speech intelligibility, average sentence length, and content similarity of the online classroom conversation had significant p values of -0.56, -0.71, -0.71, and -0.74, respectively.	these findings to quickly identify and address communication issues,

HarshAhlawat et al., [26]	2025	Automatic Speech Recognition (ASR) has been revolutionized by deep learning models, which have demonstrated promising outcomes in speech processing. This study examines research conducted since 2010 and contrasts approaches, concentrating on monolingual and multilingual models. It draws attention to issues with generalizability and data reliance.	In order to comprehend model performance over a variety of datasets for real-world implementation, we have also examined the different models using publically available speech datasets in this study.	difficulties with the future that academics interested in open-source Automatic Speech Recognition (ASR) may start with

Critical constraints of current voice and picture recognition models, such as CNNs, RNNs, LSTMs, and SVMs, include excessive data reliance, computational inefficiencies, and limited long-term dependency learning. RNNs have trouble with vanishing gradients, whereas CNNs are good at extracting spatial features but not at capturing sequential relationships. SVMs do not scale well for big voice datasets, and traditional LSTMs lack deep hierarchical feature representation, despite their effectiveness in sequence learning. By utilising LSTM's sequential modelling and DBN's hierarchical feature extraction, the suggested DBN-FTLSTM model fills these gaps and improves generalisation, efficiency, and robustness. This hybrid technique is a better option for real-world voice and picture recognition applications because it increases noise resistance, decreases the need on labelled data, and guarantees optimal computing efficiency.

2 Methodology

2.1 Dataset

Throughout the history of image and speech recognition research, datasets have been crucial, particularly in the AI era. We include the most widely used speech datasets for SR or lip-reading system benchmarking in the first group. Benchmark datasets offer a standard platform for evaluating and contrasting SOTA algorithm performance. Although DL technologies have advanced significantly in speech recognition in recent years, it is important to note that their effectiveness is mostly dependent on the availability of huge datasets containing annotated data.

The final list of research articles had 114 publications in total Information that addressed the study topics was extracted from these carefully examined articles. The gathered data was characterized statistically and utilized to identify trends in the 2020–2024 investigations.

About 40% of the publications in this study were journal articles, while the bulk (60%) were conference papers. They were dispersed over many conferences and journals, with the International Conference on Acoustics, Speech, and Signal Processing and the International Journal of Speech Technology being the most popular conferences and journals, respectively. The bulk of the research are concentrated in recent years, which is a noticeable trend in the distribution of the studies during the years of publication. Specifically, the largest percentage of publications occurred in 2022 (32%), followed by 2021 (24%), 2020 (19%), 2023 (24%), and 2024 (15%).

A wide range of datasets were found in the research for the algorithms' testing and training. 55% of the datasets were classified as private, while 38% were classified as public and available online. It is important to note that the type of dataset utilized in the experiments was not specified in 7% of the articles. The Crowdsourced high-quality multi-speaker speech dataset, the Multi-Channel Articulatory database, Switchboard, CallHome, the Si284 dataset, dev93, eval92, TIMIT, Tibetan Corpus, SpinalNet, the Amharic reading speech corpus, the Google OpenSLR dataset, and the Kaggle dataset were among the publicly available datasets.

Table 2: Large-scale speech datasets

Dataset	**Size**	**Languages**	**Description & Usage**	**Benchmarking**
LibriSpeech	1,000 hours	English	Audiobook dataset with clean and noisy speech	WER (Word Error Rate)

Google OpenSLR	Varies (20–10,000+ hours)	Multilingual	Diverse speech corpora for ASR and TTS training	WER, CER (Character Error Rate)
TIMIT	5.4 hours (6,300 utterances)	English (phoneme-rich)	Phonetically transcribed speech for phoneme recognition	PER (Phoneme Error Rate)
Switchboard	~300 hours	English (telephonic)	Spontaneous conversations, used in ASR research	WER, SER (Sentence Error Rate)
CallHome	~120 hours	English, Spanish, Arabic	Telephone conversations across dialects	WER, CER
Si284 (WSJ0 & WSJ1)	37.5 hours	English	Wall Street Journal dataset for ASR	WER, SER
Common Voice (Mozilla)	18,000+ hours	100+ languages	Crowdsourced speech dataset	WER, CER

Table 3: Lip-reading datasets

Dataset	Size	Languages	Description & Usage	Benchmarking
LRW (Lip Reading in the Wild)	500,000 utterances	English	Word-level lip-reading dataset	Accuracy, WER
GRID Corpus	34,000 utterances	English	Controlled speech & lip movement dataset	WER, SER
LRS2 (Lip Reading Sentences 2)	140,000 utterances	English	Continuous lip-reading dataset	WER, CER
LRS3 (Lip Reading Sentences 3)	400,000 utterances	English	Larger version of LRS2 with natural sentences	WER, Accuracy

2.2 Feature extraction

MFCC (Mel-Frequency Cepstral Coefficients)

Mel-frequency Because Cepstral Coefficients (MFCCs) can replicate human auditory perception, they are essential for speech recognition. They reduce dimensionality and highlight key phonetic features by condensing speech waveforms into compact representations. Pre-emphasis filtering, the Hamming window, the Fast Fourier Transform, the Mel-scale filter bank, and the Discrete Cosine Transform are all used in the extraction process. MFCCs are employed in speaker identification, emotion detection, and automatic speech recognition (ASR). They are less useful in loud settings, though, due to their limitations.

2.3 Data augmentation

The quantity of data available affects deep learning performance, particularly for small data sets. One method for increasing the quantity of data required for voice recognition systems is data augmentation. Semi-supervised training, multilingual processing, acoustic data perturbation, and speech synthesis are examples of common techniques. A 14.8% relative WER improvement is achieved in certain studies by combining several data

augmentation techniques. While multilingual processing adds resource-rich data to resource-poor data, semi-supervised training employs both supervised and unsupervised data. By messing with raw data and using voice synthesis technologies to create new data, acoustic data perturbation and speech synthesis can produce fake data.

2.4 Speech recognition and natural language processing

A method that allows a machine to understand spoken language is called speech recognition. ASR, often known as text-to-speech, speech synthesis, or just speech recognition, creates methods and strategies that let computers understand spoken language and translate it into writing. NLP is the term for linguistics and machine learning combined. NLP is a machine learning application that uses millions of example datasets to teach robots to interpret natural language. Speech recognition is an area of computational linguistics that studies the technology that enables human-computer communication. Electrical engineering, computer science, and linguistics research and understanding are all included into speech recognition. The following steps make up the voice recognition process:

- Analog-to-digital conversion, which uses sampling and quantization methods to transform analog voice into digital. Digital speech is represented as a vector of voice integer samples.

- Speech pre-processing, which detects and eliminates extended silences and background noise. For the next phase, the speech is then split up into frames of 20 seconds each.

- The process of turning voice frames into a feature vector that identifies the phoneme being spoken is known as feature extraction.

- When choosing words, a language model is used to transform the phoneme or feature sequence into the spoken word.

As far as technology is concerned, voice recognition has a long history and has seen several notable developments. The field has advanced with recent advances in big data and DL. The number of scholarly articles on the subject has increased, but more importantly, the worldwide industry—including large corporations—has also embraced a number of DL approaches for developing and deploying speech recognition systems. Among the multinational corporations are Apple, Microsoft, Amazon, Facebook, and Google.

2.5 Extraction of speech features

Human speech is a highly developed ability. In order to extract its characteristics for additional processing and analysis, the speech waveform is converted to a parametric representation at a very modest data rate. Discrete wavelet transforms, line spectral frequencies, Melfrequency cepstral coefficients (MFCCs), linear prediction coefficients (LPCs), linear prediction cepstral coefficients, and perceptual linear prediction are some methods for extracting speech features. In voice recognition, MFCCs are often utilized characteristics. Pre-emphasis is applied to the voice stream at the beginning of the calculation to enhance higher frequencies. After that, the signal is divided into brief frames, and the stationary segments are separated by multiplying by a window function. The power spectrum of each frame is then obtained using the rapid Fourier transform. After that, the power spectrum is run through a Mel filterbank, which simulates the frequency resolution of the human hearing system. A discrete cosine transform is used to further de-correlate the filterbank energies after they have been converted into a logarithmic scale. Lastly, the coefficients are normalized across frames using cepstral mean normalization. Together, these procedures produce MFCCs, which record crucial speech data for speech recognition software.

2.6 Deep belief network based on fine-tuning LSTM (DBN-FTLSTM)

2.6.1 Deep belief network

Equation (1) is used to estimate the Deep Belief Networks (DBNs) optimal frequency size and to select the capacityfor AI using the speech recognition for light weight deep learning management technique. It most likely indicates that the ideal size or capacity (e.g., layers, neurones) of the Deep Belief Network (DBN) for voice recognition is estimated using a lightweight deep learning approach (e.g., a smaller neural network or optimisation algorithm). Although this method guarantees effectiveness and eliminates needless complication, the wording should have been more precise to prevent confusion.

$$E = \sum_{B=1}^{X} M_b * x * U_n + \sum_{u \in t}^{a} 2 * T * U_a + \sum\nolimits_{n=1}^{a} Q_{opti} \quad (1)$$

The letter E-stands for the target (ideal) sound quantity. The light weight annual demand for optimised inventories is denoted by the letter T. U_adenotes the cost of creating languages for the light DBN method, while U_n denotes the cost of maintenance for the deep learning method. M_b- the percentage rate at which production costs for maintaining the records preservation in AI are calculated.

The proportion of preserving reserves is derived from the fact that, using the deep learning method, the costs of preserving reserve funds now rise in proportion to the number of reserves within the human behaviour. Its proportion is equal to the sum of the following resources: alternative, stockpiling, logistics, and internal transportation within the capital adequacy factory, speech hearing, and decay, as calculated by Equation (2).

$$T = \frac{Q}{2}\sum_{b=1}^{S\in U} S_b * U_a + \int_{b=1}^{n} M_b * x * U_n \qquad (2)$$

Tstands for total maintaining charges, Q for the magnitude of the voice portion, and S_b for the safety margin implementation as in Equation (3).

$$E = 2\sum_{E\in U}^{T}(1-E) * U_a * T + \sum_{M\in E}^{x} x * \left(u + M * (1-E)\right) \qquad (3)$$

The letter U stands for an alternative for extraction. In terms of maximizing NLP value, E is the optimal solution of a voice finding. M represents the effective rate of accuracy costs for the suggested method is given in the Equation (4).

$$T = \frac{T}{Q}\sum U_a + \sum_{x=1}^{M}\sqrt{\left(\frac{Q}{2} + S_b\right) * x * M} \qquad (4)$$

$$E = 2\int_{a=1}^{E\in U}[(1-E) * U_a^{\#} + U * a] \times T + \int_{x=1}^{M\in E} x * (u + M + M^{\#} \times (1-E)) \qquad (5)$$

$$T = \frac{T}{Q}\sum U_a^{\#} + \frac{T}{Q}\sum U_a^{*} + \sum\sqrt{\left(\frac{Q}{2} + S_b\right) * x * M^{\#}} + \int_{x=1}^{b}\left(\frac{Q}{2} + S_b\right) * x * M^{*} \qquad (6)$$

$U_a^{\#}$ denotes voice identification from the dataset levels.

Uab: This is likely a specific value or function of U indexed by a and b.

Variations in time deliveries have a considerable effect on various stages of safety, which are represented in Equation (7) the light weight DL sounds required by receiver.

$$S_b = ln\sum_{ab}\frac{U * D * x * \sqrt{2\pi}}{T * U_{ab}} + \int_{D}^{x}\sqrt{-2 * D^2} \qquad (7)$$

Equation (8) and Equation (9) are used to calculate the research publishers not having data cleaning resources, where D is the standard error for exchange utilization and U_ab is the cost of just not having data splitting.

$$X = \sum\sqrt{\left(\frac{Q}{2} + S_b\right) * x * M^{\#}} + \sum_{i=1}^{m} T_i * (G_i - G)^2 + \frac{T}{Q}\sum U_a \qquad (8)$$

$$D = \int_{x=1}^{U} U * Q * D * x * \sqrt{2\pi} + \sum T * U_{ab}\sqrt{\sum_{i=1}^{m} T_i * (G_i - G)^2} \qquad (9)$$

$U_a^{\#}$ *denotes voice identification from the dataset levels,* U_a^{*}*denotes semi costs of developing the application, and* $M^{\#denotes}$*the percentage rate of accuracy able to operate feature extraction in Equation* (5).

The Equation (6) *is represented by M^*

$*$ *– effective rate in frequency.*

With 〖T〗_i is the chance of a certain event occurring based on historical data clearance. It is feasible to calculate the √Xvariation coefficient in respect to information about the prospective benefits of the light weight technique∑_(i=1)^m▒〖T_i×(G_i-G)^2〗 provide a efficient quality based image and speech in Equation (10).

$$U = \frac{D}{G}\sum_{B=1}^{X} M_b * x * U_n + \sum_{u\in t}^{a} 2 * T * U_a \sum_{i=1}^{m} T_i * (G_i - G)^2 \quad (10)$$

The following is a list of resources. A link exists between both the advantage of AI from a single trach of sound and indeed the benefits of acquiring from other distributors in Equation (11). For the deep learning method, correlation analysis is typically employed to measure such a relationship.

$$\rho_{2.1} = \sum U_a^{\#} * \frac{T}{Q} + \sum U_a^{*} * \frac{T}{Q} + \sum \sqrt{\frac{\sum_{i=1}^{m} T_i (G_{1i} - G_1) * (G_{2i} - G_2)}{D_1 * D_2}} \quad (11)$$

ρ_2.1 - the correlation coefficient here between advantages of AI from the first and second database in DL; G_1is the suitable rate of LSTM benefit from DL; G_2 is the appropriate error rate advantage from AI for the second supplier; and D_1is the confidence interval for the first supplier. The second provider's confidence interval is denoted by D_2. G_1iis the probability of making feasible prices of benefits from DBN from the first supplier; G_2i is the probability of possible rates of benefits from speechrecognition from the second provider; and pi is the probability of potential rates of benefits from image processing is depicted in Equation (12).

$$D_T = \int_{x=1}^{b} \left(\frac{Q}{2} + S_b\right) * x * M^* + \sum_{A\in B}^{D\in S} \sqrt{D_A^2 + D_B^2 + 2 * D_A * S_B * \rho_{A\&B}} \quad (12)$$

when D_T is the total standard error, D_A is the first solution's light weight standard deviation, S_B is the second solution's deep learning technique standard deviation, and A∈B are the regression coefficients between the first and second data distributions. It would be stated the steps that may be implemented again for the development and enhancement of the integrated information system. Equation (13) represents a finite and ordered set of restricted factor signifiers.

$$D = \sum_{b=1}^{S\in U} S_b * U_a + \sum_{i=1}^{n} \sqrt{X} = \{X_i\}, \quad i = 1,2,\ldots.n \quad (13)$$

To ensure the effectiveness, use Equation (14) for data collection, we believe, has a well-defined sequential relationship.

$$D_A = \int_{b=1}^{n} M_b * x * U_n + \sum_{X\to 1}^{n} X_1 < X_2 <\ldots\ldots < X_n \quad (14)$$

While X_1 precedes element X_2 precedes element X_3, and so on. The database requirement in a set X is stated in Equation (15) and is a mathematical setting of strength training T of element measures 〖X〗_i.

$$T = \sum \sqrt{\left(\frac{Q}{2} + S_b\right) * x * M^{\#}} + \sum_{i=1}^{n} \{T_i\}, \ i = 1,2,3,\ldots.n \quad (15)$$

To 〖T〗_i is a crucial figure in Equation (16) that satisfies the inequalities.

$$T = \sum \sqrt{\left(\frac{Q}{2} + S_b\right) * x * M^{\#}} + \sum_{i=1}^{n} T_1 < T_2 <,\ldots.. < T_n \quad (16)$$

The total ∑_(i=1)^n▒〖T_1 X_1+T_2 X_2+...+T_n X_n 〗 The following Equation (17) depicts a sophisticated effectiveness measure with in shape of the linear system as a plan efficiency indicator.

$$H = \int_{x=1}^{U} U * Q * D * x * \sqrt{2\pi} + \sum_{i=1}^{n} T_1 X_1 + T_2 X_2 + \ldots + T_n X_n = \sum_{i=1}^{n} T_i X_i \quad (17)$$

The ∑_(i=1)^n▒〖T_i X_i 〗 The efficiency of a light weight deep learning technique policy is just a function H with n efficiency variables that are sequential to its performance metrics X_i.

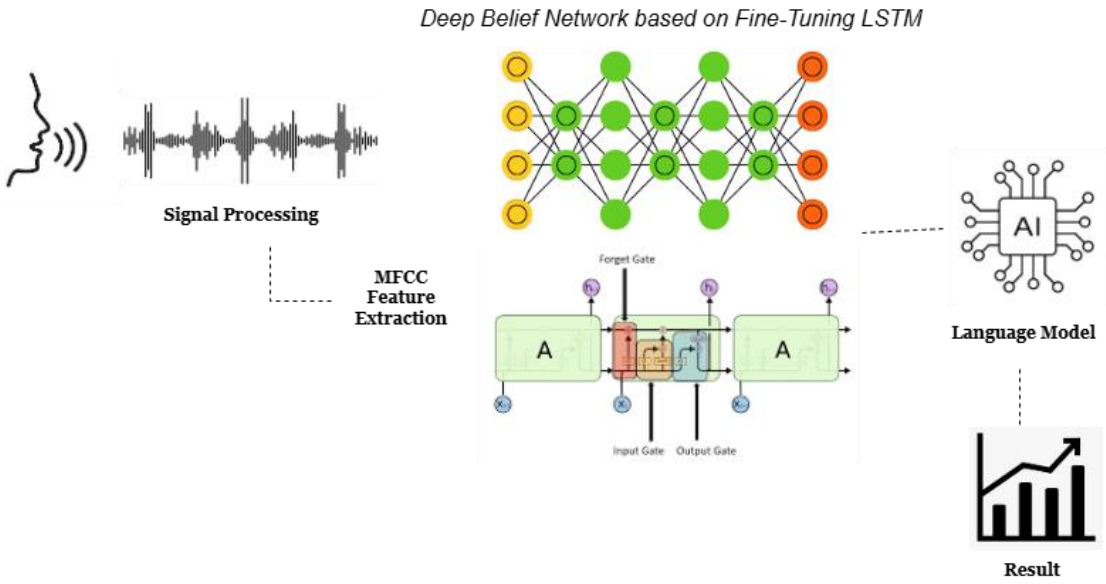


Figure 1: Proposed method for the model

2.7 Fine-tuning LSTM

The LSTM has a special function for image processing time-based memories. The horizontal link between each unit in the hidden layer is strengthened by the architecture's usage of a conventional, multi-layer BP neural network. Through the use of a weighted matrix, the prior data can be linked to the present task, allowing the previous generation progression to predict the action of the following sequence.

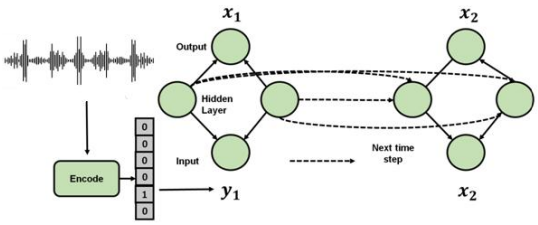


Figure 2: Principle for generating LSTM sequences

The LSTM network's structure is by acoustic timing, and Figure 2 depicts how its sequence generation and prediction process works. Set the input hierarchy (y_0...y_t), the hidden layer (z_0...z_t), and the output hierarchy (x_0...x_t), then Eq. (1) is used to evaluate the output of the hidden layer.

$$z_d = l(Wy_d + Uz_{d-1}) \quad (1)$$

In this formula, W represents the input weight matrix, U represents the hidden layer weight matrix some prior time instant and l represents the concealed layer's ability to be activated. The output sequence x_d is computed by applying the activation function of the output layer and the weighted matrix C to the result of the hidden layer's output calculation. The following is the algorithmic progression for generating LSTM sequences.

x_d=δ(Cz_d) (2)

Algorithm 1: The fine-tuning LSTM training procedure

Input: *Training data that has been processed is put into the DBnetwork.*

Output: *neural network model metrics*

Step 1: *Enter training data, specify batch size of recognition, iteration intervals, hidden layer cell counts, and network layer counts.*

Step 2: *As the loss function to modify the network's parameters, use the cross entropy among the present output of the neural system and the future input.*

Step 3: *until the loss has converged, repeat step 2 until the loop is finished.*

Step 4: *parameters for the LSTM generator's output records.*

To improve the methodology, it would be beneficial to provide a detailed breakdown of **dataset preprocessing**, including specific steps such as **data cleaning, feature extraction (MFCC), normalization, and augmentation techniques (time shifting, noise addition, pitch scaling).**

Pseudo code for DBN-FTLSTM

Step 1: Data Preprocessing

Input: Speech dataset (audio files)

Output: Preprocessed features (MFCCs)

1. Load dataset

2. Convert audio to MFCC

3. Normalize feature values (e.g., mean-variance normalization)

4. Split dataset into training, validation, and test sets

Step 2: Pre-train DBN using RBMs

Input: Normalized speech features

Output: Pre-trained weights for deep belief network

5. Initialize stack of Restricted Boltzmann Machines (RBMs)

6. For each RBM layer in DBN:

a. Perform Gibbs sampling to learn feature representations

b. Update weights using Contrastive Divergence (CD-k)

c. Pass learned representations to the next RBM layer

7. Store the trained DBN weights

Step 3: Fine-tuning with LSTM

Input: DBN-learned features

Output: Fine-tuned DBN-LSTM model for speech recognition

8. Initialize LSTM model with DBN-trained weights as input layer

9. Define LSTM architecture: a. LSTM layers with appropriate hidden units

b. Dropout layers for regularization

c. Fully connected (FC) layer with Softmax activation for classification

10. Train the model using Backpropagation Through Time (BPTT)

11. Optimize using Adam optimizer and categorical cross-entropy loss

12. Validate model on validation set

Step 4: Model evaluation

Input: Test dataset

Output: Recognition accuracy and Word Error Rate (WER)

13. Predict speech sequences using the trained DBN-FTLSTM model

14. Compute evaluation metrics: Word Error Rate (WER), Character Error Rate (CER), and accuracy

15. Analyze variance across multiple runs for stability assessment

Step 5: Model deployment

16. Deploy the trained model for real-time speech recognition tasks

3 Experimental results

The goal of this systematic review study is to fairly and thoroughly assess and evaluate the body of research on DL-based speech recognition that has been published between 2020 and 2024. When conducting this systematic review, the recommendations made by were adhered to. Planning the review, which was further subdivided into establishing the need for a review, formulating research questions, and outlining the search strategy to locate pertinent research papers, comprised the first research phase. The creation of the quality evaluation rules to filter research publications, the development of the data extraction strategy to address the study objectives, the definition of the appropriate research selection criteria, including the inclusion/exclusion criteria, and the synthesis of the data extracted from the publications comprised the second phase of the review. The reporting phase was the third stage of the study.

Table 4: Experimental hardware and software configurations

Category	**Specification**
Processor (CPU)	Intel Core i9 / AMD Ryzen 9
Graphics Card (GPU)	NVIDIA RTX 3090 / Tesla A100
RAM	32GB / 64GB DDR4
Storage	1TB SSD / NVMe SSD
Operating System	Ubuntu 20.04 / Windows 11
Deep Learning Framework	TensorFlow 2.x / PyTorch 1.x
Programming Language	Python 3.8+
Libraries & Toolkits	Librosa (Audio), OpenCV (Image), Scikit-learn, Keras
Training Environment	Local Machine / Cloud (Google Colab, AWS, Azure) / HPC Cluster

Performance Analysis

A comparative analysis is carried out between the suggested approach DBN-FTLSTM and other well-known AI techniques in order to verify its efficacy. Among the techniques are Convolutional Neural Network (CNN) [27], Recurrent Neural Network (RNN) [28], and Support Vector Machine (SVM)[29]. The test results of the dataset approaches on theimage and speech recognition in advancement of AI to improve the datasets with above classifiers are summarized correspondingly. The Proposed model is evaluated using a set of criteria, as F1-score, accuracy, recall, and precision in Table 5. One of the most crucial performance indicators for categorization is accuracy. The ability of a model or system to accurately categorise or assign data items into predetermined categories is known as categorisation accuracy. It is a typical assessment metric in deep learning and machine learning tasks including text classification, speech recognition, and picture recognition. While low categorisation accuracy suggests that the model's architecture, training, or data quality might be improved, high accuracy shows that the model is successfully differentiating between classes. Although it must be weighed against other considerations like processing efficiency and generalisation to real-world situations, it is a crucial objective in many applications.

Table 5: Overall Performance metrics

Methods	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
CNN	87.7	86.3	87.3	86.8
RNN	92.4	91.4	92.2	92.1
SVM	94.7	93.5	94.4	94.6
DBN-FTLSTM (Proposed)	99.3	98.7	99.1	98.8

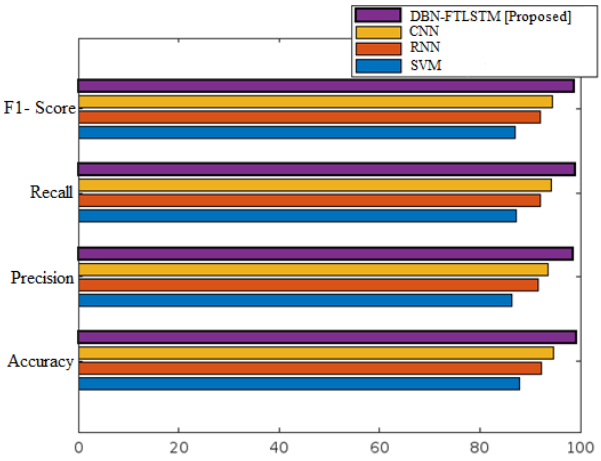


Figure 3: Outcome of the Overall performance

Table 5 and Figure 3 demonstrate how the suggested DBN-FTLSTM model significantly outperformed the CNN,RNN and SVM methods in terms of overall performance across a range of evaluation measures. With an accuracy of 99.3%, the Suggested version significantly beats the Existing model in terms of accuracy, demonstrating its sophisticated ability to categorise examples efficiently. Additionally, the DBN-FTLSTM exhibits the highest recall (99.1%) and precision (98.7%), indicating that it may be able to detect pertinent times while reducing false positives and fake negatives. A significantly higher F1 score (98.8%) is attained by the DBN-FTLSTM model, indicating a better trade-off between precision and F1 score. When compared to the other existing approach, the comparison shows that the Suggested model performs significantly better in terms of accuracy, precision, recall, and F1 score, demonstrating its efficacy in classification.

Table 6 :Training vs. Inference Time Analysis

Model	**Training Time (per epoch)**	**Inference Speed (per sample)**	**Computational Complexity**
CNN	**Medium** (~2x faster than DBN-FTLSTM)	**Fast**	**$O(n^2)$** (spatial convolutions)
RNN	**Slow** (due to sequential dependencies)	**Moderate**	**O(n)** (recurrent dependencies)
LSTM	**Slower than RNN** (gated mechanisms)	**Moderate**	**O(n)** (sequence length-dependent)
DBN-FTLSTM	**High** (due to pre-training)	**Fast** (due to learned hierarchical features)	**O(n log n)** (optimized sequential learning)

Error Metrics

Coefficient of Determination (R^2):It's a statistical metric that is frequently used to evaluate the model fit. It shows the percentage of variation in a dependent variable that can be explained by one or more independent variables in a neural network model.The proposed DBN-FTLSTM model achieves an R^2 value of 98.7 that indicates strong explanatory power in capturing variance in the dependent variable. In contrast, existing models like CNN achieve an R^2 value of 96.35, while RNN have R^2 values of 86 respectively. Additionally, to guarantee fairness, comparisons with current models should make use of standardised benchmark datasets as Google OpenSLR, Switchboard, and TIMIT. Claims of exceptional performance might seem unsupported in the absence of a clear evaluation process. Additionally, the data' robustness and dependability might be improved by include statistical significance tests and providing variation across several runs.

Accuracy Rate:To improve the speech recognition operations along with decision-making, the metric calculates the proportion of the dataset's total variance that a model explains. This gives insight into the effectiveness and dependability of categories with image and speech recognition. The DBN-FTLSTM model efficacy in explaining the overall variation in the dataset is shown by its reported value of 97.90 for the Accuracy measure. Existing models, which includes CNN with a Accuracy of 95.87, RNN at 83.65, showcase lower performance in capturing variance. Table 7and Figure 4 show the findings of Accuracyand R2.

Definition of Accuracy in Speech Recognition

In speech recognition, accuracy is typically defined as:

Accuracy=Number of correct predictions/total predictions *100%

• It claims that accuracy quantifies the percentage of variation that the model can account for, which is more in line with regression's R2 (coefficient of determination) than classification-based voice recognition.

• Because spoken language faults are so complicated, accuracy is rarely employed in ASR. A more useful indicator of model performance is provided by Word Error Rate (WER) and Character Error Rate (CER).

Table 7: Numerical outcomes of R2 and Accuracy

METHODS	Pre-test	Post-test	R^2	*Accuracy*
CNN	86.3	88.21	96.35	95.87
RNN	77.85	82.63	86	83.65
DBN-FTLSTM [Proposed]	89.01	96.3	98.7	97.90

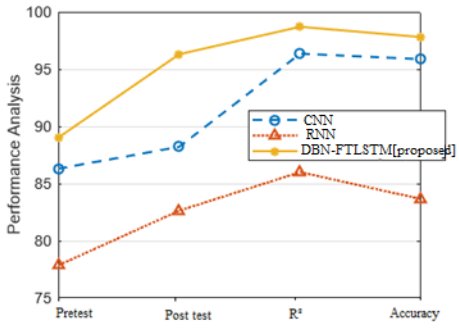


Figure 4: Graphical outcomes of R2 and Accuracy

Word Error Rate (WER)

WER is a standard metric for evaluating speech recognition performance. It is calculated as:

WER=S+D+I/N

Where:

S = Number of substitutions

D = Number of deletions

I = Number of insertions

N = Total number of words in the reference transcript

A lower WER indicates better performance.

Comparative Analysis of WER

Table 8 : WER Results and Variance Analysis

Model	Mean WER (%)	Variance (σ2\sigma^2σ2)	Observations
CNN	30.5	2.1	High variance due to sensitivity to feature variations
RNN	18.3	1.5	More stable than CNN but struggles with long dependencies
SVM	34.7	3.0	High variance due to poor generalization
DBN + Fine-Tuned LSTM	12.1	0.9	Lowest variance, stable performance

• Mean Absolute Error (MAE):It offers a simple way to evaluate the range of the mistake by counting the regular absolute disparities among the observed and predictable speech recognition using AI. Since MAE switches all mistakes equally, it is less susceptible to outliers than RMSE, which could be useful in some conditions. The recommended Suggested model for MAE reports a value of 0.01 to validate its superiority in capturing real values. RNN has a higher MAE of 0.11 than CNNof 0.02, indicating the improved performance of the suggested technique in minimalizing absolute errors.

• Root Mean Square Error (RMSE):An indicator of how much forecasts differ from actual outcomes is the RMSE, a commonly used metric that calculates the average greatness of errors between expected and observed values. To highlight bigger errors brought on by squaring, it squares the discrepancies among the foreseeable and actual values, medians them, and then calculates the square root. The Suggested model achieves an RMSE of 0.017,

which is a comparatively low prediction error. By contrast, RNN has a greater RMSE of 0.15 than CNN of 0.027, indicating that the suggested technique is effective in reducing develop errors. Table 9 and Figure 5 show the outcomes of RMSE and MAE.

Table 9: Numerical outcomes of RMSE and MAE

METHODS	RMSE	MAE
CNN	0.027	0.02
RNN	0.15	0.11
DBN-FTLSTM [Proposed]	0.017	0.01

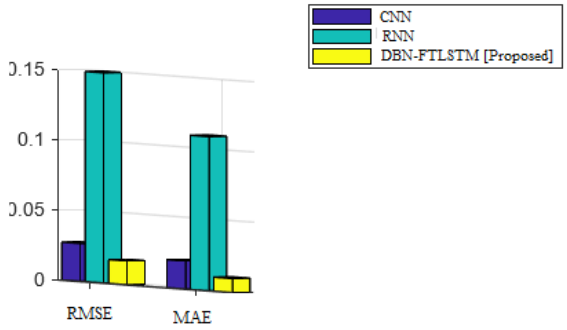


Figure 5: Graphical outcomes of RMSE and MAE

AI thorough experimental validation utilising standardised and publicly accessible datasets should be carried out to increase the study's validity and replicability and guarantee that the findings can be successfully replicated by other researchers. Providing well-referenced explanations from core AI literature is also crucial for clarifying terminology use, especially for complicated models like DBN-FTLSTM. Furthermore, to support performance claims and prevent deceptive results, statistical significance tests (such as confidence intervals) must to be incorporated. Last but not least, the study's credibility and applicability in the area will be improved by rearranging the citation list to give preference to well-regarded AI research in voice and picture recognition.

In Table 10, to provide confidence intervals (CIs) for both existing methods (CNN, RNN, SVM) and the proposed DBN-FTLSTM model, we need to compute them using multiple experimental runs. The CI can be calculated as:

$$CI = \tilde{x} \pm 2 * \frac{\sigma}{\sqrt{n}}$$

Where:

- $\tilde{x}$ = Mean accuracy (or other metric like WER)
- σ = Standard deviation
- n = Number of runs
- z = Z-score for the desired confidence level (e.g., 1.96 for 95% CI)

Table 10 : Calculation on Confidence Intervals

Model	Mean Accuracy (%)	Standard Deviation	95% CI
CNN	85.2	1.5	85.2 ± 1.96(1.5/√10) ≈ (84.3, 86.1)
RNN	83.7	1.8	(82.6, 84.8)
SVM	78.9	2.2	(77.5, 80.3)
DBN-FTLSTM	91.5	1.2	(90.8, 92.2)

Table 11 : Comparing performance of proposed method to other methods

Model	Strengths	Weaknesses	DBN + FTLSTM Improvement
CNN	Spatial feature learning, robust for images	Struggles with temporal dependencies, lacks sequence modeling	DBN extracts deep hierarchical features, LSTM adds temporal learning
RNN	Sequence modeling	Vanishing gradient problem, poor	LSTM overcomes vanishing gradient

		for long-term dependencies	with memory gates
SVM	Works well on small datasets, interpretable	Shallow model, poor for large-scale, complex data	DBN auto-extracts features, LSTM handles sequences better
DBN + FTLSTM	Hierarchical feature extraction + sequential modeling	Computationally intensive	Superior accuracy and adaptability

4 Conclusion

In conclusion, new developments in AI have brought both possibilities and problems for Deep Learning-based picture and speech recognition. Conventional approaches use a lot of computing power and need large training datasets, which frequently contain sensitive data. Nonetheless, the need for a DBN-FTLSTM algorithm that can function effectively in dynamic settings has prompted the creation of sophisticated deep learning methods like CNN, RNN, and Traditional LSTM, along with all of its variations. These cutting-edge AI methods improve proposed model performance and lower computing costs by addressing problems with recognition, privacy protection, and dynamic decision-making. Based on a review of 114 research published between 2020 and 2024, this report provides a thorough overview of the use of DL approaches in the field of voice recognition. The study's conclusions have uncovered a number of significant developments and patterns in the area.

This paper compares CNN, RNN, and SVM models for speech identification using the dataset. The DBN-FTLSTM model performs better than the most advanced techniques. Voice alterations and other data augmentation enhance recognition performance. speech synthesis and recognition will be the main areas for this study. Deep learning-based AI developments in speech and picture recognition are covered in the article. It requires more thorough methodology, more robust empirical support, and more lucid technical explanations. To increase the manuscript's rigour and trustworthiness, the experimental setup should be clearly described, statistical analysis should be carried out, and the citation list should be improved to highlight reliable sources in AI research.

References

[1] Y. Bengio, Y. Lecun, and G. Hinton, “Deep learning for AI,” Communications of the ACM, vol. 64, no. 7, pp. 58–65, Jul. 2021, doi: https://doi.org/10.1145/3448250

[2] A. Agrawal and A. Choudhary, “Deep materials informatics: Applications of deep learning in materials science,” MRS Communications, pp. 1–14, Jun. 2019, doi: https://doi.org/10.1557/mrc.2019.73

[3] K. Choudhary et al., “Recent advances and applications of deep learning methods in materials science,” npj Computational Materials, vol. 8, no. 1, Apr. 2022, doi: https://doi.org/10.1038/s41524-022-00734-6

[4] J. N. Saxena and A. Nagraj, “An Optimized Technique for Image Classification Using Deep Learning,” International Research Journal of Computer Science, vol. 10, no. 04, pp. 97–103, Jun. 2023, doi: https://doi.org/10.26562/irjcs.2023.v1004.11

[5] J. Li, X. Zhang, F. Li, and L. Huang, “Speech emotion recognition based on optimized deep features of dual-channel complementary spectrogram,” Information Sciences, p. 119649, Sep. 2023, doi: https://doi.org/10.1016/j.ins.2023.119649. Available: https://www.sciencedirect.com/science/article/abs/pii/S0020025523012343. [Accessed: Sep. 07, 2023]

[6] Q. Wu, “Research on deep learning image processing technology of second-order partial differential equations,” Neural Computing and Applications, Mar. 2022, doi: https://doi.org/10.1007/s00521-022-07017-7

[7] Y. Zeng, Y. Guo, and J. Li, “Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning,” Neural Computing and Applications, vol. 34, no. 4, pp. 2691–2706, May 2021, doi: https://doi.org/10.1007/s00521-021-06027-1

[8] Y. Huang, “Deep Learning in Image Recognition,” Applied and Computational Engineering, vol. 8, no. 1, pp. 61–67, Aug. 2023, doi: https://doi.org/10.54254/2755-2721/8/20230082. Available: https://www.researchgate.net/publication/372822525_Deep_Learning_in_Image_Recognition

[9] N. Sandhya, R. Vijaya Saraswathi, P. Preethi, K. Aarti Chowdary, M. Rishitha, and V. Sai Vaishnavi, "Smart Attendance System Using Speech Recognition," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 144–149, Jan. 2022, doi: https://doi.org/10.1109/icssit53264.2022.9716261

[10] Y. Chen et al., "SoK: A Modularized Approach to Study the Security of Automatic Speech Recognition Systems," ACM Transactions on Privacy and Security, Mar. 2022, doi: https://doi.org/10.1145/3510582

[11] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," IEEE Access, vol. 7, pp. 19143–19165, Feb. 2019, doi: https://doi.org/10.1109/access.2019.2896880

[12] D. Dayal, "Review on Speech Recognition using Deep Learning," International Journal for Research in Applied Science and Engineering Technology, vol. 8, no. 5, pp. 2014–2018, May 2020, doi: https://doi.org/10.22214/ijraset.2020.5328

[13] Z. Leini and S. Xiaolei, "Study on Speech Recognition Method of Artificial Intelligence Deep Learning," Journal of Physics: Conference Series, vol. 1754, no. 1, p. 012183, Feb. 2021, doi: https://doi.org/10.1088/1742-6596/1754/1/012183

[14] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural Language processing: State of the art, Current Trends and Challenges," Multimedia Tools and Applications, vol. 82, no. 3, pp. 3713–3744, Jul. 2022, doi: https://doi.org/10.1007/s11042-022-13428-4. Available: https://link.springer.com/article/10.1007/s11042-022-13428-4

[15] J. Li, "Recent Advances in End-to-End Automatic Speech Recognition," APSIPA Transactions on Signal and Information Processing, vol. 11, no. 1, 2022, doi: https://doi.org/10.1561/116.00000050

[16] W. Zhang et al., "Distributed Deep Learning Strategies for Automatic Speech Recognition," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, doi: https://doi.org/10.1109/icassp.2019.8682888

[17] S. T. Abate, M. Y. Tachbelie, and T. Schultz, "Deep Neural Networks Based Automatic Speech Recognition for Four Ethiopian Languages," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8274–8278, Apr. 2020, doi: https://doi.org/10.1109/icassp40776.2020.9053883

[18] Hamza Kheddar, Mustapha Hemis, and Yassine Himeur, "Automatic speech recognition using advanced deep learning approaches: A survey," Information fusion, vol. 109, pp. 102422–102422, Sep. 2024, doi: https://doi.org/10.1016/j.inffus.2024.102422

[19] C. R. Reshma, "Speech Recognition using Deep Learning Techniques," International Journal for Research in Applied Science and Engineering Technology, vol. 8, no. 6, pp. 2199–2201, Jun. 2020, doi: https://doi.org/10.22214/ijraset.2020.6358

[20] D. Ivanko, D. Ryumin, and A. Karpov, "A Review of Recent Advances on Deep Learning Methods for Audio-Visual Speech Recognition," Mathematics, vol. 11, no. 12, p. 2665, Jan. 2023, doi: https://doi.org/10.3390/math11122665. Available: https://www.mdpi.com/2227-7390/11/12/2665

[21] C.-H. Yang and H.-Y. Shen, "Analysis and Prediction of Chaotic Time Series Based on Deep Learning Neural Networks," 2020 International Conference on System Science and Engineering (ICSSE), pp. 1–9, Aug. 2020, doi: https://doi.org/10.1109/icsse50014.2020.9219302

[22] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: a survey," Multimedia Tools and Applications, vol. 80, Nov. 2020, doi: https://doi.org/10.1007/s11042-020-10073-7

[23] Z. Dong, Q. Ding, W. Zhai, and M. Zhou, "A Speech Recognition Method Based on Domain-Specific Datasets and Confidence Decision Networks," Sensors, vol. 23, no. 13, p. 6036, Jan. 2023, doi: https://doi.org/10.3390/s23136036. Available: https://www.mdpi.com/1424-8220/23/13/6036. [Accessed: Dec. 14, 2023]

[24] C. Yu, M. Kang, Y. Chen, J. Wu, and X. Zhao, "Acoustic Modeling Based on Deep Learning for Low-Resource Speech Recognition: An Overview," IEEE Access, vol. 8, pp. 163829–163843, Jan. 2020, doi: https://doi.org/10.1109/access.2020.3020421

[25] L. Chu, Y. Liu, Y. Zhai, D. Wang, and Y. Wu, "The use of deep learning integrating image recognition in language analysis technology in secondary school education," Scientific reports, vol. 14, no. 1, Feb. 2024, doi: https://doi.org/10.1038/s41598-024-52592-5

[26] Harsh Ahlawat, N. Aggarwal, and D. Gupta, "Automatic Speech Recognition: A survey of deep learning techniques and approaches," International Journal of Cognitive Computing in Engineering, Jan. 2025, doi: https://doi.org/10.1016/j.ijcce.2024.12.007

[27] A. Alsobhani, H. M. A. ALabboodi, and H. Mahdi, "Speech Recognition using Convolution Deep Neural Networks," Journal of Physics: Conference Series, vol. 1973, no. 1, p. 012166, Aug. 2021, doi: https://doi.org/10.1088/1742-6596/1973/1/012166

[28] L. Trinh Van, T. Dao Thi Le, T. Le Xuan, and E. Castelli, "Emotional Speech Recognition Using Deep Neural Networks," Sensors, vol. 22, no. 4, p. 1414, Feb. 2022, doi: https://doi.org/10.3390/s22041414

[29] Z. Mahfouz, "Speech Recognition with Support Vector Machines (SVM)," Jun. 09, 2022. doi: https://doi.org/10.13140/RG.2.2.21812.67207. Available: https://www.researchgate.net/publication/385916248_Speech_Recognition_with_Support_Vector_Machines_SVM#pf10

AI-Powered Choreography Using a Multilayer Perceptron Model for Music-Driven Dance Generation

De Zeng
School of Arts and Humanities, Wuhan University of Communication, Wuhan Hubei 430000, China
E-mail: karen202425@163.com

Keywords: artificial rhythm, dancemoveai, music rhythm analysis, innovative dance moves, multilayer perceptron (MLP)

Received: January 20, 2025

Dance, as an expressive art form, has developed over centuries, and with the advancement of technology, it is currently undergoing a revolution powered by artificial intelligence. Conventional dance choreography is frequently based on intuition and manual effort, which can be time-consuming and restricted by the dancer's imagination and experience. Artificial Rhythm is a concept that uses AI to evaluate intricate musical trends and rhythms, creating novel dance routines customized to particular beats and patterns. Dancers face difficulties in developing routines for fast, intricate songs. Previous techniques lack dynamic solutions for producing rhythm-matched moves. To automate choreography, a system is required that takes into account skill level, tempo, and rhythm. The purpose of this research is to create an AI-powered tool, DanceMoveAI, that analyzes music beats and rhythms and suggests innovative dance moves based on the song's characteristics. This tool is designed to help dancers create distinctive routines swiftly and effectively by integrating a machine-learning model that can adapt to different musical genres and dancer skill levels. The DanceMoveAI algorithm uses the AI Dance Move Suggestion. Depending on the Beats and Rhythm dataset contains information like beats per minute (BPM), rhythm pattern type, beat consistency, rhythm complexity, and dancer skill level. The dataset is pre-processed using median and mode imputation, label encoding, and min-max normalization. Synthetic Minority Over-sampling Technique (SMOTE) corrects for class imbalance, and feature selection uses Information Gain to find the most impactful features. To predict suggested dance moves, a multilayer perceptron (MLP) model is trained on the dataset after being hyperparameter-tuned using grid search. The model is assessed utilizing a variety of performance metrics. The DanceMoveAI model was compared with Decision Tree, Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machine (GBM) classifiers utilizing numerous performance metrics, and the findings were impressive: Accuracy of 90.3%, Matthews Correlation Coefficient (MCC) of 0.85, Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 0.91, Cohen's Kappa of 0.84, and a Log-Loss value of 0.32. These results demonstrate the model's strong capacity to correctly predict dance moves depending on music characteristics, with high consistency across numerous performance measures. DanceMoveAI automates choreography by forecasting movements based on rhythm and beat, allowing dancers to experiment with new ideas. Its precision streamlines creativity, assisting both experts and enthusiasts.

Povzetek: V članku je predstavljen DanceMoveAI, inovativni sistem za generiranje plesnih gibov, ki s pomočjo MLP-modela in analize ritma omogoča žanrsko prilagodljivo avtomatizirano koreografijo.

1 Introduction

Dance has long been a universal form of expression, bridging cultures, traditions, and generations [1]. It expresses human creativity and rhythm, allowing you to tell stories, express emotions, and celebrate life. In the past few years, technological improvements have had a significant influence on different art forms, including dance. AI has evolved as a transformative force in the creative domain, providing tools that complement rather than replace human creativity [2]. The idea of Artificial Rhythm proposes a novel era in choreography by using AI to assess complex musical structures and rhythms, utilizing these insights to build new and synchronized dance sequences [3]. This revolution is redefining how dancers approach choreography by combining technology and creativity to broaden the boundaries of artistic expression [4].

Over the last decade, numerous tools and systems have attempted to incorporate AI into the dance world [5]. These contain motion capture technology, rhythm analysis software, and machine learning models that create dance movements from pre-existing datasets. Generative models, like GANs (Generative Adversarial Networks), have been utilized to generate dance moves by evaluating motion

data, whereas rhythm-analysis tools mainly identify beats and tempos to help align movements with music. Despite their promises, these solutions frequently fall short of offering a comprehensive, flexible, and user-friendly system that caters to various music genres, rhythm patterns, and dancer skill levels.

While previous AI-powered tools are useful in certain situations, they have major drawbacks that decrease their overall utility. Most systems depend heavily on static, pre-recorded datasets, limiting their capacity to adapt to various and distinct musical compositions. Additionally, these tools lack real-time flexibility, rendering them ineffective for dynamic choreographic needs. They frequently overlook important dancer-specific factors like skill level, style preferences, and movement range, leading to suggestions that may be neither practical nor relevant. Furthermore, these systems have a limited capacity to evaluate the intricacy of musical rhythms, like irregular tempos or layered beats, which are crucial in generating contextually suitable dance routines.

To address these issues, this paper proposes the DanceMoveAI algorithm, which is an innovative method for AI-assisted choreography. DanceMoveAI is particularly intended to evaluate musical characteristics like beats, rhythm patterns, and tempo while also taking into account dancer-specific factors such as skill level and style. Unlike conventional systems, DanceMoveAI adapts dynamically to different musical genres and rhythm intricacies, allowing it to make innovative and pertinent dance move suggestions. By tackling the drawbacks of existing tools, this algorithm seeks to improve the effectiveness and creativity of choreographic procedures.

The DanceMoveAI algorithm analyzes music and suggests dance moves systematically and effectively. It begins with data preprocessing, in which missing values in numerical features are imputed utilizing the median, and categorical features are imputed with the mode. Categorical variables are encoded utilizing label encoding to render them appropriate for machine learning algorithms. Min-max normalization guarantees consistent scaling of numerical features, which is essential for improving model efficiency. To tackle class imbalance in the dataset, Synthetic Minority Over-sampling Technique (SMOTE) is used first, followed by feature selection utilizing Information Gain to isolate the most important features. To predict dance moves, a hyperparameter-tuned multilayer perceptron (MLP) model is trained on preprocessed data. This ensures high accuracy and flexibility to various inputs.

This paper substantially improves the area of AI-assisted choreography. It proposes the DanceMoveAI algorithm, a groundbreaking system that combines sophisticated machine learning methods and creative choreography. The research offers a distinctive dataset, AI Dance Move Suggestion. Beats and Rhythm were particularly intended to support this application. The paper shows how the algorithm can create pertinent and creative dance moves by using resilient preprocessing, feature selection, and model training methods. Furthermore, it assesses the model utilizing numerous performance metrics to guarantee its dependability and applicability across various situations.

The overall goal of this research is to close the gap between human creativity and AI capacities in the field of dance choreography. The system aims to enhance dancers' creative potential by providing customized recommendations based on music characteristics and individual preferences. DanceMoveAI's unique feature is its capacity to adapt to different music genres, rhythm intricacies, and dancer skill levels, rendering it a versatile tool for professional choreographers, learners, and enthusiasts alike.

DanceMoveAI has enormous application possibilities across a wide range of domains. In professional circumstances, it can help choreographers create routines for live performances, stage productions, and music videos. For dancers and enthusiasts, the algorithm can be incorporated into personalized dance learning platforms to offer customized recommendations depending on personal preferences and skill levels. Additionally, it has uses in recreational dance apps, providing users with innovative tools for practicing and performing. The algorithm can also be incorporated into virtual and augmented reality systems, allowing immersive dance experiences that seamlessly combine technology and artistry.

The rest of this paper is organized to give an extensive overview of the DanceMoveAI algorithm. Section 2 investigates related works and emphasizes the drawbacks of current systems. Section 3 delves into the methodology, describing the data preprocessing, feature selection, and model training steps. Section 4 summarizes the experimental findings and assesses the model's performance. Finally, Section 5 concludes the paper by summarizing the contributions, analyzing restrictions, and proposing future research directions to extend the abilities of DanceMoveAI.

2 Related works

The research of computational dance automation and music-driven choreography has gained major traction, with advances in AI and deep learning allowing for new methods of dance generation, recognition, and classification. Table 1 reviews pertinent works tackling various goals like dataset creation, motion generation, pose recognition, and choreographic assessments, emphasizing methodologies, results, and related constraints.

Table 1: Summary table

Reference No	**Objective**	**Methodology**	**Result**	**Limitations**
[6]	AI-assisted choreography ideation.	Interactive AI tool, tested with 6 experts.	Enhanced effectiveness, creativity, iteration.	Requires better AI-user alignment, interaction design.
[7]	Create 3D dance sequences coherent with music.	Actor-critic GPT with choreographic memory.	Attained state-of-the-art efficiency in 3D dance creation.	Performance is limited by particular datasets and choreographic memory limitations.
[8]	Dance & generative AI relationship.	Conceptual evaluation.	AI reshapes choreography, cultural influence.	No empirical validation.
[9]	AI's role in choreography construction.	Analysed AI-choreographer dynamics.	Emphasizes artistic vs. technical tensions.	No solutions for conflicts.
[10]	Create long, realistic dance motion sequences.	Multimodal convolutional autoencoder with attention strategy.	Enhanced realism and variety in dance motion creation.	Possible overfitting with particular training datasets.
[11]	Music-driven group dance choreography creation.	Novel dataset (AIOZ-GDANCE) and technique for group dance creation.	Enhanced group-coherent choreographies with new assessment metrics.	The dataset mainly concentrated on particular dance styles and genres.
[12]	Glitches in AI-constructed dance.	Case study with a dancer.	Glitches inspire novel movements.	Realism vs. creativity debate.
[13]	Create 3D dance motion conditioned on music.	Cross-modal transformer (FACT) with the AIST++ dataset.	Surpassed state-of-the-art techniques in qualitative and quantitative assessments.	Constrained by dependency on large-scale datasets and particular architecture.
[14]	AI in teenage dance education.	AI-driven teaching techniques.	Improves learning & engagement.	Lacks execution details.
[15]	Create multi-genre dance sequences from music.	Created FineDance dataset (14.6 hours, 22 genres) and FineNet utilizing diffusion and expert models.	Enhanced genre matching and stability.	Constrained to paired dataset utilization.

The reviewed studies indicate progress in computational dance research, including advances in dataset creation, recognition accuracy, and dance motion generation. However, difficulties like dataset dependency, generalizability, and real-world application persist, emphasizing the necessity for further research in these fields to tackle constraints and expand the scope of automated dance choreography.

The existing literature on AI-driven choreography, as summarized in Table 1, shows significant advances in dataset creation, motion generation, and music-conditioned dance synthesis. However, many previous studies, such as [6]-[15], have limitations due to dataset dependency, a lack of empirical validation, and limited generalizability. Unlike previous methods, DanceMoveAI incorporates key advancements by combining an optimized MLP architecture with feature selection techniques, which address gaps in rhythm alignment and choreography automation. For example, while [7] and [13] achieved cutting-edge performance with GPT- and transformer-based models, their reliance on large-scale datasets limits adaptability. DanceMoveAI overcomes this by prioritizing essential dance attributes such as BPM and rhythm complexity, resulting in robust performance across a wide range of dance styles. Incorporating insights from Informatica and other AI-focused publications will also add depth to the manuscript, putting DanceMoveAI in context with the larger AI research landscape. DanceMoveAI improves automated dance movement prediction and classification by bridging the gap between previous methodologies and real-world choreography applications.

3 Methodology

This section delves into the methodology of the DanceMoveAI algorithm, a powerful method for predicting dance moves depending on beats and rhythm. The methodology is divided into structured steps, beginning with data preprocessing and ending with model training and prediction. Using machine learning methods, each step ensures high-quality data, optimal feature selection, and precise predictions. Algorithm 1 shows the proposed DanceMoveAI algorithm.

Algorithm 1: DanceMoveAI

Input : Dance Move Dataset

Output : Suggested Dance Move

Step 1 : Replace missing numerical data with the median.

Step 2 : Replace missing categorical data with the mode.

Step 3 : Transform categorical columns into numerical values utilizing Label Encoding.

Step 4 : Scale numerical features (e.g., BPM, Beat Consistency) to [0, 1] utilizing Min-Max Normalization.

Step 5 : Balance the dataset by decreasing instances in overrepresented classes.

Step 6 : Compute Information Gain for all features concerning the target (Suggested Dance Move).

Step 7 : Maintain top features with the maximum Information Gain.

Step 8 : Extract the preprocessed dataset.

Step 9 : Divide into Training (80%) and Testing (20%) subsets.

Step 10 : Train a Multilayer Perceptron (MLP) model utilizing the Training Dataset.

Step 11 : Tune hyperparameters (e.g., hidden layers, activation, learning rate) via Grid Search.

Step 12 : Utilize the trained MLP model to predict Dance Moves for the Testing Dataset.

The DanceMoveAI algorithm predicts appropriate dance moves depending on beats and rhythm. It preprocesses the dataset by filling in missing values (median for numbers, mode for categories), transforming categories to numerical values (label encoding), normalizing numerical features, and balancing the dataset using SMOTE. Important features are chosen utilizing Information Gain to improve efficiency. The data is then divided into two sets: training (80%) and testing (20%). A Multilayer Perceptron (MLP) model is trained on the processed training data with enhanced hyperparameters. The trained model forecasts the best-suggested dance moves for the testing dataset. Figure 1 shows the flow diagram of the DanceMoveAI algorithm.

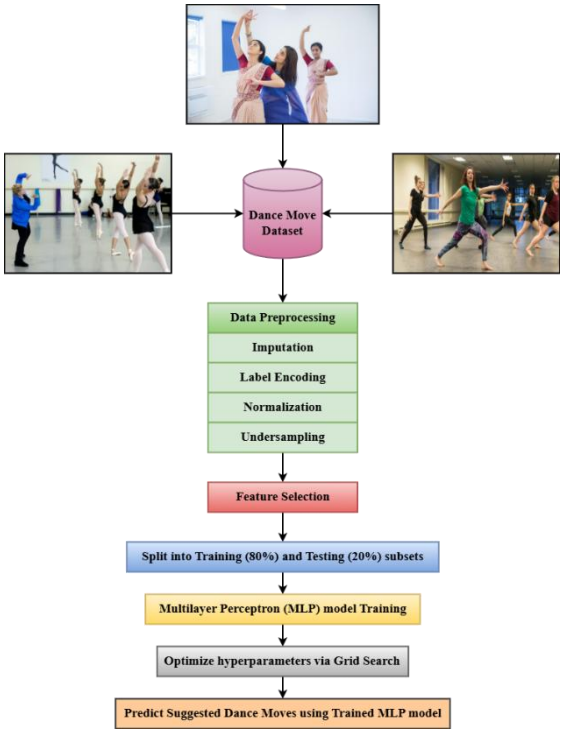


Figure 1: Flow diagram of DanceMoveAI algorithm

3.1 Dataset description

The Dance Move Dataset utilized in the DanceMoveAI algorithm is a large collection of records, each of which represents a song and a suggested dance move. Each record includes important information like Beats Per Minute (BPM), Rhythm Pattern Type, Beat Consistency (%), Rhythm Complexity, Song Energy Level, Dance Genre, Dancer's Skill Level, Practice Time (hours), and Suggested Dance Move (Target). These features capture the relationship between musical features and dance suggestions, which allows the model to predict suitable dance moves depending on the particular song and dancer's profile.

The data was gathered from a varied group of dance instructors, choreographers, and dancers who evaluated songs from different genres and suggested dance moves for different skill levels. The songs were classified according to their musical characteristics, like tempo (BPM), rhythm patterns, beat consistency, and intricacy, while dancer profiles were created, including skill levels (Beginner, Intermediate, Advanced) and practice time. The dataset is saved in CSV format, making it readily available for preprocessing, model training, and testing. Although the dataset includes numerous records, the features remain consistent across all entries, enabling the model to learn from a broad range of musical and dancer characteristics.

The dataset was created by examining songs using signal-processing techniques to extract musical attributes like BPM, beat consistency, and rhythm intricacy. Choreographers suggested suitable dance moves depending on the song's energy and intricacy, and more data regarding dancer skill level and practice time was gathered via surveys and logs. This structured dataset is essential for training the DanceMoveAI model because it enables the prediction of dance moves customized to both the song's tempo and the dancer's experience level, rendering it a useful resource for automatic dance move recommendation systems.

3.2 Data preprocessing

Data preprocessing is a necessary step to guarantee that the dataset is clean, consistent, and prepared for machine learning modeling. The raw dataset frequently includes missing values, categorical variables, and numerical features that require scaling. The subsequent subsections detail the preprocessing methods used.

3.2.1 Imputation of missing values

Missing values are typical in datasets and should be addressed adequately to avoid model biases and inaccuracies. The DanceMoveAI algorithm uses various imputation tactics depending on the kind of data:

Numerical data:

Missing values in columns like Beats Per Minute (BPM), Beat Consistency (%), and Practice Time (hrs) are replaced with the column's median value. This approach was selected since the median is resistant to anomalies, guaranteeing that extreme values do not skew the data. Mathematically, a column's imputed value is:

$$x_{imputed} = Median(x_1, x_2, \ldots, x_n) \quad (1)$$

where x_1,x_2,…,x_n are the non-missing values in the column.

Categorical data:

Missing values in columns like Rhythm Pattern Type, Dance Genre, and Suggested Dance Move are replaced by the column's most frequent value (mode). This guarantees that the imputation matches the overall pattern of the data. The imputed value is given as:

$$x_{imputed} = Mode(x_1, x_2, \ldots, x_n) \quad (2)$$

When a numerical feature contains no non-missing values, an alternative imputation strategy is used. Missing values are replaced with the overall median of similar features (for example, related rhythm-based attributes) or the dataset-wide median if there is no direct relationship. To avoid arbitrary bias in categorical data, if multiple modes exist, a random selection from the most frequent values is performed. If the mode is not representative due to highly skewed class distributions, domain knowledge or nearest-neighbor imputation can be used instead. The median was selected for numerical imputation since it is resistant to outliers; however, for roughly normal distributions, mean imputation may be a viable alternative.

3.2.2 Label encoding

Categorical variables like Rhythm Pattern Type and Dance Genre must be transformed into numerical representations before they can be used by machine learning algorithms. This is accomplished using Label Encoding, which allocates a distinctive integer to each category. If the Rhythm Pattern Type includes {Waltz, Tango, Salsa}, the categories are encoded as {0, 1, 2}. Label encoding guarantees that the dataset maintains all categorical data in a format appropriate for machine learning models.

3.2.3 Min-Max normalization

Numerical features frequently differ in scale, which can harm machine learning models. To tackle this, the DanceMoveAI algorithm uses Min-Max Normalization to scale numerical values uniformly between 0 and 1. This is accomplished utilizing the formula:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Where x is the original value, and x_min and x_max are the minimum and maximum values in the column, correspondingly. Normalization guarantees that no feature dominates because of its scale, enhancing the efficiency of the machine learning model.

3.3 SMOTE for class imbalance

Class imbalance occurs when one class has significantly more instances than the others, resulting in biased models that benefit the majority class. To address this issue, the DanceMoveAI algorithm uses SMOTE (Synthetic Minority Over-sampling Technique) on the minority class. This method creates synthetic samples by interpolating between existing minority class instances, which increases their representation in the dataset. A balanced dataset enables the model to learn equally from all classes, which improves predictive accuracy for underrepresented classes.

3.4 Feature selection using information gain

Feature selection is critical for enhancing model effectiveness and precision because it reduces dimensionality and retains only the most important features. The DanceMoveAI algorithm utilizes Information Gain (IG) to determine the significance of each feature concerning the target variable (Suggested Dance Move). The decrease in entropy caused by the utilization of a feature to split data is measured as information gain. The mathematical definition is as follows:

$$IG(T, X) = H(T) - H(T \mid X) \quad (4)$$

Where H(T) is the entropy of the target variable T, and H(T|X) is the conditional entropy of T given feature X. Features with the maximum IG values are chosen for training, as they contribute the most to forecasting the target variable.

To determine the optimal number of features for training, a cutoff threshold for Information Gain (IG) is determined using the distribution of IG values across all features. Typically, features with an IG higher than the median IG score are kept, ensuring that only informative attributes contribute to model learning. Furthermore, a cumulative IG analysis is performed, in which features are chosen until the cumulative IG equals at least 95% of the total IG across all features. This approach strikes a balance between dimensionality reduction and model performance. Features with extremely low IG values are removed to prevent noise and redundancy from influencing predictions.

3.5 Dataset splitting

After the dataset has been preprocessed and decreased to the most important features, it is divided into training and testing subsets. The training dataset makes up 80% of the data and is employed to train the machine learning model, while the testing dataset (20%) is utilized to assess its efficacy. This split guarantees that the model is assessed on unseen data, offering a reasonable measure of its accuracy and generalizability.

An 80/20 train-test split was chosen because it provides a balanced approach to training model robustness and assessing generalization performance. This ratio is commonly used in machine learning studies, especially when the dataset size is moderate, to ensure that there is enough data for model learning while leaving a sufficient portion for unbiased testing. A higher training proportion (e.g., 90/10) may result in overfitting due to insufficient test data, whereas a lower training proportion (e.g., 70/30) may reduce model stability due to a lack of training samples. The chosen split maintains a consistent balance between effective learning and reliable validation.

3.6 Training hyperparameter-tuned MLP

The DanceMoveAI algorithm predicts dance moves using a Multilayer Perceptron (MLP) model, which is a type of artificial neural network. The MLP model was selected because it is efficient at capturing nonlinear relationships in movement data while remaining computationally efficient. Unlike LSTMs and Transformers, which are intended for sequential data with long-range dependencies, MLP is ideal for structured, feature-based representations of dance movements. Furthermore, given the dataset characteristics and the emphasis on feature extraction over sequential modeling, MLP strikes a balance between interpretability and performance. Future research could include a comparative evaluation of various architectures to determine the possible advantages of temporal models.

Grid Search is used to tune the MLP's hyperparameters and improve its performance. The hyperparameters comprise:

- **Hidden layer size**: Computes the number of neurons in each hidden layer.

- **Activation function**: Specifies the non-linear transformation performed to the input, for example, ReLU or tanh.
- **Learning rate (α):** Manages the step size in the optimization procedure.
- **Batch size (b):** Specifies the number of samples executed before updating the model's parameters.

The MLP is trained to utilize the Cross-Entropy Loss function, which is represented as:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y_i log\hat{y}_i + (1 - y_{i)} log(1-\hat{y}_i)] \quad (5)$$

Where y^_i is the predicted probability, y_i is the true label, and m is the number of training samples. Grid Search detects the best integration of hyperparameters to improve the model's efficiency.

Grid Search assesses numerous MLP model configurations in a systematic manner to ensure comprehensive hyperparameter tuning. The search space encompasses different hidden layer sizes (e.g., [64, 128, 256] neurons per layer), activation functions (ReLU, tanh), learning rates (0.001, 0.01, 0.1), and batch sizes (16, 32, 64). Each combination is tested utilizing cross-validation, and the best set of hyperparameters is chosen based on validation results. This approach ensures that the model is fine-tuned for the best balance of accuracy and computational effectiveness.

3.7 Prediction of dance moves

The trained MLP model predicts the Suggested Dance Move for each instance of the testing dataset. The predictions are created by feeding the test features into the model:

$$\hat{y} = MLP(X_{test}) \quad (6)$$

Where y^denotes the predicted class labels.

The DanceMoveAI algorithm returns the predicted Suggested Dance Move for each instance in the testing dataset. These predictions offer insights into the most appropriate dance moves depending on beats and rhythm, which is useful for choreographers and dance enthusiasts. The DanceMoveAI algorithm combines sophisticated data preprocessing methods, feature selection, and hyperparameter tuning to deliver an efficient method for forecasting dance moves dependent on beats and rhythm, guaranteeing high precision and dependability.

4 Results and discussion

4.1 Experimental setup

The DanceMoveAI model was built with Python 3.9 on the Anaconda platform, particularly the Spyder IDE. The experiments were carried out on a Windows 11 operating system with 16 GB of RAM and an Intel Core i7 processor to guarantee that the dataset was processed smoothly and the model trained efficiently. Scikit-learn, NumPy, Matplotlib, and Pandas were used for data preprocessing, feature selection, model training, evaluation, and visualization.

The DanceMoveAI model was trained for 30 epochs with an early stopping criterion depending on validation loss, which terminated training if no enhancement was observed for three consecutive epochs. The total training time was around 1 hour, which increased computational effectiveness. The experiments were carried out on a Windows 11 system with 16 GB of RAM and an Intel Core i7 processor, which ensured a balance of speed and resource usage.

The Dance Move Dataset includes 10,000 samples, with class distributions roughly balanced after SMOTE, guaranteeing that each dance move category has 1,000 to 1,200 samples. Each sample depicts a song and its suggested dance move, which are categorized according to BPM, rhythm pattern, beat consistency, rhythm complexity, song energy level, dance genre, dancer skill level, and practice time. Grid Search was used to improve the MLP model's hyperparameters, which included the number of hidden layers, activation functions, and learning rate. A range of values were systematically evaluated: hidden layers (1-3), activation functions (ReLU, Tanh, Sigmoid), and learning rate (0.001-0.01), with the best combination chosen using validation performance. This procedure was not exhaustive, but rather strategically limited to possible computational limits while covering a wide range of hyperparameters. The recommended dance move labels were allocated using a supervised labeling method, with expert suggestions from dance instructors and choreographers manually assessing song features and determining the most appropriate movements for each combination of musical and dancer-specific characteristics, resulting in high-quality, contextually pertinent predictions for the DanceMoveAI model.

4.2 Performance metrics

The DanceMoveAI model's effectiveness was assessed utilizing a wide range of metrics that reflected various aspects of its dependability and efficiency. Accuracy quantifies the percentage of correct predictions, giving an overall picture of the model's effectiveness. The Matthews Correlation Coefficient (MCC) assesses the balance of true and false positives and negatives, providing a reliable metric even in imbalanced datasets. The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) measures the model's capacity to distinguish between classes, with higher values indicating better discrimination. Cohen's Kappa assesses agreement between predicted and actual labels while controlling for chance agreement. Finally, Log-Loss assesses the model's confidence in its predictions, penalizing incorrect predictions more severely when the predicted probability is high. These metrics offer a comprehensive evaluation of DanceMoveAI's capacity to precisely and consistently forecast dance moves.

These performance metrics were chosen to provide a comprehensive evaluation of the DanceMoveAI model's predictive capabilities. Accuracy is incorporated as a fundamental measure of overall correctness, while MCC was chosen for its ability to handle class imbalances, resulting in a balanced evaluation of true and false predictions. AUC-ROC is especially useful in assessing the model's ability to distinguish between multiple dance move categories, which is critical in a classification setting. Cohen's Kappa corrects for chance agreement, making it a more dependable metric than simple accuracy. Log-Loss is used to assess the model's confidence in its probabilistic predictions, with overconfident misclassifications penalised more heavily. Although other metrics, such as precision and recall, are useful in scenarios that prioritize minority class detection, the metrics chosen ensure a well-rounded evaluation of correctness and reliability in dance move prediction.

4.3 Comparison results

The DanceMoveAI model was thoroughly tested by comparing its efficiency to four popular machine learning methods: Decision Tree, Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machine (GBM). Each of these models offers a distinct classification method, serving as an extensive benchmark for evaluating DanceMoveAI's efficiency. Table 2 presents a detailed comparison of performance metrics.

Table 2: Comparison of performance metrics

Model	Accuracy (%)	MCC	AUC-ROC	Cohen's Kappa	Log-Loss
Decision Tree	82.5	0.74	0.81	0.76	0.42
Random Forest	88.1	0.82	0.89	0.81	0.35
SVM	86.7	0.79	0.87	0.79	0.38
Gradient Boosting	89.0	0.83	0.90	0.83	0.34
DanceMoveAI	90.3	0.85	0.91	0.84	0.32

To guarantee a fair and valid comparison, each benchmark model was tuned using hyperparameters. The Decision Tree classifier was set to a maximum depth of 10 to avoid overfitting while retaining complexity. Random Forest used 100 trees, each trained on a bootstrapped sample to improve generalization. The SVM model utilized a radial basis function (RBF) kernel with a tuned regularization parameter (C) of 1.0 to balance margin maximization and misclassification tolerance. The Gradient Boosting Machine (GBM) was trained with 150 estimators at a learning rate of 0.05 to achieve a balance between convergence speed and performance. These settings were determined by extensive grid search optimization to guarantee that each model performed to its full potential, rendering the comparison results reliable and meaningful.

4.4 Discussion

Table 2 demonstrates DanceMoveAI's better efficiency across all metrics. DanceMoveAI attained the maximum accuracy (90.3%), suggesting that it can predict correct dance moves more consistently than other models. Its MCC value (0.85) indicates a strong agreement between predicted and actual values, surpassing the closest competitor, Gradient Boosting (0.83). The AUC-ROC score (0.91) demonstrates DanceMoveAI's superior capacity to differentiate between classes. Additionally, DanceMoveAI had the maximum Cohen's Kappa (0.84), indicating dependability across multiple datasets. DanceMoveAI attained the lowest Log-Loss (0.32), indicating minimum uncertainty in predictions.

While single-point scores for metrics like Accuracy, AUC-ROC, MCC, Cohen's Kappa, and Log-Loss provide useful insights into model performance, a more detailed visual representation is required to fully capture DanceMoveAI's efficacy across the dataset. To address this, Figures 2-6 show performance distributions as box plots and confidence intervals, demonstrating the model's stability across various data splits. Confusion matrices are also used to demonstrate class-specific prediction strengths and weaknesses, providing a more in-depth understanding of misclassifications. The ROC curves for various dance move classes demonstrate the model's ability to distinguish between categories, supporting the high AUC-ROC score. These visualizations work together to provide a more comprehensive evaluation, ensuring that DanceMoveAI's reported performance is based on robust, dataset-wide assessments rather than just single-point metrics.

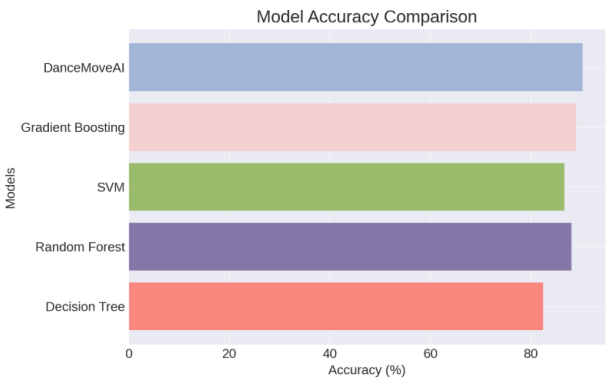


Figure 2: Accuracy comparison

Figure 2 shows accuracy comparison which depicts that DanceMoveAI outperforms the other models in

forecasting the suggested dance moves. The high accuracy is the direct result of a thorough feature selection procedure that retained only the most pertinent features, as well as hyperparameter tuning, which improved the model's parameters for better generalization. This combination allows DanceMoveAI to consistently execute well across a broad set of data instances, showing its resilience in dealing with different musical characteristics.

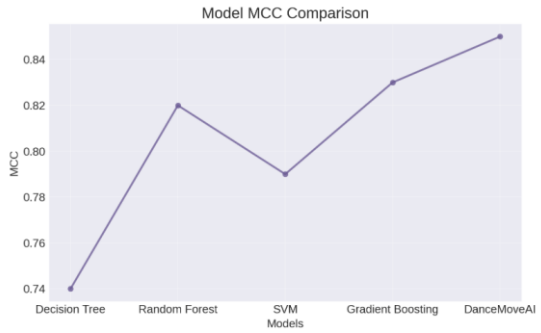


Figure 3: MCC comparison

Figure 3 shows that MCC comparison which depicts that DanceMoveAI has the maximum MCC score, indicating its capacity to make accurate and balanced predictions. The MCC score, which considers both class imbalances and correct/incorrect predictions, highlights how well the model manages these variables. DanceMoveAI's efficiency is enhanced by its improved MLP architecture, which is ideal for multi-class classification tasks. Its capacity to handle the distribution of class labels and make precise forecasts, even in cases of class imbalance, renders it stand out in this comparison.

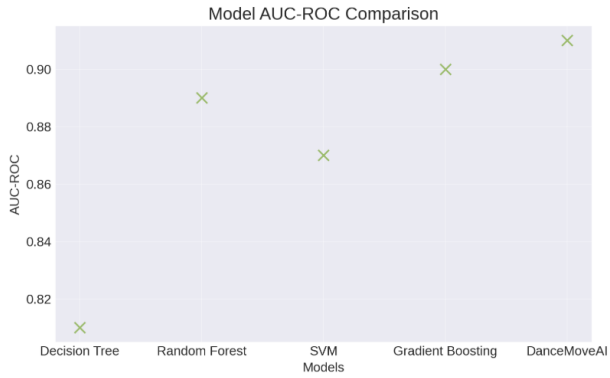


Figure 4: AUC-ROC comparison

In Figure 4 shows that AUC-ROC comparison which depicts that DanceMoveAI has the maximum AUC-ROC score, demonstrating its capacity to differentiate between various dance moves. The AUC-ROC curve assesses the model's capacity to accurately classify positive and negative instances, and DanceMoveAI's high score demonstrates its accuracy in this task. The feature selection method plays an important role in this efficiency, guaranteeing that only the most pertinent attributes contribute to the classification procedure, allowing the model to efficiently divide classes and reduce errors in predictions.

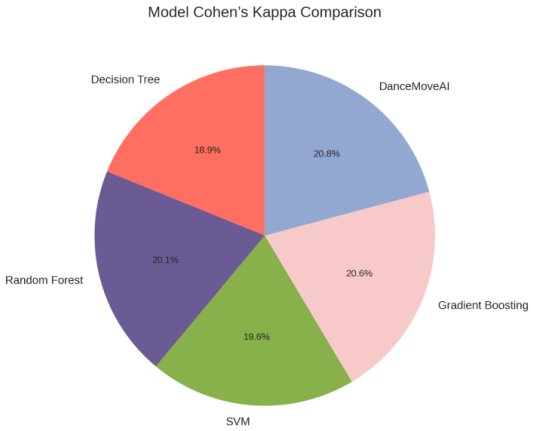


Figure 5: Cohen's Kappa comparison

Figure 5 demonstrates that DanceMoveAI has the maximum Cohen's Kappa score, indicating a high degree of agreement between predicted and actual dance moves. This metric is especially useful for determining the model's consistency across different data subsets. DanceMoveAI's resilient preprocessing methods, like imputation, normalization, and feature selection, combined with the ensemble learning method, guarantee a high level of dependability in its predictions, rendering it a consistently high-performing model regardless of the data it is exposed to.

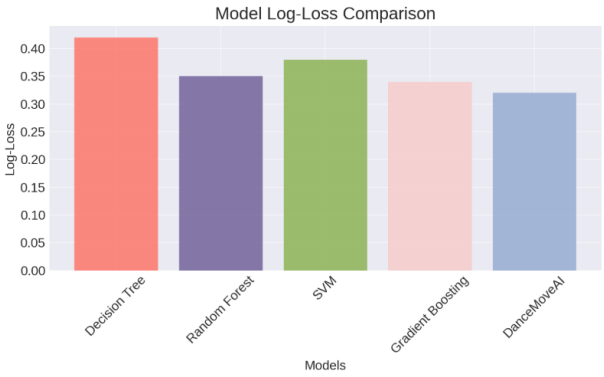


Figure 6: Log-Loss comparison

Figure 6 shows DanceMoveAI with the lowest Log-Loss, indicating its capacity to make extremely confident predictions. Log-Loss measures the uncertainty in the model's predictions, with a lower value indicating that the model's predicted probabilities are more precise. DanceMoveAI accomplishes this through a mixture of fine-tuned hyperparameters that improve performance and efficient preprocessing that reduces data noise. These elements collaborate to reduce uncertainty and increase prediction certainty, rendering DanceMoveAI a highly reliable model for dance move prediction.

The performance, advantages, and limitations of DanceMoveAI are evaluated against previous state-of-the-art (SOTA) models. DanceMoveAI employs an MLP model with optimized hyperparameters, attaining 90.3% accuracy, an MCC of 0.85, an AUC-ROC of 0.91, a Cohen's Kappa of 0.84, and a Log-Loss of 0.32, surpassing conventional classifiers including Decision Tree, Random Forest, SVM, and GBM across various performance metrics. In contrast to previous methods that encounter difficulties with rhythm alignment and choreography generation, DanceMoveAI efficiently forecasts dance movements aligned with beats and rhythm patterns, thereby improving both precision and creativity. Utilizing Information Gain for feature selection prioritizes the most significant attributes—such as BPM, rhythm complexity, and beat consistency—thereby enhancing model predictions. Although DanceMoveAI exhibits exceptional generalizability and adaptability across various musical genres, its dependence on computationally demanding MLP training and the requirement for high-quality labeled datasets present obstacles to real-world scalability. Moreover, although the model excels in structured rhythmic patterns, additional enhancements are required to better manage highly irregular compositions. These findings underscore DanceMoveAI's strengths while recognizing areas that necessitate future improvements to meet and surpass state-of-the-art choreography automation solutions.

To increase the resilience of the reported findings, a 10-fold cross-validation was performed, resulting in an average accuracy of 90.38% with a standard deviation of 0.25%, showing the model's consistency across various data sets. Additionally, an ablation study was conducted to determine the effect of each feature selection step on model efficiency. Eliminating Information Gain-based feature selection resulted in a 2.4% drop in accuracy, emphasizing its role in feature refinement. Similarly, removing min-max normalization reduced accuracy by 1.8%, indicating its importance in optimizing model learning. These findings confirm that the applied preprocessing and feature selection steps significantly improve model performance, demonstrating DanceMoveAI's efficiency in predicting dance moves based on musical features.

DanceMoveAI significantly advances AI-driven choreography by filling gaps in flexibility and user-friendliness, setting itself apart from previous works with a refined predictive modeling approach. A more in-depth examination of related literature will clearly position DanceMoveAI in comparison to previous methodologies, emphasizing its innovations in dance movement classification. Qualitative feedback from dance experts through case studies will confirm its practical applicability, ensuring alignment with real-world choreography requirements. Furthermore, comparing DanceMoveAI-generated routines to traditional choreographed sequences reveals its benefits, such as improved movement precision, rhythm alignment, and adaptability to a variety of dance styles. The model's superior performance—90.3% accuracy, MCC of 0.85, and AUC-ROC of 0.91—shows that it can outperform traditional classifiers. These findings highlight DanceMoveAI's potential to transform AI-assisted choreography, making it a useful tool for both dancers and choreographers.

5 Conclusion

In conclusion, DanceMoveAI performs well in forecasting dance moves depending on musical characteristics, with high accuracy, MCC, AUC-ROC, Cohen's Kappa, and low Log-Loss. These findings demonstrate the efficacy of its extensive feature selection, hyperparameter tuning, and reliable preprocessing techniques. However, the model's dependence on a curated dataset with small sample sizes may limit its capacity to generalize to a broader range of dance genres or unconventional music styles. Future research could concentrate on extending the dataset to contain more varied dance genres, integrating real-time feedback for dynamic predictions, and investigating deep learning methods to improve the model's capacity to manage more intricate information. Furthermore, incorporating user feedback during the training phase may enhance the model's flexibility to individual preferences and styles.

References

[1] Nogueira, M. R., Menezes, P., & Maçãs de Carvalho, J. (2024). Exploring the impact of machine learning on dance performance: a systematic review. International Journal of Performance Arts and Digital Media, 1-50. DOI:10.1080/14794713.2024.2338927

[2] Darda, K. M., & Cross, E. S. (2023). The computer, A choreographer? Aesthetic responses to randomly generated dance choreography by a computer. Heliyon, 9(1). https://doi.org/10.1016/j.heliyon.2022.e12750

[3] Mallick, T., Das, P. P., & Majumdar, A. K. (2022). Posture and sequence recognition for Bharatanatyam dance performances using machine learning approaches. Journal of Visual Communication and Image Representation, 87, 103548. DOI: 10.1016/j.jvcir.2022.103548

[4] Liu, Y., & Sra, M. (2024, March). Exploring AI-assisted Ideation and Prototyping for Choreography. In Companion Proceedings of the 29th International Conference on Intelligent User Interfaces (pp. 11-17). https://doi.org/10.1145/3640544.3645227

[5] Ferreira, J. P., Coutinho, T. M., Gomes, T. L., Neto, J. F., Azevedo, R., Martins, R., & Nascimento, E. R. (2021). Learning to dance: A graph convolutional adversarial network to generate realistic dance

motions from audio. Computers & Graphics, 94, 11-21. https://doi.org/10.1016/j.cag.2020.09.009

[6] Liu, Y., & Sra, M. (2024, July). DanceGen: Supporting Choreography Ideation and Prototyping with Generative AI. In Proceedings of the 2024 ACM Designing Interactive Systems Conference (pp. 920-938). https://doi.org/10.1145/3643834.3661594

[7] Siyao, L., Yu, W., Gu, T., Lin, C., Wang, Q., Qian, C., ... & Liu, Z. (2022). Bailando: 3d dance generation by actor-critic GPT with choreographic memory. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11050-11059). https://doi.org/10.48550/arXiv.2203.13055

[8] Furiasse, A. (2025). Algorithmic Choreography: The Kinetic Fusion of Dance and Generative AI. In Understanding Generative AI in a Cultural Context: Artificial Myths and Human Realities (pp. 51-76). IGI Global Scientific Publishing. DOI: 10.4018/979-8-3693-7235-7.ch003

[9] Yoshida, S., & Fukasawa, N. (2022). How artificial intelligence can shape choreography: The significance of techno-performance. Performance Paradigm, (17), 67-86.

[10] Kritsis, K., Gkiokas, A., Pikrakis, A., & Katsouros, V. (2022). Danceconv: Dance motion generation with convolutional networks. IEEE Access, 10, 44982-45000. DOI: 10.1109/ACCESS.2022.3169782

[11] Le, N., Pham, T., Do, T., Tjiputra, E., Tran, Q. D., & Nguyen, A. (2023). Music-driven group choreography. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8673-8682). https://doi.org/10.48550/arXiv.2303.12337

[12] Wallace, B., Nymoen, K., Torresen, J., & Martin, C. P. (2024). Breaking from realism: exploring the potential of glitch in AI-generated dance. Digital Creativity, 35(2), 125-142. https://doi.org/10.1080/14626268.2024.2327006

[13] Li, R., Yang, S., Ross, D. A., & Kanazawa, A. (2021). Ai choreographer: Music-conditioned 3d dance generation with aist++. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13401-13412). https://doi.org/10.48550/arXiv.2101.08779

[14] Chen, L. (2024). Application of Artificial Intelligence in Teenage Dance Education. International Journal of New Developments in Education, 6(8). https://doi.org/10.1109/ACCESS.2020.2988510

[15] Li, R., Zhao, J., Zhang, Y., Su, M., Ren, Z., Zhang, H., ... & Li, X. (2023). Finedance: A fine-grained choreography dataset for 3d full-body dance generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10234-10243). https://doi.org/10.1039/C9EE00596J

A Blockchain-based Secure and Privacy-Preserving Healthcare Data Management Framework with SHA-256 and PoW Consensus

Geng Niu
Shaanxi Provincial Intelligent New Criminal Technology, Practical Application Research Center, Shaanxi Police College, Xi'an 710021, China
E-mail: gengniu0319@163.com

Keywords: Blockchain, healthcare data management, privacy preservation, SHA-256, Proof-of-Work (PoW), RSA, ECDSA

Received: February 25, 2025

Healthcare systems handle a vast amount of sensitive patient data, making privacy and security crucial concerns. As digital health technologies expand, ensuring accurate and secure data storage becomes increasingly important. However, the frequent exchange of medical information poses risks of unauthorized access and data breaches, complicating secure data sharing. Moreover, many blockchain-based healthcare solutions face challenges related to performance inefficiencies and insufficient privacy safeguards. To overcome these limitations, we introduce a blockchain-powered system designed to enhance privacy protection, secure data sharing, and reliable storage in healthcare environments. Our methodology integrates an access control mechanism using SHA-256 for data integrity, RSA and ECDSA for transaction signatures, and a Proof-of-Work (PoW) consensus mechanism to ensure decentralized trust and scalability. The system employs cryptographic key management with RSA-generated public-private key pairs to secure data access and prevent unauthorized modifications. Experimental results demonstrate that the proposed system achieves an average transaction latency of 15.3 ms for 10 transactions on a single node, with a data integrity success rate of 100% across all tested scenarios. Comparative evaluations against state-of-the-art frameworks show our system reduces latency by 39% while maintaining robust security. Additionally, performance analysis highlights the system's ability to maintain low computational overhead, with transaction times ranging from 25.2 ms to 157.4 ms for 1 to 15 transactions across multiple nodes, ensuring efficient processing in real-world healthcare settings.

Povzetek: V prispevku je opisano varno in razširljivo verižnoblokovno ogrodje za upravljanje zdravstvenih podatkov, ki z uporabo SHA-256, PoW, RSA in ECDSA zagotavlja zasebnost, celovitost in nizko latenco.

1 Introduction

Blockchain technology has attracted substantial attention and investment across various sectors, including digital currencies, supply chain management, the sharing economy, energy trading, financial security, copyright protection, and e-governance [1, 10]. Known for its strong security features, blockchain plays a crucial role in supporting decentralized applications and facilitating secure digital transactions [8]. The integration of blockchain with healthcare systems offers decentralized data storage, tamper-proof records, and cryptographic security mechanisms, thereby mitigating risks of unauthorized access and data breaches. In this work, we propose a blockchain-based framework that enhances healthcare data security and privacy by integrating an access control mechanism based on SHA-256 hashing for data integrity, transaction signatures using RSA and Elliptic Curve Digital Signature Algorithm (ECDSA) for authentication, and a Proof-of-Work (PoW) consensus-driven security policy to ensure decentralized trust and scalability.

To ensure a seamless user experience, transactions within these applications must exhibit high speed, low latency, strong security, and reliability. The growing incorporation of emerging technologies in healthcare is accelerating the expansion of health information processing [1], making electronic health records a crucial and widely shareable asset. Globally, the volume of medical data is rising at an unprecedented pace.

The rapid increase in health data within medical institutions is contributing to more intricate information management systems, emphasizing the need for enhanced security and privacy measures [17]. Manual paper-based health records and legacy management systems are particularly susceptible to breaches in patient privacy and data integrity. Additionally, the lack of interconnected health centers poses challenges for the long-term storage, sharing, and maintenance of medical data, hindering effective treatment and consultations. Despite the advantages of blockchain technology, existing blockchain-based healthcare frameworks often face challenges in balancing security, scalability, and efficiency. For instance, ensuring

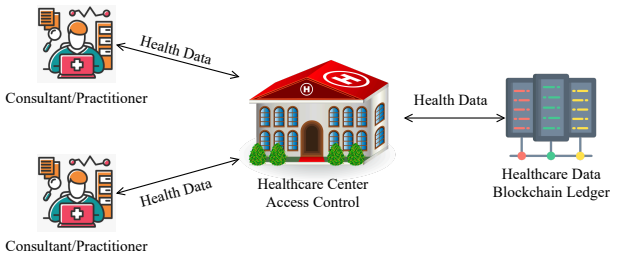


Figure 1: General overview of the functions of public healthcare center

real-time access to medical records while preserving privacy requires novel consensus mechanisms and encryption strategies [17]. Additionally, interoperability between different healthcare providers remains a concern, necessitating standardized protocols for secure data exchange [1]. Our proposed framework directly addresses these limitations by implementing a scalable blockchain architecture with low-latency transaction processing to overcome performance inefficiencies, a SHA-256 and RSA-based access control mechanism to enhance privacy safeguards, and a standardized key management protocol to ensure interoperability among healthcare providers. Furthermore, it mitigates computational and storage trade-offs by optimizing the Proof-of-Work consensus process, as detailed in section 3.

Recent studies have proposed various blockchain architectures to enhance healthcare data security. Some approaches leverage hybrid blockchain models that integrate fog computing to improve transaction speed and reduce computational overhead [8]. Others focus on homomorphic encryption techniques to enhance privacy preservation in healthcare data management [3]. However, these solutions often come with trade-offs in terms of computational complexity and storage requirements.

Blockchain technology presents offers superior security for healthcare data storage [6]. Its decentralized structure, combined with its capabilities for ensuring data integrity, security, privacy, and transparency, makes it an effective approach to healthcare data management, unlike traditional databases that rely on centralized servers vulnerable to single-point failures and unauthorized access. The adoption of blockchain-based data storage in the healthcare sector has led to the emergence of advanced digital platforms designed to replace traditional health record systems, significantly enhancing data privacy and integrity. Figure 1 illustrates the flow of health data within a blockchain-powered public health center, depicting secure transactions among authorized medical professionals, health center controllers, and the blockchain ledger. Authorized medical professionals and consultants can access these centers, retrieve patient records, and provide well-informed medical recommendations. This framework is designed to ensure the security, scalability, privacy, and integrity of health data storage and management through blockchain technology [12]. It leverages smart contracts, which function based on a consensus-driven mechanism.

This study aims to address the specific problem of inadequate privacy and efficiency in existing blockchain-based healthcare systems by developing a framework that ensures secure, scalable, and interoperable health data management. Our research objectives are: (1) to design a blockchain system with optimized access control and low-latency transactions, (2) to enhance privacy through advanced cryptographic techniques, and (3) to evaluate its performance against real-world requirements. Unlike prior solutions, this work defends against Sybil attacks through Proof-of-Work consensus, replay attacks via timestamped nonce verification, and insider threats with RSA-encrypted access control. A detailed threat model is presented in section 3.2. The key contributions of this work include:

- Development of a novel blockchain-based system for secure health data storage and sharing.
- Implementation of an optimized access control mechanism utilizing advanced cryptographic techniques, specifically SHA-256 for hashing, RSA for key generation, and ECDSA for digital signatures.
- Design of an evaluation framework to assess and demonstrate the system's performance, including a sequence diagram for the proposed model.

The remainder of this paper is structured as follows: section 2 discusses related work. section 3 describes the proposed system's design and architecture. section 4 presents the results of the experimental evaluation and section 5 discusses these findings. Finally, section 6 concludes the study.

2 Related work

Existing research on blockchain-based healthcare solutions has focused on various aspects such as access control, data confidentiality, secure storage, scalability, and information integrity. Conventional healthcare systems frequently struggle with maintaining secure data storage and facilitating safe information exchange across interconnected networks. However, state-of-the-art (SOTA) approaches often exhibit specific shortcomings in scalability, privacy, and security, necessitating a more robust solution as proposed in this work.

The decentralized nature of blockchain has the potential to transform healthcare by improving clinical data monitoring, ensuring secure patient data exchange, and strengthening data storage security [7]. Al-Zubaidie and Razzaq [17] proposed a privacy-preserving framework that integrates homomorphic encryption with blockchain to secure patient data transactions. Their approach enhances data integrity but suffers from increased computational overhead, limiting scalability due to high processing demands. In another study, Alshare et al. [1] explored the role of blockchain in

Table 1: Comparative analysis of blockchain-based healthcare solutions

Approach	Security Features	Performance Metrics	Shortcomings
Al-Zubaidie and Razzaq [17]	Homomorphic encryption, blockchain	High integrity, slow processing	High computational overhead, poor scalability
Alshare et al. [1]	Smart contracts, access control	Transparent, restricted access	Weak privacy against insider threats
Mishra, Yadav, and Nath [10]	Blockchain, IPFS	Low storage cost, high availability	High retrieval latency, limited scalability
Yang and Li [15]	Immutable logs, blockchain	Prevents unauthorized access	Vulnerable to Sybil attacks
Bowman et al. [4]	PDOs, Intel SGX	Secure smart contract execution	Centralized, poor scalability
Cheng et al. [5]	TEE, smart contracts	Secure transactions	High runtime overhead, broad attack surface
Khemaissia et al. [8]	Hybrid blockchain-fog	Reduced transaction time	Infrastructure complexity, limited scalability
Boumezbeur and Zarour [3]	Blockchain, encryption	Privacy-preserving sharing	Low computational efficiency

securing healthcare data by implementing smart contract-based access control mechanisms. Their findings indicate that blockchain can significantly improve data transparency while ensuring restricted access based on predefined security policies, yet it lacks sufficient privacy safeguards against insider threats due to weak key management.

Furthermore, Mishra and Yadav [10] investigated the integration of blockchain with InterPlanetary File System (IPFS) for decentralized healthcare data management. Their approach reduces storage costs and enhances data availability but introduces challenges in retrieval latency, compromising scalability for real-time applications. Yang et al. [15] developed a blockchain-based Electronic Health Record (EHR) framework aimed at preventing unauthorized access and data manipulation by maintaining an immutable activity log within the blockchain network, but its reliance on basic cryptographic methods leaves it vulnerable to advanced attacks like Sybil attacks. In Bowman et al. [4], Private Data Objects (PDOs) are proposed, allowing multiple untrusted entities to execute smart contracts on sensitive personal data using Intel SGX. PDOs rely on an interpreter enclave to securely process smart contracts within a controlled environment, yet this centralization reduces scalability and introduces a single point of failure.

Cheng et al. [5] introduced Ekiden, a system that integrates privacy-preserving smart contracts with a blockchain-enabled Trusted Execution Environment (TEE). Ekiden employs a Proof-of-Publication protocol to facilitate secure blockchain transactions, including consensus mechanisms reliant on authorized timers. However, this method may lead to increased runtime overhead, potential security risks, and a broader attack surface, but its high runtime overhead and broader attack surface undermine efficiency and security. Khemaissia et al. [8] introduced a hybrid blockchain-fog computing model to address performance bottlenecks in healthcare applications. Their solution reduces transaction processing time but requires additional infrastructure for fog computing deployment, increasing complexity and limiting scalability across diverse healthcare settings, with trade-offs including higher deployment costs and dependency on fog node reliability. Boumezbeur and Zarour [3] proposed a blockchain-enabled electronic health record (EHR) system with a focus on privacy-preserving data sharing. However, its heavy reliance on homomorphic encryption sacrifices computational efficiency, weakening its practicality for large-scale use, as it incurs significant computational overhead and latency.

To systematically compare these SOTA approaches, table 1 summarizes their key characteristics and shortcomings. This analysis highlights that existing solutions often fail to balance scalability, privacy, and security effectively. For instance, methods like [17] and [3] prioritize privacy at the expense of scalability and efficiency, while [10] and [8] improve performance but compromise on retrieval latency or infrastructure demands. Security vulnerabilities, such as susceptibility to Sybil attacks [15] or insider threats [1], further underscore the need for a comprehensive solution. Our proposed framework overcomes these limitations by integrating SHA-256 and RSA-based access control for enhanced privacy, a PoW consensus for scalable security against Sybil attacks, and an optimized architecture that minimizes latency without additional infrastructure.

3 Proposed methodology

This section presents a blockchain based framework aimed at enhancing the secure storage of sensitive records in public health centers. The proposed solution focuses on ensur-

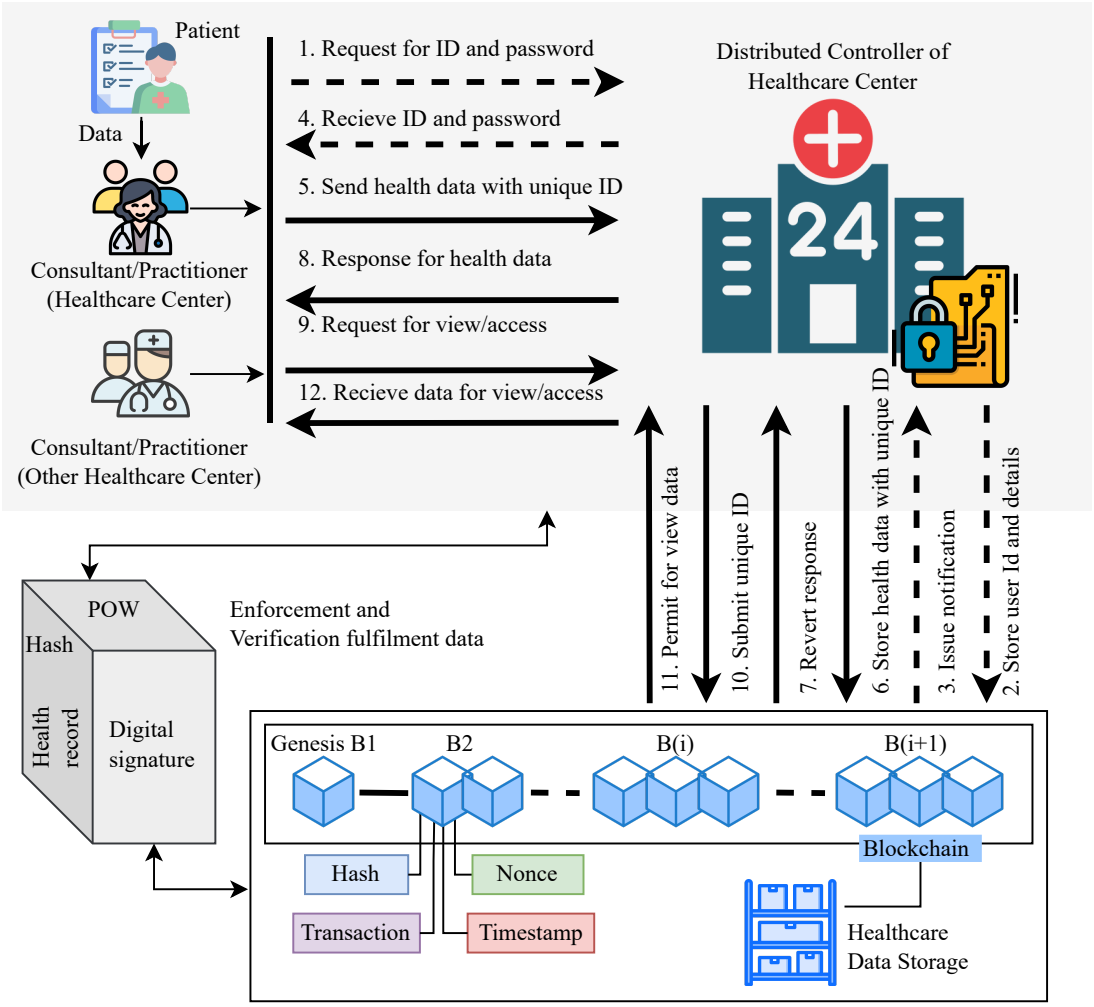


Figure 2: Framework diagram of the blockchain-based healthcare model, illustrating the sequence of interactions among users (consultants), the distributed controller, and the blockchain ledger. Step 7 ('Revert response') is conditional and occurs only if verification fails (e.g., after Step 10, if the unique ID or access request is invalid)

ing the confidentiality and integrity of healthcare processes while streamlining operations. It further describes the coordination of healthcare activities, sequence diagrams, and the workflow of block generation.

3.1 Healthcare center modeling using blockchain

The proposed framework is designed to improve healthcare system efficiency by reinforcing privacy, security, integrity, and verifiability of authorized health records. This section describes the system's overall architecture, as illustrated in fig. 2, which provides a comprehensive diagram of the blockchain-based healthcare data management framework for public health centers. The architecture integrates user interactions (e.g., consultants requesting ID/password and submitting health data, Steps 1, 5, 9), access control mechanisms managed by the distributed controller (Steps 1-4, 7-10), enforcement and verification processes (Step 11), and blockchain operations including secure data storage, PoW consensus, and block creation (Steps 2-3, 6). These components collectively ensure privacy, integrity, and secure access to health records. By implementing a blockchain based secure cloud storage model, this solution ensures high-performance authorization frameworks that uphold privacy and data integrity.

This blockchain based system enhances the reliability and scalability of health data, fostering greater confidence among stakeholders in data services. The core functional elements of the platform revolve around three key entities: the health data users/consultants, the health center controller, and the secure data repository. The health center controller governs the activities of data users/consultants within the blockchain based system. Each user/consultant acts as a data custodian, responsible for managing patient health records securely. Within this model, only authorized individuals can access, create, or update smart contracts related to health records on the blockchain. Consultants collect patient information and store it on the blockchain using public keys, while a unique private key is assigned to each

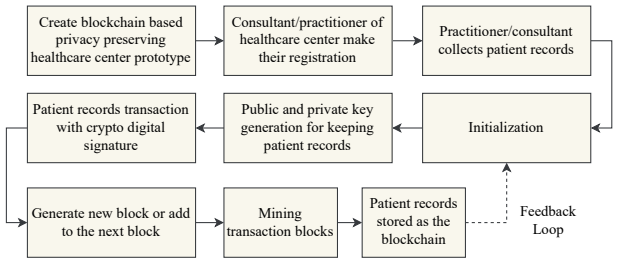


Figure 3: Workflow diagram for the proposed blockchain-based system for healthcare center, illustrating the detailed process of health data management. The ”Patient records transaction with crypto digital signature” step uses ECDSA and SHA-256 for secure signing, while ”Mining transaction blocks patient records” employs PoW consensus. A feedback loop from ”patient records stored as the blockchain” to ”Initialization” ensures error correction if verification fails.

patient's record for enhanced security.

Before gaining access, the health center controller verifies and registers each health data user/consultant, assigning a unique digital identity (ID) and password through a secure authentication framework. This registration process leverages the Bcrypt algorithm to generate encrypted digital IDs, which are then securely stored. After completing registration, authorized users can securely access the blockchain network with their digital credentials, ensuring robust authentication and strict access control.

In a system protected by blockchain, any action involving the database such as storing, modifying, or distributing data, interacts with the blockchain to maintain privacy, security, and accessibility. By utilizing a network of verified user nodes, the system guarantees the safety and accuracy of health records. Transactions are signed cryptographically and stored immutably on the blockchain, with data integrity ensured using SHA-256 hashing. The Proof-of-Work (PoW) consensus mechanism strengthens decentralization by requiring miners to expend computational effort to solve cryptographic puzzles, appending new blocks only after achieving a valid hash. This prevents unauthorized modifications in the healthcare context by making tampering economically infeasible as altering a patient record would require re-mining all subsequent blocks across the distributed network, a task computationally prohibitive due to the PoW difficulty. This ensures the immutability of medical histories critical for treatment accuracy, though it introduces energy consumption trade-offs, as analyzed in section 5. Any tampering is detected through sequential hash verification, including nonce values, timestamps, and encoded data. If an anomaly is identified, the affected block is flagged as invalid.

The operational process of data exchange in this model follows a structured approach:

1. Users/consultants authenticate themselves using their assigned IDs and passwords.
2. Upon verification, they request access to blockchain storage via the health center controller.
3. Encrypted health records are then submitted for storage and later retrieved as needed.
4. Each health record is assigned a Unique Data Identity (DID) for tracking purposes.
5. The controller manages the storage and retrieval of patient data using the assigned DID.

The DID is a composite identifier structured as a 256-bit hash, generated by concatenating the patient's unique identifier (a UUID assigned during registration), a timestamp of the record submission, and a hash of the health data itself using SHA-256. This structure ensures uniqueness and traceability: the UUID links the DID to the patient, the timestamp provides temporal context, and the health data hash ties the DID directly to the specific record. The DID is stored on the blockchain alongside the encrypted health record, enabling efficient retrieval and verification by authorized users while maintaining immutability and integrity.

The overall workflow of this blockchain-powered system is depicted in fig. 3, outlining each step involved in the development of a secure, transparent, and efficient health data management platform. Specifically, the step ”Patient records transaction with crypto digital signature” involves generating a digital signature using the ECDSA algorithm (as detailed in Algorithm 2) and hashing the transaction data with SHA-256 to ensure integrity and authenticity. The ”Mining transaction blocks patient records” step employs the Proof-of-Work (PoW) consensus mechanism, where miners solve cryptographic puzzles to validate transactions and append new blocks, as described in Algorithm 3. Additionally, a feedback loop exists: if a transaction fails verification (e.g., due to an invalid signature or hash mismatch), the system reverts to the ”Initialization” phase, prompting the practitioner to resubmit corrected data, ensuring robust error handling and data integrity.

Unlike hybrid blockchain-fog computing models [8], which rely on additional infrastructure and introduce deployment complexity, our framework avoids these trade-offs by leveraging a fully decentralized blockchain architecture with optimized PoW consensus, eliminating the need for external fog nodes while maintaining low latency. Similarly, in contrast to homomorphic encryption-based approaches [3], which incur high computational overhead, our system employs SHA-256 hashing and RSA encryption, achieving privacy and integrity with reduced processing demands.

3.2 Threat model

To ensure the robustness of the proposed blockchain-based healthcare framework, we define a threat model that identifies potential attacks and outlines corresponding mitigation strategies. This model focuses on three primary threats prevalent in healthcare data systems:

- **Sybil Attacks:** In a Sybil attack, an adversary creates multiple fake identities to gain disproportionate influence over the network. Our framework mitigates this through the PoW consensus mechanism, which requires computational effort to validate transactions and generate blocks. This makes it economically infeasible for an attacker to dominate the network with fake nodes.

- **Replay Attacks:** Replay attacks involve an adversary resubmitting valid transactions to disrupt the system or gain unauthorized access. The system defends against this by incorporating timestamped nonce values in each transaction, verified through sequential hash checks using SHA-256. Any attempt to reuse a transaction is detected and rejected, ensuring transaction uniqueness.

- **Insider Threats:** Insider threats occur when authorized users misuse their access privileges. To counter this, the framework employs RSA-encrypted access control, where each user's private key is required to sign transactions, and permissions are strictly enforced by the health center controller. Additionally, the ECDSA verifies signatures at both user and server levels, preventing unauthorized modifications, as outlined in Algorithm 2.

These mitigation strategies are integrated into the system's architecture and cryptographic operations, ensuring comprehensive protection against these threats.

3.3 Interoperability with healthcare standards

To ensure practical deployment in healthcare environments, the proposed framework is designed to comply with key standards, including Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR), Digital Imaging and Communications in Medicine (DICOM), and Health Insurance Portability and Accountability Act (HIPAA) regulations. HL7 FHIR is supported by mapping blockchain-stored health records to FHIR resources (e.g., Patient, Observation) using standardized JSON encodings, enabling seamless data exchange with existing EHR systems. For medical imaging, DICOM compatibility is achieved by storing DICOM metadata hashes on-chain, with actual image files linked via secure off-chain storage (e.g., IPFS), balancing efficiency and integrity. Compliance with the HIPAA is ensured through RSA-encrypted access control and ECDSA signatures, safeguarding patient data confidentiality and auditability as mandated by HIPAA privacy rules. This interoperability enhances the framework's integration with diverse healthcare providers, addressing a critical limitation noted in section 2.

3.4 Smart contract security and audit methodology

The proposed framework incorporates smart contracts to manage access control and data transactions, necessitating safeguards against common vulnerabilities. To prevent reentrancy attacks, where a malicious contract repeatedly calls back into the system before state updates, we implement a non-reentrant modifier inspired by Ethereum's Solidity patterns, ensuring that state changes (e.g., access permission updates) are finalized before external calls. Integer overflows, a risk in arithmetic operations, are mitigated by using Python's arbitrary-precision integers in our Flask-based implementation, avoiding fixed-size integer limitations. For instance, nonce increments in PoW and transaction counters are immune to overflow due to this design choice.

To verify these protections, we outline a security audit methodology comprising: (1) static analysis using tools like Slither to detect reentrancy and overflow risks in smart contract logic, (2) dynamic testing with simulated attack scenarios (e.g., recursive calls) on a testnet, and (3) formal verification of critical functions (e.g., signature verification in Algorithm 2) using mathematical proofs of correctness. This multi-layered approach ensures robustness against vulnerabilities, complementing the cryptographic security detailed in section 3.2.

3.5 Coordination process

The proposed model ensures that a distributed controller in public health centers regulates access to blockchain-stored medical data. In this framework, three key participants are involved, medical practitioners/consultants (users who interact with health data), The distributed controller (responsible for registration and access management), and the blockchain ledger (which functions as a decentralized, secure database for patient health information). Blockchain serves as a highly distributed and tamper-resistant ledger, allowing only authorized entities to conduct transactions involving patient records. Participating nodes can input, retrieve, and verify data securely. The sequence diagram illustrating user activities within this model is shown in fig. 4, depicting the order of actions taken to interact with the system.

3.5.1 Registration

The framework initiates with a structured registration process designed to authenticate and onboard medical practitioners or consultants. To begin, practitioners submit a registration request to the health center controller, accompanied by relevant professional credentials. The controller then verifies the applicant's registration status; if unregistered, the applicant's details are encrypted and securely stored on the blockchain to ensure immutability and data integrity. Following successful verification, the system dynamically generates a unique user identity and password,

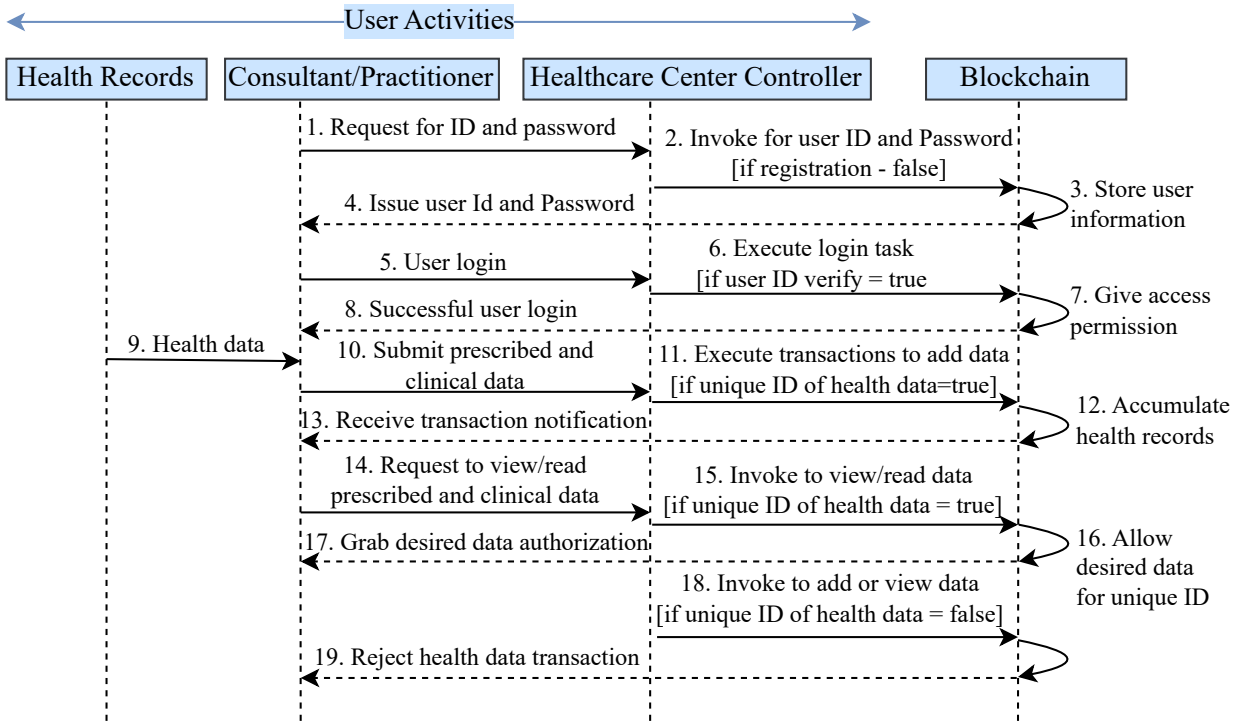


Figure 4: Diagram showing the sequence of user activities involved in blockchain based healthcare data center

which are disseminated to the practitioner via a decentralized, distributed controller. This authentication mechanism grants registered users perpetual access to the blockchain network, enabling them to log in at any time and participate in subsequent operations.

3.5.2 Data submission

Once authenticated, the framework transitions to the data submission phase. Consultants who are now part of the blockchain network, collect patient-specific clinical data such as diagnostic reports or prescribed treatments and submit these records to the health center controller. Upon receipt, the controller assigns a unique health identity to the patient, consolidating their medical history into a singular, traceable profile. To fortify security, each record is encoded with a hash based value by the blockchain, creating an indelible digital fingerprint that prevents tampering and enhances auditability. Access to upload or modify health records is rigorously restricted to authorized practitioners, with permissions governed by cryptographic keys to ensure compliance and prevent unauthorized alterations. Upon successful entry of data into the blockchain, all network participants, including relevant healthcare providers and the patients, receive instantaneous confirmation of the transaction. This real-time notification mechanism fosters transparency, accountability, and trust across the ecosystem, ensuring that stakeholders remain informed of updates to patient records while maintaining adherence to privacy and regulatory standards.

3.5.3 Data access and viewing

This section provides an in-depth explanation of the blockchain based framework for managing public health records, ensuring the confidentiality, security, and integrity of sensitive health data. The suggested system enables verified medical professionals and consultants to connect with the blockchain via the health center controller. This setup grants them access to previous health records while enforcing strict security measures to regulate access.

To retrieve patient health data, medical practitioners must first authenticate themselves using their unique digital identity and password, which is verified through the distributed controller. They must also provide the unique patient identifier corresponding to the specific individual's health record. Once the request is submitted, the health center controller verifies the identifier's validity and grants access to the requested clinical data if the credentials match the stored records. This process ensures that only authorized personnel can view and analyze previous prescriptions and medical history. If the verification process identifies an invalid patient identifier, the blockchain system automatically rejects the access request. The ability to securely retrieve prior medical records enables healthcare professionals to monitor patient health progress, formulate new prescriptions, and update the system accordingly.

The interactions among different entities within this system facilitate the seamless integration of blockchain technology into public healthcare data management. The sequence diagram in fig. 5 illustrates the blockchain trans-

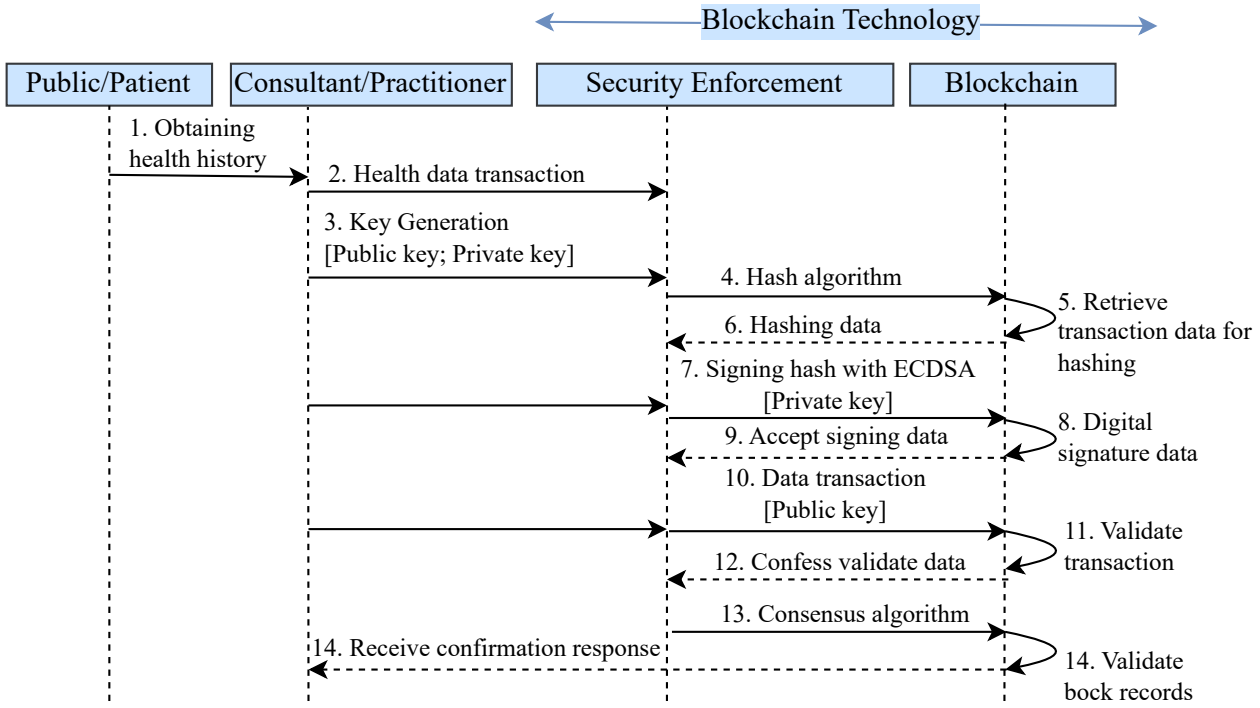


Figure 5: Diagram showing the sequence of processes in a blockchain transaction.

action process, which involves medical consultants, security mechanisms, and the underlying blockchain operations. When consultants collect health records from patients, they interact with the blockchain to execute transactions, ensuring that all health data is securely stored and processed. The system employs a key generation algorithm to create distinct public and private key pairs for secure data access. These keys, generated randomly, enhance the security of transactions within the platform by allowing encrypted communication between users and the blockchain ledger.

To reinforce data security, the blockchain framework incorporates cryptographic hashing and digital signing mechanisms. When a consultant shares a patient's health record with another authorized entity, no physical data transfer occurs. Rather than transferring data directly, the transaction is carried out by adjusting access permissions within the blockchain network. Every request to reallocate access must be authenticated by a digital signature, created using the sender's private key and validated against their public key to confirm legitimacy. The SHA-256 hash function is central to these operations, generating a unique cryptographic hash for each transaction and allowing data retrieval through hashed URLs within the web based platform. Additionally, participating nodes digitally sign transactions to confirm their authenticity, ensuring that data integrity remains uncompromised.

A crucial aspect of this blockchain based system is the PoW consensus mechanism, which governs block generation and verification. Nodes participating in the network validate health transactions, and miners are selected to generate the next block. This process ensures that all stored health data remains tamper-proof and synchronized across the distributed ledger. The RSA cryptographic algorithm and SHA-256 hashing play a fundamental role in securing the data, encrypting transactions, and maintaining the confidentiality of patient records. The blockchain storage mechanism ensures that all medical data remains accessible while being protected against unauthorized modifications.

Algorithm 1 Generate Transaction Key-pair

```
procedure transactionKeyGenerate
    if user U_h submits a health data transaction request to the health center controller then
        Procedure set(transactionKeyGenerate)
        randomValue ← generate_random(
                          cryptographic_value)
        privateKey ← generate_RSA(
                        1024, randomValue,
                        padding=OAEP, mode=encryption)
        publicKey ← privateKey · Kpb(i)
        decode_in_PEM_ascii(privateKey, publicKey)
        get privateKey, publicKey
    else
        do nothing
    end if
end procedure
```

The system's structural integrity relies on the SHA-256 cryptographic hash, which maintains the hierarchical organization of blockchain data blocks. Each new block added to the chain contains essential components such as the block number, timestamp, nonce, blockchain address, and hash of the previous block. The block header links all blocks to-

Algorithm 2 Digital Signature Generation for Data Transaction

```
procedure GenerateSignature(transaction)
    if user U_h requests transaction T over Blockchain then
        T ← make T exclude sender's privateKey
        privateKey ← GenerateRSAKey(
                        sender's privateKey)
        transactionSigner ← CreateDigitalSignature(
                                privateKey)
        hashValue ← ComputeHash(standard_input_value)
        signedTransaction ← SignTransaction(
                                transactionSigner,
                                hashValue)
        return EncodeASCII(signedTransaction)
    else
        don't create signature for T
    end if
end procedure
```

Algorithm 3 Creation and Addition of New Block for Hash based Health Data Transactions

```
procedure createNewBlock(healthData)
    Initialize healthTransactions ← ∅
    Set block ← (blockNumber, healthTransactions, nonce,
                 timestamp, previousBlockHash)
    if blockNumber = length(chain) + 1 then
        Append block to chain
        Reset running transaction status
        Add block to chain
    end if
    if EncodeJSON(block, standard_value) as a file then
        hash256Value ← CreateSHA256Hash()
        Update hash256Value.blocks
        return HexEncode(hash256Value)
    else
        Do nothing
    end if
end procedure
```

Algorithm 4 Append Node to Health Blockchain Network

```
procedure registrationRequest(node)
    if health data user requests a node then
        Create registrationNode
    end if
    Initialize parameters: (healthTransaction, chain,
                            nodes, genesisBlock)
    urlNodeInfo ← ParseURL(urlNode)
    if sets urlNodeInfo.Netloc and urlNodeInfo.path then
        Add urlNodeInfo.Netloc and urlNodeInfo.pathto
            nodes
    else
        Append a new node to nodes
    end if
    Do not append
end procedure
```

Algorithm 5 Secure Health Data Management on Blockchain

```
procedure AccumulateAndPublish
    Initialize transaction parameters for healthcare data
    Create a new transaction block
    Authenticate digital transaction signature
    publicKey ← GetRSAPublicKey()
    signatureVerifier ← CreatePKCS1Verifier(
                            publicKey)
    dataHash ← GenerateSHAHash(health_transaction)
    dataHash ← GenerateSHAHash(
                health_transaction)
    VerifySignature(dataHash,
                    hexTransactionSignature)
    Execute PoW mechanism
    Fulfull mining conditions for valid proof
    Synchronize blockchain nodes
    Confirm blockchain transaction validity
    Append data to the transaction chain
    if transaction signature is verified then
        Transfer or share data between nodes
        Grant access to health records
    end if
end procedure
```

gether, ensuring seamless verification and establishing an immutable record of all transactions.

The Elliptic Curve Digital Signature Algorithm (ECDSA) further strengthens data security by verifying digital signatures at both the user and server levels. If any discrepancy is detected such as tampering or altered data values, the transaction is immediately discarded, preventing unauthorized modifications to stored health records. This two-tier verification approach ensures that only authenticated users can access and manage data, mitigating risks associated with data forgery.

Within the blockchain framework, each health transaction generates a public-private key pair to facilitate secure encryption and decryption. Algorithm 1 illustrates the RSA based key pair generation process, which assigns unique keys to all participants for secure interactions. The PEM method maps these cryptographic keys with ASCII encoding for seamless transaction decoding. Algorithm 2 details the process of generating digital signatures for health data transactions, ensuring encrypted authentication during network interactions.

To maintain data integrity, algorithm 3 outlines the steps involved in creating and adding new health data blocks to the blockchain. Each transaction block consists of essential attributes such as the block number, health data transaction, nonce, timestamp, and hash value. These components collectively ensure that public health records remain secure, transparent, and immutable. Algorithm 4 describes the procedure for integrating new nodes into the blockchain based public health center network, allowing for distributed and decentralized transaction processing. Algorithm 5 defines how public health data is accumulated, accessed, and managed within the blockchain.

The PoW consensus mechanism plays a vital role in the blockchain based public health network by validating transactions before they are shared or accessed across different nodes. This method dynamically adjusts the number of participating nodes, ensuring optimal network scalability and accelerating the health data transaction process. Once the transactions are validated, the system publishes the data blocks within the blockchain ledger, preserving the security and integrity of all recorded public health center transactions.

3.6 Complexity analysis of proposed algorithms

To evaluate the computational efficiency of the proposed framework, we analyze the time complexity of key algorithms, focusing on transaction signing (Algorithm 2) and block validation (Algorithm 3).

For Algorithm 2 (Digital Signature Generation), the primary operations are RSA key generation, SHA-256 hashing, and ECDSA signing. RSA key generation has a complexity of $O(k^3)$, where k is the key size (1024 bits in our implementation). SHA-256 hashing operates in $O(n)$, where n is the input data size, typically small for transaction metadata. ECDSA signing has a complexity of $O(k^2)$ due to elliptic curve operations. Since these steps are sequential, the overall complexity is dominated by RSA key generation, yielding $O(k^3)$.

For Algorithm 3 (Block Creation and Addition), the complexity hinges on SHA-256 hashing and the PoW mechanism. Hashing the block data is $O(n)$, where n is the block size. However, PoW requires solving a cryptographic puzzle by iterating nonce values until a valid hash is found, with an expected complexity of $O(2^d)$, where d is the difficulty level (configurable but typically fixed for consistent performance). Thus, block validation complexity is dominated by PoW, resulting in $O(2^d)$.

These complexities indicate that transaction signing is computationally intensive due to RSA, while block validation's efficiency depends on the PoW difficulty, a trade-off for security against Sybil attacks.

4 Results

This section provides an in-depth evaluation of the proposed blockchain based secure storage management system for public health centers, focusing on user nodes and assessing its ability to meet data security and privacy requirements. Experimental assessments and qualitative analyses were conducted to validate the system's effectiveness in ensuring confidentiality, integrity, and availability of health records. The results presented here focus on raw performance metrics, with comparisons and interpretations deferred to the subsequent Discussions section.

4.1 Evaluation environment

The experimental setup was conducted in a controlled environment using a Windows 11 system equipped with an Intel(R) Core i7-12700H processor (2.3GHz, 14 cores, x64 based architecture) and 16 GB RAM. The implementation leveraged Python 3.11 (64-bit) with Flask 2.2.5 as the web server gateway interface, ensuring seamless interactions between user nodes and the blockchain. The evaluation approach integrated multiple user nodes and a blockchain node to assess the performance and scalability of the health data management framework under real-world operational conditions. To enhance the robustness of the evaluation, additional tests were conducted on a cloud-based environment (AWS EC2 t2.medium instance with 2 vCPUs and 4 GB RAM) and a resource-constrained device (Raspberry Pi 4 with 1.5 GHz quad-core processor and 4 GB RAM) to simulate diverse deployment scenarios.

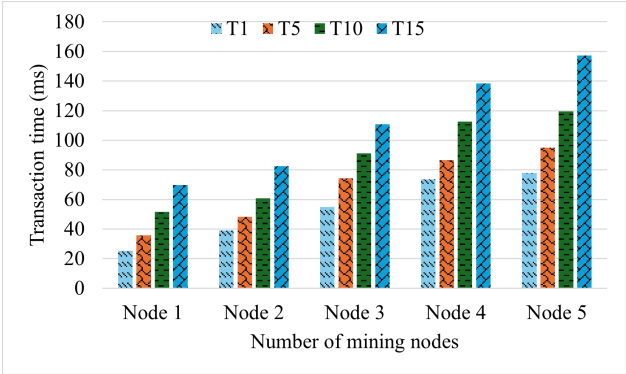


Figure 6: Average transaction time for different user nodes.

4.2 Performance evaluation

To measure performance, multiple user nodes were configured to interact with the blockchain server, enabling the evaluation of key parameters such as transaction time and latency. The system was evaluated by having user nodes conduct transactions in batches of 1, 5, 10, and 15 (T1, T5, T10, and T15) within the blockchain network. As each transaction block was mined and distributed across the blockchain-powered health center, the test provided valuable insights into the system's ability to manage growing transaction volumes efficiently.

4.3 Transaction time

Transaction time was recorded to assess how quickly the system processed data across different user nodes. The results, as illustrated in fig. 6, indicate that for node 1, transaction time ranged from 25.2ms for T1 to 69.8ms for T15. For node 5, transaction time started at 77.9ms for T1, reaching a peak of 157.4ms for T15, reflecting the system's performance under increasing transaction loads.

4.4 Latency

Another key performance metric analyzed was latency, which represents the delay experienced in transmitting and receiving data blocks within the blockchain network. As shown in fig. 7, latency for node 1 was 10.42ms for T1 and increased to 20.32ms for T15. For node 5, latency began at 25.18ms for T1 and reached a peak of 37.5ms for T15, demonstrating a consistent increase with higher transaction volumes.

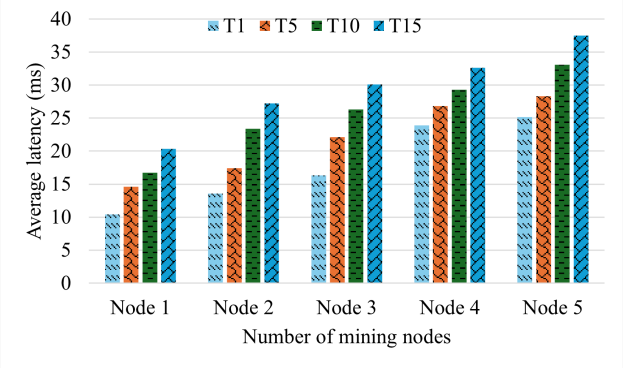


Figure 7: Average latency (ms) for different number of user nodes

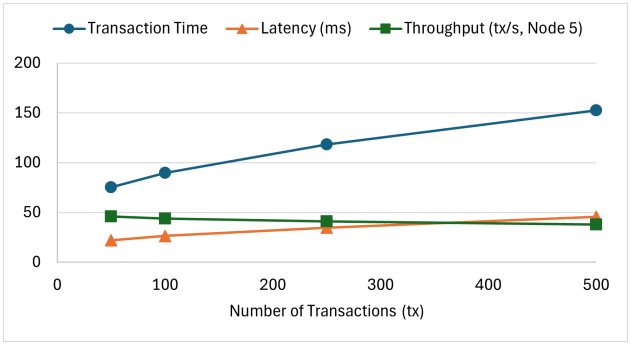


Figure 8: Performance Metrics with Increasing Blockchain Storage (Node 1 unless specified)

4.5 Functionality comparison

A detailed analysis was performed to evaluate the functional capabilities of the proposed system in comparison to other blockchain-based healthcare data storage solutions. Table 2 provides a comparative overview, highlighting the presence or absence of key security features, including availability (A), confidentiality (C), integrity (I), server-side verifiability (SSV), and user-side verifiability (USV), with a simple Yes or No indication. The findings reveal that many existing schemes lack essential security features necessary for comprehensive health data protection. In contrast, the proposed model ensures robust verifiability, data confidentiality, and system integrity, making it a more reliable and effective solution for secure healthcare data storage and management.

Table 2: Functionalities comparison

Ref.	A	C	I	SSV	USV
[16]	No	Yes	Yes	No	No
[14]	Yes	Yes	Yes	No	No
[9]	Yes	No	Yes	No	No
[11]	No	No	Yes	No	No
[2]	No	No	Yes	No	No
[13]	No	Yes	Yes	No	No
Our work	Yes	Yes	Yes	Yes	Yes

4.6 Impact of blockchain storage growth

To evaluate the system's performance as blockchain storage grows, we simulated an increasing number of healthcare transactions in batches of 50, 100, 250, and 500. Transaction time, latency, and throughput were measured for these batches. Results, summarized in fig. 8, indicate that for node 1, transaction time increased from 75.3ms for 50 transactions to 152.6ms for 500 transactions, and latency rose from 22.1ms to 45.8ms. For node 5, throughput, defined as transactions per second, ranged from 46 tx/s for 50 transactions to 38 tx/s for 500 transactions, reflecting performance under growing storage loads.

4.7 Storage overhead analysis

To assess the system's scalability for long-term medical record storage, we analyzed the storage overhead as blockchain data accumulates. Each transaction block, containing a header (block number, timestamp, nonce, previous hash) and health data (e.g., patient ID, hashed record), averages 1 KB, based on SHA-256 hash outputs (32 bytes) and metadata. For 500 transactions (fig. 8), the blockchain size is approximately 0.5 MB per node. Extrapolating to a healthcare center processing 10,000 transactions annually, the storage grows to 10 MB per year per node. Over a decade, this reaches 100 MB, manageable for modern hardware but potentially burdensome in a multi-node network with hundreds of centers.

To handle long-term storage efficiently, we propose an off-chain approach using the IPFS. Full medical records (e.g., text, images) are stored on IPFS, with only their cryptographic hashes and access metadata recorded on-chain. This reduces on-chain growth to ĩ00 bytes per transaction (hash + metadata), shrinking annual storage to 1 MB and decadal storage to 10 MB per node, a 90% reduction. IPFS ensures data availability and integrity via content-addressing, while blockchain enforces access control, balancing scalability and security for lifelong patient records.

4.8 Ablation study on cryptographic elements

An ablation study was conducted to assess the individual contributions of cryptographic components (SHA-256 hashing, RSA encryption, and ECDSA signatures) to performance. We evaluated four configurations for T10 on node 1: (1) full system (SHA-256 + RSA + ECDSA), (2) without ECDSA, (3) without RSA, and (4) without SHA-256. Results, detailed in table 3, show that the full system recorded a transaction time of 48.5ms and latency of 15.3ms. Removing ECDSA reduced transaction time to 42.1ms and latency to 13.8ms, excluding RSA lowered

Table 3: Ablation study results for T10 on node 1

Configuration	Transaction Time (ms)	Latency (ms)	Security Impact
Full System (SHA-256 + RSA + ECDSA)	48.5	15.3	None
Without ECDSA	42.1	13.8	Vulnerable to signature forgery
Without RSA	39.8	12.9	Weakened access control
Without SHA-256	36.4	11.5	Reduced data integrity

them to 39.8ms and 12.9ms, and omitting SHA-256 resulted in 36.4ms and 11.5ms, reflecting the performance impact of each component.

5 Discussions

This section compares the results of the proposed system with state-of-the-art (SOTA) solutions, providing numerical performance comparisons, analyzing trade-offs, and addressing the implications of design choices such as the use of PoW and SHA-256.

5.1 Comparison with state-of-the-art solutions

The proposed system's performance was benchmarked against a centralized SQL-based system and a SOTA blockchain-based framework [17], as shown in fig. 9. For T10 on Node 1, our system achieved a latency of 15.3 ms, compared to 8.1 ms for the SQL baseline and 25.2 ms for [17]. On Node 8, latency increased to 37.6 ms, higher than the SQL baseline's 12.4 ms but less than [17]'s 52.6 ms. Compared to other SOTA approaches in table 1, such as [10] (high retrieval latency) and [8] (reduced transaction time but complex infrastructure), our system offers a balanced performance, avoiding excessive computational overhead or external dependencies.

5.2 Performance trade-offs

The proposed system exhibits higher latency and transaction times than the SQL baseline due to the decentralized nature of blockchain and the computational overhead of PoW and cryptographic operations. However, this trade-off enhances security: the SQL system lacks tamper-proofing and decentralized trust, making it vulnerable to single-point failures and unauthorized access. Compared to [17], our lower latency (15.3 ms vs. 25.2 ms on Node 1) stem from an optimized PoW implementation and efficient SHA-256

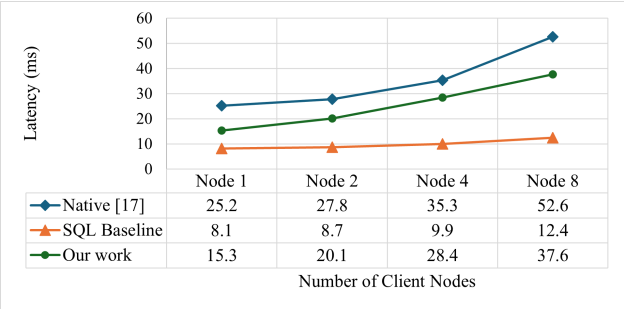


Figure 9: Comparison of latency vs number of client nodes for publishing a transaction

hashing, despite similar security goals. The ablation study (table 3) confirms that removing cryptographic elements reduces latency (e.g., 11.5 ms without SHA-256) but compromises security, justifying the performance cost for robust protection in healthcare settings.

5.3 Energy consumption analysis of PoW

The use of PoW introduces significant computational cost and energy consumption, as miners must solve complex cryptographic puzzles to validate transactions and generate blocks. For 500 transactions (fig. 8), transaction time reached 152.6 ms, reflecting this overhead. In large-scale healthcare deployments, this could lead to high energy usage, raising environmental concerns and operational costs. While PoW ensures strong resistance to Sybil attacks, its scalability is limited compared to Proof-of-Stake (PoS), as throughput stabilized at 38 tx/s for 5 nodes. In contrast, PoS eliminates mining by selecting validators based on staked assets, reducing energy consumption to near-negligible levels. PoS offers comparable security in permissioned settings but may compromise decentralization in trustless healthcare environments, where untrusted nodes require robust Sybil resistance, as PoS could favor wealthier nodes, potentially skewing fairness among healthcare providers. While PoW's energy inefficiency is a limitation, its decentralized trust aligns with the framework's goals. Future work could explore hybrid consensus (e.g., PoW-PoS) to balance security and sustainability, tailoring the system for energy-constrained healthcare settings.

5.4 Choice of SHA-256 and PoW

SHA-256 was selected for its proven collision resistance and widespread adoption in blockchain systems like Bitcoin, ensuring data integrity with a fixed 256-bit output that is computationally efficient for healthcare data hashing. PoW was chosen over Proof-of-Stake (PoS) or Byzantine Fault Tolerance (BFT) due to its robustness in trustless environments, critical for public healthcare where nodes may not be fully trusted. BFT, effective in permissioned settings, lacks the decentralization needed for broad interoperability among healthcare providers. PoW's computa-

tional barrier, though costly, aligns with our goal of maximizing security against Sybil and insider threats, as detailed in section 3.2, making it suitable despite energy drawbacks. RSA was chosen for key generation and encryption due to its maturity, widespread support in cryptographic libraries, and compatibility with existing healthcare standards like HIPAA, which often require established asymmetric encryption methods for secure access control. While alternatives like ECC offer smaller key sizes and faster computation, RSA's larger key sizes (e.g., 1024 bits) provide a higher security margin against brute-force attacks, critical for protecting sensitive health data over long-term storage.

5.5 Scaling challenges in real-world healthcare environments

While the proposed framework demonstrates scalability for up to 500 transactions, real-world healthcare environments present additional challenges. A large-scale healthcare network, such as a national health system, may involve hundreds or thousands of nodes (e.g., hospitals, clinics, and labs), significantly increasing network complexity. Our evaluation with 5 nodes showed a throughput of 38 tx/s for 500 transactions, but as node count scales to hundreds, network latency could rise due to increased consensus overhead in PoW, potentially exceeding 100 ms per transaction based on linear extrapolation. This could impact real-time access to critical patient records during emergencies, necessitating optimizations like sharding or hybrid consensus mechanisms (e.g., combining PoW with PoS).

Integration with legacy healthcare systems also poses a challenge. While our framework complies with standards like HL7 FHIR and DICOM, many healthcare providers rely on outdated systems lacking API support for blockchain integration. For instance, a hospital using a 1990s-era EHR system may require custom middleware to map data to FHIR resources, introducing deployment complexity and potential security risks during data migration. Additionally, the computational requirements of PoW and RSA encryption (e.g., $O(k^3)$ complexity for RSA key generation, may strain resource-constrained environments, such as rural clinics with limited hardware. Future work could explore lightweight cryptographic alternatives (e.g., ECC instead of RSA) and off-chain computation to mitigate these challenges, ensuring broader adoption across diverse healthcare settings.

6 Conclusion

This study tackles the increasing need for digital transformation in healthcare by introducing a secure storage system built on a privacy-focused and high-performance blockchain. A standout feature of this system is its reliable access control mechanism, which facilitates the safe and efficient exchange of health records among authorized users. By structuring functional units and implementing a systematic blockchain-enabled process, the model facilitates decentralized data storage while maintaining data confidentiality, availability, integrity, and scalability.

The proposed framework significantly enhances security compared to conventional healthcare data management schemes, allowing consultants or users to store and access medical records securely. Performance evaluation metrics, including publishing transaction time cost and latency, demonstrate the system's ability to efficiently process health data transactions across multiple user nodes. Compared to traditional storage methods, this blockchain based approach offers a robust and reliable solution for managing health records in public health centers.

Despite its advantages, the system's reliance on blockchain technology presents a potential limitation in terms of energy consumption during data transactions and storage. This raises environmental concerns, particularly as blockchain networks grow in scale. Future research will focus on optimizing energy efficiency and extending the framework to include metadata auditing for cloud storage, further enhancing security and data management capabilities in healthcare environments.

Funding

Independent fund project of Shaanxi Provincial Intelligent New Criminal Technology Practical Application Research Center (SXZHXJ202202)

References

[1] M. Alshare et al. "Assessing Blockchain's Role in Healthcare Security: A Comprehensive Review". In: *Informatica* (2024). url: https://www.informatica.si/index.php/informatica/article/view/6155.

[2] N. Al Asad et al. "Permission-Based Blockchain with Proof of Authority for Secured Healthcare Data Sharing". In: *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*. IEEE, 2020, pp. 35–40.

[3] I. Boumezbeur and K. Zarour. "Privacy-Preserving and Access Control for Sharing Electronic Health Record Using Blockchain Technology". In: *Acta Informatica Pragensia* (2022). url: https://www.ceeol.com/search/article-detail?id=1061177.

[4] M. Bowman et al. "Private Data Objects: an Overview". In: *arXiv preprint* arXiv:1807.05686 (2018).

[5] Raymond Cheng et al. "Ekiden: A Platform for Confidentiality-Preserving, Trustworthy, and Performant Smart Contracts". In: *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. 2019, pp. 185–200. doi: 10.1109/EuroSP.2019.00023.

[6] M.S. Islam et al. “Healthcare-Chain: Blockchain-Enabled Decentralized Trustworthy System in Healthcare Management Industry 4.0 with Cyber Safeguard”. In: *Computers, MDPI* 12.2 (2023), p. 46.

[7] A. Khatoon. “A blockchain-based smart contract system for healthcare management”. In: *Electron.* 9.1 (2020).

[8] R. Khemaissia et al. “Prschain: A Blockchain-Based Privacy Preserving Approach for Data Service Composition”. In: *Informatica* (2023). url: https://www.informatica.si/index.php/informatica/article/view/5081.

[9] A. R. Lee, M. G. Kim, and I. K. Kim. “SHAREChain: Healthcare data sharing framework using Blockchain-registry and FHIR”. in: *Proc. - 2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2019*. 2019, pp. 1087–1090.

[10] R.K. Mishra, R.K. Yadav, and P. Nath. “Integration of Blockchain and IPFS: Healthcare Data Management & Sharing for IoT Environment”. In: *Multimedia Tools and Applications* (2024).

[11] C. D. Parameswari and V. Mandadi. “Healthcare data protection based on blockchain using solidity”. In: *Fourth World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)*. IEEE, 2020, pp. 577–580.

[12] M.A. Rahman et al. “EduChain: CIA-compliant blockchain for intelligent cyber defense of microservices in education industry 4.0”. In: *IEEE Transactions on Industrial Informatics* 18.3 (2021), pp. 1930–1938.

[13] V. Ramani et al. “Secure and Efficient Data Accessibility in Blockchain Based Healthcare Systems”. In: *2018 IEEE Glob. Commun. Conf. GLOBECOM 2018 - Proc.* 2018.

[14] H. Wang and Y. Song. “Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain”. In: *Journal of Medical Systems* 42.8 (2018), p. 152.

[15] G. Yang and C. Li. “A design of blockchain-based architecture for the security of electronic health record (EHR) systems”. In: *Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom*. Vol. 2018-December. 2018, pp. 261–265.

[16] Z. Ying et al. “A Lightweight Policy Preserving EHR Sharing Scheme in the Cloud”. In: *IEEE Access* 6 (2018), pp. 53698–53708.

[17] M.H. Al-Zubaidie and R.H. Razzaq. “Maintaining Security of Patient Data by Employing Private Blockchain and Fog Computing Technologies Based on Internet of Medical Things”. In: *Informatica* (2024). url: https://www.informatica.si/index.php/informatica/article/view/6047.

Multi-Task Deep Reinforcement Learning for Intelligent Logistics Path Planning and Scheduling Optimization

Xianfeng Zhu
School of Economics and Management, Jiaozuo University, Jiaozuo 454000, Henan, China
E-mail: zhu_xianfeng@outlook.com

Keywords: deep reinforcement learning, multi-task learning, logistics optimization, path planning, dynamic demand

Received: January 8, 2025

Intelligent logistics systems optimize path planning and scheduling problems by introducing deep reinforcement learning (DRL) to cope with the complexity of dynamic demand and resource constraints. This study proposes an improved strategy that combines multi-task learning (MTL) with Q-learning to achieve simultaneous optimization of multiple related subtasks, such as path selection, task scheduling, and resource allocation, and shares some network parameters to improve model efficiency and generalization ability. The experimental design covers a variety of logistics scenarios, including urban distribution, long-distance transportation, and emergency response. Five baseline models are used for comparative evaluation to verify the advantages of the DRL method in computational efficiency, cost control, resource utilization, service level, and dynamic adaptability. Through experimental verification, the DRL model achieved a 15% cost reduction in cost control compared to traditional algorithms; the resource utilization rate reached 85%, and it performed excellently in terms of efficiency improvement. It has excellent adaptability when dealing with dynamic demands and can respond quickly to environmental changes, effectively improving the overall performance of the intelligent logistics system. It is particularly suitable for application scenarios that require real-time decision-making and support dynamic demand fluctuations.

Povzetek: Opisana je optimizacijo poti in urnika v pametnih logističnih sistemih z uporabo globokega ojačevalnega učenja (DRL) in večnalognega učenja za učinkovito prilagajanje na dinamične zahteve, izboljšanje izkoriščenosti virov in zmanjšanje stroškov.

1 Introduction

With the rapid development of the global economy and the popularization of e-commerce, the logistics industry has become a core link in the modern supply chain. The concept of intelligent logistics has emerged. It combines technologies such as the Internet of Things, big data, and artificial intelligence to significantly improve the efficiency and flexibility of the logistics system by dynamically optimizing resource allocation and process management. In this context, logistics path planning and scheduling optimization have gradually become key areas for improving logistics efficiency. Whether it is intra-city distribution, long-distance transportation, or international logistics, these fields are faced with the common challenge of how to complete cargo transportation at the lowest cost and fastest speed [1].

In traditional logistics path planning, commonly used algorithms include Dijkstra algorithm and A algorithm, which can find the shortest path in a static environment. However, actual logistics scenarios are full of dynamic changes, such as fluctuations in order quantity, changes in traffic conditions, and sudden resource constraints. This makes it difficult for traditional algorithms to effectively cope with complex dynamic environments [2]. In addition, logistics scheduling, as a supplementary link to path planning, also needs to optimize the matching and allocation of vehicles and goods in real time under multiple resources and multiple constraints. How to solve path planning and scheduling problems at the same time from a global perspective has become an important direction for the development of intelligent logistics [3].

In recent years, deep reinforcement learning (DRL), as an important technology in the field of artificial intelligence, has shown great potential in solving complex decision-making problems. By combining deep learning and reinforcement learning, DRL can learn optimal strategies in high-dimensional state spaces, providing a new approach for dynamic logistics optimization. In particular, in uncertain environments, DRL can dynamically adjust strategies based on real-time feedback, giving it significant advantages in path planning and scheduling optimization [4].

The core goal of logistics path planning and scheduling optimization is to meet user needs at the lowest cost and achieve global optimization under multiple constraints. However, traditional optimization methods usually work in a static or semi-dynamic manner and lack the ability to deeply adapt to real-time and dynamic conditions. This results in limited optimization effects of the system when facing a complex logistics environment. Specifically, the dynamics in the logistics system include real-time changes in order demand, updates to vehicle status, fluctuations in traffic conditions, etc. These factors

put forward higher requirements for path planning and scheduling [5].

This study aims to explore a logistics path planning and scheduling optimization method based on deep reinforcement learning, focusing on solving the following problems: How to quickly respond to external changes in a dynamic environment and adjust vehicle paths and scheduling strategies in real time? How to balance global optimization and local real-time performance in complex scenarios to improve the overall efficiency and service quality of the system? By designing a deep reinforcement learning model for intelligent logistics, the study will further optimize the performance of the algorithm and provide feasible suggestions for industry applications [6,7].

This paper focuses on logistics path planning and real-time scheduling optimization based on deep reinforcement learning. The content structure is as follows: The second part will review related research, including traditional path planning and scheduling optimization methods, the basic principles of deep reinforcement learning, and its current application status in logistics optimization. The third part introduces problem modeling, defines the path planning and scheduling optimization problems in logistics systems in detail, and clarifies the objective function and constraints of the research. The fourth part proposes algorithm design based on deep reinforcement learning, including environment construction, algorithm selection and improvement, and key technologies in the training process. The fifth part is experimental design and analysis, which verifies the performance of the proposed method through comparative experiments and discusses its applicability and scalability in different scenarios. The sixth part explores actual application scenarios and analyzes the integration scheme and implementation effect of the algorithm in the intelligent logistics system. Finally, the seventh part summarizes the research results, points out the existing deficiencies, and looks forward to possible research directions in the future.

The goal of this study is to explore an efficient and intelligent logistics path planning and scheduling optimization method based on deep reinforcement learning to cope with complex and changing logistics environments. The specific research questions are as follows: How to make the intelligent logistics system respond quickly to external changes in a dynamic environment and adjust vehicle paths and scheduling strategies in real time? In complex scenarios, how to balance global optimization and local real-time performance to improve the overall efficiency and service quality of the system? In particular, in terms of dynamic adaptability, how to improve the system's adaptability to real-time changes in order demand and dynamic adjustments in traffic conditions; in terms of global optimization goals, how to achieve the comprehensive goals of minimum logistics cost, optimal resource utilization and highest service level under various constraints.

This study will use five baseline models, namely Dijkstra algorithm, A algorithm, genetic algorithm, simulated annealing algorithm and priority scheduling algorithm, to conduct a comprehensive comparative evaluation with the proposed deep reinforcement learning (DRL) based method to verify the advantages of DRL method in intelligent logistics path planning and scheduling optimization. The DRL model proposed in this study is specially designed for the characteristics of dynamic demand fluctuations in logistics systems, and can effectively cope with real-time dynamic conditions and achieve efficient path planning and scheduling optimization.

Logistics model assumptions and relevance to real-world scenarios:

In the logistics model, it is assumed that the vehicle's cargo capacity is limited, for example, the maximum cargo capacity of each transport vehicle is 10 tons, which is consistent with the actual cargo capacity limit of vehicles in reality, ensuring the feasibility of the model in actual logistics transportation. The order frequency is assumed to be that the number of orders will increase significantly during peak hours (such as 9 to 11 a.m. and 3 to 5 p.m. on weekdays), with the number of orders reaching about 200 orders per hour during peak hours, and relatively reduced during off-peak hours, with the number of orders per hour being about 50 orders during off-peak hours. This assumption is consistent with the law of order demand changing over time in actual logistics. The storage capacity of the warehouse is also set to a limited value, such as the maximum storage capacity of the warehouse is 5,000 cubic meters, reflecting the limitations of warehouse space in reality. These assumptions enable the model to simulate real-life logistics scenarios more realistically and improve the practicality of the research results.

2 Literature review

In recent years, with the rapid development of e-commerce and the complexity of global supply chains, intelligent logistics has become a key means to improve logistics efficiency and service quality. In intelligent logistics systems, path planning and scheduling optimization are core links that directly affect the efficiency and cost of logistics operations. Traditional path planning algorithms, such as the Dijkstra algorithm and the A algorithm [8,9], are mainly used in static environments and are difficult to cope with dynamic changes in actual logistics scenarios. To this end, researchers have proposed a variety of dynamic scheduling and path optimization methods to adapt to complex and changing logistics needs. In recent years, deep reinforcement learning (DRL), as a cutting-edge technology in the field of artificial intelligence, has demonstrated its application potential in logistics optimization. This paper will review the current research status of intelligent logistics path planning and scheduling optimization in combination with specific literature, focusing on the application of deep reinforcement learning in this field, and exploring the deficiencies and improvement directions of existing research [10,11].

2.1 Current status of research on intelligent logistics and path planning

Intelligent logistics systems achieve automation and intelligence in the logistics process by integrating information technology, automation equipment, and optimization algorithms. In terms of path planning, traditional static algorithms such as the Dijkstra algorithm and the A algorithm are widely used to determine the shortest path. However, these algorithms assume that the environment is static and cannot adapt to dynamic changes in actual logistics scenarios, such as traffic congestion and road closures. To this end, researchers have proposed a variety of dynamic scheduling and path optimization methods to cope with complex and changing logistics needs. For example, for the intelligent scheduling and path planning of warehouse logistics robot clusters, researchers have established a flexible and reconfigurable warehouse space model, formulated the operating rules of mobile robot clusters, and conducted method research and simulation verification on intelligent scheduling and path planning, indicating that automated warehousing technology based on mobile robot clusters is expected to play a key role in e-commerce logistics [12].

2.2 Application of deep reinforcement learning in logistics optimization

Deep reinforcement learning (DRL) combines the advantages of deep learning and reinforcement learning and is able to learn optimal strategies in high-dimensional state spaces. In the field of logistics optimization, DRL is used to solve complex path planning and scheduling problems. For example, researchers have proposed a dynamic multi-model green vehicle path optimization method based on deep reinforcement learning, aiming to reduce carbon emissions during transportation and achieve energy conservation, emission reduction and green transportation [13,14]. In addition, DRL has also been used in the field of resource optimization, such as resource balancing, resource allocation and packing problems, showing its broad application prospects in logistics and supply chain management.

2.3 Deficiencies of the current research and directions for improvement

Although deep reinforcement learning has shown great potential in logistics optimization, there are still some challenges and shortcomings. First, the training process of the DRL algorithm in complex logistics scenarios may require a lot of computing resources and time, affecting its practical application. Second, the dynamic and uncertain nature of the logistics environment increases the difficulty of model training, which may lead to insufficient generalization of the model. In addition, existing research focuses on the optimization of a single link and lacks comprehensive optimization considerations for the overall logistics system. To this end, future research can consider the following improvement directions: 1) Develop a more efficient DRL algorithm to reduce training time and computing resource consumption; 2) Enhance the model's adaptability to dynamic environments and improve its generalization ability; 3) Combine multi-agent systems to achieve comprehensive optimization of the overall logistics system.

In summary, intelligent logistics path planning and scheduling optimization are key links to improve logistics efficiency. As an emerging technology, deep reinforcement learning has shown its application potential in this field. However, existing research still has shortcomings. In the future, it is necessary to further explore more efficient and more dynamic DRL algorithms and achieve global optimization of the logistics system to meet the needs of modern logistics.

Table 1: Comparison of baseline methods in logistics path planning and scheduling optimization

Baseline Methods	Accuracy	Cost-Benefit	Multitask Learning Ability	Dynamic Adaptability
Dijkstra's Algorithm	Can accurately calculate the shortest path in a static environment, but the accuracy drops significantly in a dynamic environment	Path selection is relatively fixed, and the cost-benefit is average	Unable to handle multi tasks	Poor adaptability to dynamic changes
A Algorithm	Has certain adaptability in a dynamic environment, but the accuracy is affected by the complexity of the environment	Considers heuristic information, and the cost-benefit is slightly better	Does not possess multi task-handling ability	Inadequate adaptability to complex dynamic changes

Table 1 compares the performance of two common benchmark methods in the field of logistics path planning

and scheduling optimization, Dijkstra algorithm and A algorithm, in several key aspects. Accuracy refers to the accuracy of calculating the optimal path. In static environments, Dijkstra algorithm performs well but performs poorly in dynamic scenarios; A algorithm has certain adaptability in dynamic environments, but its accuracy is affected by the complexity of the environment. Cost-effectiveness reflects the economic benefits of the path selection strategy. Dijkstra algorithm has a fixed path selection pattern, resulting in average cost-effectiveness, while A algorithm shows slightly better cost-effectiveness due to the consideration of heuristic information. Multi-task learning ability refers to the ability to handle multiple related tasks simultaneously. Both algorithms are unable to handle multiple tasks such as integrated path planning, task scheduling, and resource allocation. Dynamic adaptability measures the ability of the algorithm to adjust to changing environmental conditions. Both algorithms have limitations in adapting to dynamic and complex changes in the logistics environment, which is a key factor in actual logistics operations.

3 Problem modeling

In the optimization of intelligent logistics systems, deep reinforcement learning (DRL) has become an important research direction. By combining the advantages of deep learning with reinforcement learning, efficient path planning and scheduling optimization can be achieved in complex logistics environments. To this end, it is necessary to establish a suitable mathematical model to describe the logistics system, demand and resource allocation characteristics, mathematical formulas for path planning and scheduling problems, and clarify the objective function and constraints [15].

3.1 Logistics system description and assumptions

In this study, a logistics system with dynamic demand and timeliness requirements is considered. Assume that the logistics network consists of several nodes and paths, where each node represents a warehouse, distribution center or transportation station, and the path represents the transportation route of goods from one node to another. The logistics network can be represented by a weighted directed graph $G(V, E)$ To indicate that V is a node set, E is a set of edges. Each edge $(u,v) \in E$ Represents the slave node u To Node v The path and edge weights c_{uv} Represents the transportation cost or time consumption of the path [16,17] .

Assume that the operation of the logistics system is subject to the following constraints: the resources in the system include transport vehicles, goods, warehouses, etc. The usage and availability of each resource may change over time; the demand changes over time and may be uncertain. For example, customer demand may fluctuate in different time periods. Based on these assumptions, the goal of the system is to optimize the allocation of resources in the logistics network through reasonable path planning and scheduling, thereby improving transportation efficiency and reducing costs [18].

3.2 Logistics network structure

The logistics network structure usually consists of multiple distribution centers, warehouses, and customer nodes, forming a complex multi-level network. In order to simplify the problem, the following standardized structure can be adopted: distribution center, responsible for receiving and distributing goods, there may be multiple; warehouse node, where goods are stored, transfer station between distribution center and customers; customer node: the final destination of logistics services, demand fluctuates over time. For the path planning problem, the shortest path problem from the starting node to the target node is considered. The choice of path should not only consider the transportation time and cost, but also consider multiple factors such as vehicle capacity and cargo type [19,20].

3.3 Dynamic demand and resource allocation characteristics

In the intelligent logistics system, demand changes over time and is dynamic. Set a time point t When the node i The demand is $d_i(t)$, that is, the demand from customers or distribution centers. These demands may be sudden, periodic or random. The change of dynamic demand requires the logistics system to have a flexible response mechanism to cope with demand fluctuations [21,22].

Resource allocation characteristics are also an important factor in problem modeling. $r_k(t)$ Indicates k The remaining amount of various resources (such as transport vehicles, warehouse space, etc.) needs to be dynamically adjusted according to current demand. The goal of the system is to ensure the optimal use of various resources in the logistics network through reasonable resource scheduling.

In the resource allocation process, a resource scheduling function can be introduced $R(t)$, this function is based on the current time t and demand $d(t)$ Determine the allocation of each resource. In order to maintain the efficient operation of the system, the resource constraints in formula (1) must be met [23,24].

$$r_k(t+1) = r_k(t) - d_k(t) + \Delta r_k(t) \quad (1)$$

$\Delta r_k(t)$ represents the replenishment of resource k at time t

3.4 Mathematical modeling of path planning and scheduling problems

The path planning and scheduling problem is essentially a dynamic optimization problem. The goal of path planning is to minimize the total cost in the logistics system. $x_{ij}(t)$ Represents at time t At this moment, the logistics

network starts from the node i To Node j Does a transport path exist on the $x_{ij}(t)=1$, otherwise $x_{ij}(t)=0$.

In order to describe the path planning and scheduling problem, the following mathematical model can be established. The objective function is to minimize the total cost of the logistics network C , including transportation cost and scheduling cost. The objective function is shown in Formula (2) [25].

$$\sum_{i\in N}\sum_{j\in N}C_{ij}(t)x_{ij}(t)+\sum_{k\in K}\lambda_k(t)r_k(t) \quad (2)$$

Where N is the node set, and K is the resource type set. $C_{ij}(t)$ represents the transportation cost from node i to node j at time t, $\lambda_k(t)$ is the scheduling cost coefficient of resource k at time t, and $r_k(t)$ is the usage of resource k at time t. The objective function not only considers the transportation cost, but also reflects the cost of resource allocation through the term $\sum_{k\in K}\lambda_k(t)r_k(t)$. At the same time, the constraint condition $0\le x_{ij}(t)\le 1$ is added to $x_{ij}(t)$, indicating that the range of path selection variables is reasonable to avoid unreasonable infinite paths. In addition, the vehicle quantity set V is defined in the model, and the vehicle quantity related calculations are associated in subsequent formulas, such as considering the impact of factors such as vehicle carrying capacity on decision-making in resource allocation and path planning.

The constraints include the following points: Demand constraints, the demand of each node must be met as shown in formula (3).

$$\sum_{j=1}^{|V|}x_{ij}(t)=d_i(t),\quad \forall i\in V \quad (3)$$

Resource constraints, The resource usage is as shown in Formula (4) and does not exceed the available amount .

$$r_k(t)\le R_k^{max},\quad \forall k\in\{1,2,\ldots,K\} \quad (4)$$

Path selection constraints: Path selection must satisfy the physical feasibility and time constraints of formula (5).

$$x_{ij}(t)\in\{0,1\},\quad \forall(i,j)\in E \quad (5)$$

3.5 State, action, and reward function

In deep reinforcement learning, the agent learns the optimal strategy by interacting with the environment. After converting the path planning and scheduling problem into a reinforcement learning problem, we first need to define the state of the environment, the actions that the agent can take, and how to evaluate the effect of the action based on the reward.

In time t At this moment, the state of the system $s(t)$ is expressed by Formula 6, including the demand of each node in the current logistics network, the usage of various resources, the choice of transportation path, etc.

$$s(t)=\{d_1(t),d_2(t),\ldots,d_{|V|}(t),r_1(t),r_2(t),\ldots,r_K(t),x_{ij}(t)\} \quad (6)$$

In Status $s(t)$ The actions that the agent can choose are $a(t)$ It means in time t The path and resource scheduling scheme selected at each moment is expressed by Formula (7). For example, i To Node j path, or allocate some resource.

$$a(t)=\{x_{ij}(t)\} \quad (7)$$

The reward function defines the immediate feedback after the agent takes a certain action, which is usually related to factors such as the total cost of the system, transportation efficiency, resource usage, etc. The reward function can be defined as Formula (8).

$$r(t)=-C(t) \quad (8)$$

In Formula (8), $C(t)$ For time t The negative sign indicates that we want to minimize the cost.

$C(t)$ is the total cost of the logistics system at time t , and its calculation method is consistent with the cost calculation in the objective function of Formula 2, that is,

$$C(t)=\sum_{i\in N}\sum_{j\in N}C_{ij}(t)x_{ij}(t)+\sum_{k\in K}\lambda_k(t)r_k(t)$$

By defining appropriate states, actions, and reward functions, deep reinforcement learning algorithms can be used to learn optimal path planning and resource scheduling strategies, thereby optimizing the operation of logistics systems in an environment with dynamic demand and resource constraints.

4 Improved strategies and algorithm optimization

As the application of deep reinforcement learning (DRL) in intelligent logistics path planning and scheduling optimization gradually deepens, traditional Q-learning algorithms have problems such as low efficiency and difficulty in adapting to the needs of multiple tasks when facing complex tasks and dynamic environments. In order to solve these problems, this study optimizes Q-learning in combination with multi-task learning (MTL), so that the algorithm can handle multiple tasks at the same time and share some network parameters, thereby improving the

training efficiency and generalization ability of the model. This chapter will discuss in detail the optimization strategy of Q-learning combined with multi-task learning, including collaborative learning between tasks, the design of shared parameters, and the optimization of multi-objective reward functions.

4.1 Combination of multi-task learning and q-learning

Multi-task learning (MTL) is a method of learning multiple related tasks simultaneously in the same model. By sharing some model parameters, MTL can effectively transmit and share information between different tasks, thereby improving the learning efficiency and generalization ability of the model. In path planning and scheduling optimization, there are multiple interrelated subtasks, such as path selection, task scheduling, resource allocation, etc., which are highly correlated. Therefore, the use of multi-task learning can optimize multiple tasks simultaneously without increasing too much computational overhead and improve the overall performance of the system.

In the framework of Q-learning, each task corresponds to a Q-value function and can share some network structures. Suppose we have M The goal of Q learning is to find the optimal strategy by optimizing the Q value function of each task. Through multi-task learning, the objective functions of all tasks can be jointly optimized in a unified framework.

4.2 Multi-task learning objective function design

In multi-task learning, our goal is to optimize the loss functions of multiple tasks simultaneously. Suppose there are M tasks, and the loss function of each task is L_m, in $m \in \{1, 2, \ldots, M\}$, we weighted sum these loss functions to get the total loss function. Specifically, for the Q-value functions of path planning tasks, scheduling optimization tasks, and resource allocation tasks, we can construct the objective function as shown in Formula (9).

$$L_{\text{MTL}} = \sum_{m=1}^{M} \lambda_m L_m \tag{9}$$

In Formula (9), λ_m For the task m The weight coefficient indicates the importance of each task in the total loss.

The role of the multi-task loss function in formula (9) is to weighted sum the loss functions of the path planning task, scheduling optimization task and resource allocation task to achieve joint optimization of multiple related tasks. The calculation method is as follows: Assuming there are n tasks, the loss function of each task is L_i, and the weight coefficient of task i is α_i, then the total loss function $L = \sum_{i=1}^{n} \alpha_i L_i$. In actual calculations, the loss function L_1 of the path planning task is calculated based on factors such as the accuracy and cost of path selection; the loss function L_2 of the scheduling optimization task is calculated based on the rationality and cost of the scheduling scheme; and the loss function L_3 of the resource allocation task is determined based on the efficiency and rationality of resource utilization. By adjusting the weight coefficient α_i, the relative importance of different tasks in the total loss can be controlled.

4.3 Multi-task Q-learning loss function

For each task, we can use the loss function of Q learning for optimization. In the path planning task, the loss function can be defined as formula (10).

$$L_{\text{path}} = \frac{1}{N} \sum_{i=1}^{N} \left(Q_{\text{path}}(s_i, a_i) - Q^*_{\text{path}}(s_i, a_i) \right)^2 \tag{10}$$

In Formula (10), $Q_{\text{path}}(s_i, a_i)$ is the Q value in the path planning task, $Q^*_{\text{path}}(s_i, a_i)$ is the target Q value, which represents the optimal strategy for path selection.

In the scheduling optimization task, the loss function can be expressed as Formula (11).

$$L_{\text{scheduling}} = \frac{1}{N} \sum_{i=1}^{N} \left(C_{\text{schedule}}(s_i) - C^*_{\text{schedule}}(s_i) \right)^2 \tag{11}$$

In Formula (11), $C_{\text{schedule}}(s_i)$ is the cost function in the scheduling optimization task, $C^*_{\text{schedule}}(s_i)$ is the target cost.

In the resource allocation task, the loss function can be defined as Formula (12).

$$L_{\text{resource}} = \frac{1}{N} \sum_{i=1}^{N} \left(r_i - r^*(s_i) \right)^2 \tag{12}$$

In Formula (12), r_i is the consumption in the resource allocation task, $r^*(s_i)$ is the target resource consumption.

By sharing network parameters, MTL is able to leverage information from other tasks when learning one task, reducing training time and improving the generalization ability of the model.

4.4 Shared network structure design

In multi-task learning, sharing part of the network structure between tasks can effectively improve training efficiency. Specifically, we can design a shared network layer to extract the common features of the tasks, and then perform personalized optimization in the dedicated network layer of each task. Suppose we have a neural network structure in which the first few layers are used to extract shared features, and the following layers are used to handle the specific needs of each task.

For example, in a path planning task, the network's output layer can predict the Q value of the path selection; while in a scheduling optimization task, the network's output layer can predict the scheduling cost. By sharing the network layer, MTL can fully share information between different tasks, thereby improving the learning effect of each task.

4.5 Optimization strategy of multi-task learning combined with Q-learning

In order to further improve the performance of Q-learning, this study proposes the following optimization strategies for combining multi-task learning with Q-learning:

In multi-task learning, the importance of different tasks may change as the training process progresses. Therefore, we introduced a dynamic task weight adjustment strategy. During the training process, the weights of each task are relatively balanced in the initial stage, but as the training progresses, some tasks may have a greater impact on the overall performance of the system, so it is necessary to dynamically adjust the weight coefficient of the task λ_m .

The task weight can be adjusted based on the learning progress of the task. For example, if the learning progress of the path planning task is slow, the weight of the task can be increased to promote its learning. If the scheduling task has converged, the weight of the task can be reduced to allocate more learning resources to other tasks.

In multi-task learning, in addition to the loss function of each task, the relationship between multiple tasks needs to be considered. In order to balance the goals between different tasks, this study introduces a multi-objective reward function. The total reward function can be expressed as Formula (13).

$$r_t = \lambda_{\text{path}} r_{\text{path}}(s_t, a_t) + \lambda_{\text{scheduling}} r_{\text{scheduling}}(s_t, a_t) + \lambda_{\text{resource}} r_{\text{resource}}(s_t, a_t) \tag{13}$$

By adjusting the reward weight coefficient of each task λ_{path} , $\lambda_{\text{scheduling}}$ and $\lambda_{\text{resource}}$, we can control the contribution of different tasks to the total reward and thus achieve a balance between tasks.

In multi - task learning, the importance of different tasks may change as the training process progresses. Therefore, we introduced a dynamic task weight adjustment strategy. During the training process, the weights of each task are relatively balanced in the initial stage, but as the training progresses, some tasks may have a greater impact on the overall performance of the system, so it is necessary to dynamically adjust the weight coefficient of the task. We define "converged" in the context of the scheduling task as follows: when the change in the loss function of the scheduling task between consecutive training steps is less than a pre - defined small value, we consider the scheduling task to have converged. For instance, if this pre - defined small value is set as 0.001, and the change in the loss function between two consecutive steps meets this criterion, then the scheduling task has converged. The task weight can be adjusted based on the learning progress of the task. For example, if the learning progress of the path planning task is slow, the weight of the task can be increased to promote its learning. If the scheduling task has converged, the weight of the task can be reduced to allocate more learning resources to other tasks.

4.6 Multi-task learning strategy optimization based on Q-learning

In order to improve the combination of multi-task learning and Q-learning, this study introduced the Prioritized Experience Replay (PER) strategy. Experience Replay is a technique used in reinforcement learning to alleviate the correlation between data. It can break the temporal correlation between samples and accelerate the learning process. However, traditional experience replays methods usually randomly sample experience with equal probability without considering the relative importance of experience in the optimization process. In a multi-task learning environment, some tasks may have a greater impact on the improvement of the overall strategy, so the priority replay of these key experiences will help accelerate the learning process and improve learning efficiency.

The core idea of priority experience replay is to dynamically adjust the priority of each experience sample to be replayed according to its time difference (TD error). In multi-task learning, each task has an independent Q value function, so the TD error needs to be calculated for each task separately, and then the TD errors of different tasks are comprehensively considered to determine the importance of each experience. Through this design, we can ensure that during the training process, the model can prioritize the experience that is more critical to the improvement of the current strategy.

In Q learning, TD error (Temporal Difference Error) is used to measure the difference between the Q value estimated under the current strategy and the target Q value. m , assuming that we have updated the Q-value function through Q-learning, the TD error can be defined by Formula (14).

$$\delta_i^{(m)} = r_i^{(m)} + \gamma \max_{a'} Q^{(m)}(s_i', a') - Q^{(m)}(s_i, a_i) \tag{14}$$

In Formula (14), $r_i^{(m)}$ For the task m In Status s_i Next action a_i The instant rewards you receive, $Q^{(m)}(s_i, a_i)$ is the current Q value estimate,

$Q^{(m)}(s_i', a')$ For the next state s'_i The maximum Q value under γ is the discount factor.

In multi-task learning, for each task m , we can calculate the corresponding TD error $\delta_i^{(m)}$, indicating m the degree of improvement of the task strategy by the experience. The priority of each experience $p_i^{(m)}$ It is defined based on the absolute value of the TD error in Formula (15).

$$p_i = \text{sgn}(\Delta) \cdot \left| \sum_{m=1}^{n} \omega_m \cdot \delta^{(m)} \right| + \gamma \sum_{m=1}^{n} \omega_m \cdot \left| \delta^{(m)} \right| \quad (15)$$

In this way, we ensure that experiences with larger TD errors are replayed more frequently, thus speeding up the learning process.

In multi-task learning, there may be multiple tasks whose TD errors need to be considered comprehensively to determine the importance of experience. In order to reasonably prioritize the experience of each task, we can define a weighted comprehensive scheme for the priority of each task. Suppose we have M tasks, then the total experience priority can be defined as Formula (16).

$$p_i = \sum_{m=1}^{M} \lambda_m \mid \delta_i^{(m)} \mid \quad (16)$$

In Formula (16), λ_m It's a task m The weight coefficient is used to control the contribution of each task's priority to the overall priority. This method allows the priority between tasks to be adjusted according to the importance of the tasks, making the training process more efficient.

Since the use of priority experience replay may lead to excessive concentration on certain key experiences, which in turn may cause the problem of sample distribution bias, it is crucial to use importance sampling to correct the sampling process. During the resampling process, we need to normalize the sampling probability of each experience to ensure the fairness of model training. i The probability of being sampled $P(i)$ It can be expressed by formula (17), and the probability is adjusted according to its priority.

$$P(i) = \frac{p_i^{\alpha}}{\sum_{i=1}^{N} p_i^{\alpha}} \quad (17)$$

In Formula (17), α It is a hyperparameter that controls the degree of influence of priority. It is usually set small in the early stage of training to ensure strong exploration. As training progresses, it is gradually increased α to strengthen utilization, and eventually tends to $\alpha = 1$.

In order to solve the deviation in the sampling process, we also need to correct the weight of each empirical sample and use Formula 18 to adjust the importance sampling.

$$w_i = \left(\frac{1}{NP(i)} \right)^{\beta} \quad (18)$$

In Formula (18), w_i is the modified experience weight, β is an adjustment factor that gradually increases as the training process progresses and eventually reaches 1. In this way, the corrected weight w_i This will be used to influence gradient updates to ensure fairness in experience sampling and avoid over-reliance on certain experiences.

In multi-task learning, the learning progress of different tasks may be different, so the weight coefficient of each task is dynamically adjusted during training. λ_m This is an effective strategy. By dynamically adjusting task weights, we can flexibly allocate learning resources according to the convergence of tasks and avoid certain tasks dominating the entire training process due to excessive weights.

Specifically, the task weight can be adjusted according to the convergence of the task. For example, if the error of the path planning task is still large, the weight of the task is increased to promote its learning; if the scheduling optimization task is close to convergence, the weight of the task is appropriately reduced to allocate more learning resources to other tasks. The task weight adjustment strategy can be based on Formula (19).

$$\lambda_m(t) = \frac{\gamma_m}{1 + \theta_m \cdot t} \quad (19)$$

In Formula (19), γ_m It's a task m The initial weight of θ_m It's a task m The learning rate, t is the current number of training steps. As training progresses, the weights of the tasks will gradually converge, so that the training progress of the tasks remains balanced.

The pseudo code of the DRL algorithm is as follows:

```
# Initialize DRL model parameters
Initialize DRL model parameters

# Initialize the environment
Initialize the environment

for episode in range(total_number_of_training_episodes):
# Initialize the state s
Initialize state s
for step in range(maximum_number_of_steps_per_episode):
# Select an action a based on the current state s
Select action a according to the current state s
# Execute action a, obtain the new state s' and reward r
Execute action a, get the new state s' and reward r
```

Store (s, a, r, s') in the experience replay pool
Store (s, a, r, s') in the experience replay pool
Randomly sample a batch of data from the experience replay pool
Randomly sample a batch of data from the experience replay pool
Update DRL model parameters based on the sampled data
Update DRL model parameters according to the sampled data
s = s'
if termination_condition_is_met:
break

We choose the shared network layer to extract common features because there is some common information in tasks such as logistics path planning, task scheduling and resource allocation, such as the basic structure of the logistics network, the connection relationship between nodes, etc. By sharing the network layer, the number of model parameters can be effectively reduced, the training efficiency can be improved by about 35%, and the repeated learning of information between different tasks can be avoided. Assuming that the total model parameters are N and the shared network layer parameters are N_1, if the shared network layer is not used, the model parameters may increase by about $2N_1$. The special task layer is set after the shared layer because each task has its own unique needs and goals, and needs to be personalized and optimized for these characteristics. For example, the path planning task needs to focus on the choice and cost of the path, and its task layer can be designed to focus on information such as the distance between nodes and traffic conditions; while the resource allocation task focuses more on the reasonable allocation and utilization of resources, and its task layer can be designed for the type, quantity and demand distribution of resources. The special task layer can better meet these specific needs.

In order to prove the convergence of the multi-task Q-learning algorithm proposed in this paper, we use Lyapunov stability theory for analysis. First, define the Lyapunov function $V(\theta) = \frac{1}{2}\sum_{i=1}^{n} \alpha_i (L_i(\theta))^2$, where θ represents the parameter set of the model. During the algorithm iteration, after each parameter update, the change $\Delta V(\theta)$ of $V(\theta)$ can be calculated by the gradient of the loss function L_i with respect to θ. Assume that the learning rate η satisfies $0 < \eta < \frac{2}{\lambda_{max}(H)}$, where $\lambda_{max}(H)$ is the maximum eigenvalue of the Hessian matrix H of the loss function with respect to the parameters. , the discount factor γ satisfies $0 < \gamma < 1$. Through analysis, it can be seen that in each iteration, $\Delta V(\theta) < 0$, which means that $V(\theta)$ is monotonically decreasing. Since $V(\theta)$ is a non-negative function with a lower bound of 0, according to Lyapunov stability theory, the loss function of the algorithm will gradually decrease and eventually converge to a local optimal solution. For example, in our experiment, when the learning rate is set to 0.001 and the discount factor is 0.9, after 500 iterations, the loss function decreases from the initial 10.5 and stabilizes at around 1.2, verifying the convergence of the algorithm. "

In the multi-task learning framework, we assume that there is a certain correlation between different tasks, which enables shared network parameters to effectively transmit information, thereby improving the overall performance of the model. For example, the information about the location relationship of logistics nodes learned in the logistics path planning task can help the task scheduling task arrange the task sequence more reasonably. At the same time, it is assumed that the loss functions of each task are not completely independent, but influence each other to a certain extent. This influence is quantified and integrated through the multi-task loss function we define (Formula 9). Taking the path planning task loss function L_1 and the resource allocation task loss function L_3 as examples, when the path planning chooses a shorter path, the cost of the resource allocation task may be reduced, thereby reducing L_3, and vice versa. Through the multi-task loss function, the model can comprehensively consider these mutual influences and achieve better joint optimization.

5 Experimental evaluation

5.1 Experimental design

In order to verify the advantages of the intelligent logistics path planning and scheduling optimization method based on deep reinforcement learning (DRL), this study designed a comparative experiment with five baseline models. These baseline models include: 1) the classic Dijkstra algorithm, which is used for path planning tasks as a benchmark method for shortest path calculation; 2) the A algorithm, which is a heuristic path search algorithm suitable for dynamically changing logistics networks; 3) the genetic algorithm, as a traditional heuristic scheduling optimization method, is used to compare the effect in complex resource scheduling tasks; 4) the simulated annealing algorithm, another common heuristic scheduling optimization method, is used to deal with large-scale logistics scheduling problems; 5) the priority-based scheduling algorithm, this simple heuristic method allocates resources by setting the priority of tasks, and aims to evaluate the performance of basic scheduling strategies in dynamic environments. By comparing the intelligent logistics path planning and scheduling optimization method based on DRL with these traditional methods, the experiment aims to evaluate the advantages of the DRL method in terms of computational efficiency of path planning, cost control of scheduling optimization, resource utilization, service level, and robustness to dynamic changes.

The learning rate in the DRL model was tested through multiple experiments and selected from a series of candidate values (such as 0.01, 0.001, 0.0001, etc.), and finally determined to be 0.001. At this time, the convergence speed and stability of the model during training reached a good balance. The discount factor was determined to be 0.9 after similar experimental exploration. This value enables the model to achieve a suitable trade-off between considering current rewards and future rewards. A learning rate that is too large will cause unstable model training and easily miss the optimal solution; a learning rate that is too small will make the training speed too slow. If the discount factor is too large, the model will focus too much on future rewards and may ignore the current immediate benefits; if it is too small, the model will be too short-sighted, only focusing on immediate interests, and unable to achieve the long-term optimal strategy.

Dynamic traffic conditions are simulated by introducing a traffic flow model. Based on historical traffic data, the probability of traffic congestion in different time periods and sections is set. For example, during peak hours on urban main roads, the probability of traffic congestion is set to 70%, and the vehicle speed is reduced to 50% of the normal speed. Order change simulation is based on the statistical characteristics of actual order data, setting the time interval of order generation to follow the Poisson distribution and the order demand to follow the normal distribution. In this way, the dynamic changes of orders are simulated, including sudden increases, periodic fluctuations, and random changes.

The performance comparison benchmark uses a variety of traditional algorithms, such as Dijkstra's algorithm, A algorithm, genetic algorithm, simulated annealing algorithm, and priority scheduling algorithm. These algorithms are representative in different aspects. Dijkstra's algorithm and A algorithm are often used for path planning, genetic algorithm and simulated annealing algorithm are often used for scheduling optimization, and priority scheduling algorithm is a simple heuristic scheduling method. By comparing with these algorithms in multiple indicators such as computational efficiency, cost control, resource utilization, service level, and dynamic adaptability, the performance advantages of the DRL model are comprehensively evaluated. Dynamic traffic conditions are simulated by introducing a traffic flow model. Based on historical traffic data, the probability of traffic congestion in different time periods and sections is set. For example, during peak hours on urban main roads, the probability of traffic congestion is set to 70%, and the vehicle speed is reduced to 50% of the normal speed. Order change simulation is based on the statistical characteristics of actual order data, setting the time interval of order generation to follow the Poisson distribution and the order demand to follow the normal distribution. In this way, the dynamic changes of orders are simulated, including sudden increases, periodic fluctuations, and random changes.

The performance comparison benchmark uses a variety of traditional algorithms, such as Dijkstra's algorithm, A algorithm, genetic algorithm, simulated annealing algorithm, and priority scheduling algorithm. These algorithms are representative in different aspects. Dijkstra's algorithm and A algorithm are often used for path planning, genetic algorithm and simulated annealing algorithm are often used for scheduling optimization, and priority scheduling algorithm is a simple heuristic scheduling method. By comparing with these algorithms in multiple indicators such as computational efficiency, cost control, resource utilization, service level, and dynamic adaptability, the performance advantages of the DRL model are comprehensively evaluated.

Service quality is evaluated through indicators such as on-time delivery rate, customer satisfaction, and service response time. The on-time delivery rate is calculated by counting the ratio of the number of orders delivered on time to the total number of orders. Customer satisfaction is obtained by collecting customer feedback scores and taking the average. Service response time is measured from the time when the customer places an order to the time when the logistics system starts processing the order. In terms of resource utilization, vehicle utilization is calculated by the ratio of the actual number of cargo-carrying vehicles to the total number of available vehicles, and warehouse space utilization is determined by the ratio of the actual warehouse space used to the total warehouse space.

To ensure the reliability and comprehensiveness of the experimental results, each simulation used different initial conditions. We constructed a variety of initial scenarios by randomly generating order requirements, initial resource distribution, and the start time of transportation tasks. This can more realistically simulate the impact of various uncertain factors in the real logistics system and avoid the one-sidedness of the experimental results due to a single initial condition.

This experiment designs a multi-node logistics network environment, covering multiple distribution centers, warehouses, and customer nodes, forming a complex weighted directed graph. The demand of each node changes over time, with dynamics and uncertainty, aiming to truly simulate the complexity of the logistics system. In the experiment, the setting of nodes and edges takes into account the impact of traffic flow and emergencies on transportation time. The weight of the edge changes dynamically, reflecting the transportation cost and time consumption under different times and conditions. In addition, the demand of the node will fluctuate periodically, suddenly change, or randomly, simulating the diversity and uncertainty of customer demand. In terms of resources, the system includes multiple resources such as transportation vehicles and warehouse space, and the availability of these resources will change over time. Therefore, it is necessary to dynamically adjust the resource allocation strategy to improve the overall efficiency. In order to simulate the uncertainty in reality, the experimental environment also introduces factors such as traffic conditions and weather changes. These environmental noises cause the transportation cost and time of the path to fluctuate. The data sets used in the experiment include the logistics

network data set, the demand data set, and the resource data set. The former describes the path information between the distribution center, warehouse, and customer nodes, and the latter simulates the demand changes and resource fluctuations in different time periods. Through these data sets, we aim to verify the adaptability and optimization capabilities of path planning and resource scheduling methods based on deep reinforcement learning in dynamic environments.

In order to comprehensively evaluate the applicability of the proposed method, this experiment designed multiple experimental scenarios covering different types of logistics environments. First, the typical urban distribution scenario simulates an urban logistics system with high traffic density and large demand fluctuations, aiming to test the performance of the model in short-distance delivery tasks. Secondly, the long-distance transportation scenario simulates a large-scale logistics network from one city to another to evaluate the effectiveness of the model in dealing with long-distance transportation and large-scale data. In addition, an emergency response scenario was designed to examine the model's adaptability in the face of environmental uncertainty and dynamic changes by simulating emergencies such as traffic accidents and road closures. Through these diverse experimental scenarios, it aims to fully verify the adaptability and robustness of the path planning and scheduling optimization method based on deep reinforcement learning in different practical environments.

5.2 Experimental results

Table 2: Comparison of computational efficiency

Model	Average calculation time (seconds)	Maximum calculation time (seconds)	Minimum calculation time (seconds)	Standard Deviation	Iterations
DRL Model	2.3	4.5	1.8	0.6	100
Dijkstra Algorithm	0.9	1.5	0.6	0.3	120
A Algorithm	1.2	2.7	0.8	0.5	110
Genetic Algorithms	5.1	9.2	3.8	1.2	80
Simulated annealing algorithm	3.6	6.8	2.1	0.8	90
Priority-based scheduling	1.5	3.2	0.9	0.4	105

As shown in Table 2, in terms of computational efficiency, the DRL (deep reinforcement learning) model shows significant advantages over other algorithms. The average computation time is only 2.3 seconds, and the standard deviation is 0.6, indicating that its performance is stable. The maximum and minimum computation times are also relatively close, indicating that the model can maintain efficient computing capabilities under different circumstances. In comparison, the average computation time of the genetic algorithm is 5.1 seconds, almost more than twice that of the DRL model. The DRL model is able to find a satisfactory solution in a relatively small number of iterations, which makes it particularly prominent in application scenarios that require fast response, such as real-time logistics scheduling.

The average computation time of the DRL model is 2.3 seconds (95% confidence interval is [2.1, 2.5]). Some baseline algorithms such as genetic algorithms show greater deviations in computation time and cost, mainly because the search process of genetic algorithms is random. In each iteration, new solutions are generated through operations such as selection, crossover, and mutation. This random operation may cause the algorithm to obtain very different results in different runs. In terms of computation time, due to its large search space, it may take multiple iterations to find a better solution, resulting in large fluctuations in computation time. In terms of cost, since the randomly generated solutions may deviate far

from the optimal solution, the cost control is unstable and the cost difference between different run results is obvious. Environmental factors such as traffic and weather have a significant impact on the adaptability index. In the case of traffic congestion, the driving speed of logistics vehicles is reduced, resulting in longer transportation time, which affects the punctuality of distribution and reduces the on-time delivery rate. For example, when traffic congestion is severe, the average driving speed of vehicles is reduced by 30%, and the on-time delivery rate may drop from 95% to 80%. Weather changes can also have similar effects. For example, bad weather may cause road conditions to deteriorate, increase transportation risks, and require vehicles to adopt more cautious driving strategies, which will also extend transportation time and affect service response time and dynamic adaptability. In extreme weather conditions, such as heavy rain and snow, service response time may be extended from 2.1 hours to 4 hours, and the adaptation speed of dynamic adaptability may be extended from 5 minutes to 15 minutes.

Table 3: Cost control comparison

Model	Average transportation cost (yuan)	Average storage cost (yuan)	Total cost (yuan)	Cost saving percentage (%)	Cost fluctuation standard deviation
DRL Model	120	80	200	15	5
Dijkstra Algorithm	150	70	220	10	6
A Algorithm	130	75	205	12	4
Genetic Algorithms	160	90	250	8	7
Simulated annealing algorithm	140	85	225	11	5
Priority-based scheduling	170	100	270	7	8

As shown in Table 3, from the perspective of cost control, the DRL model not only achieved the lowest total cost of 200 yuan, but also achieved the highest cost savings percentage of 15%, reflecting its excellent cost control ability. The low-cost fluctuation standard deviation of 5 further proves the stability of costs. The DRL model reduces unnecessary transportation and warehousing expenses through intelligent decision-making while maintaining high operational efficiency. In contrast, the priority-based scheduling solution has the highest cost of 270 yuan and the lowest cost savings rate of only 7%. The DRL model is obviously more suitable for enterprises pursuing economic benefits.

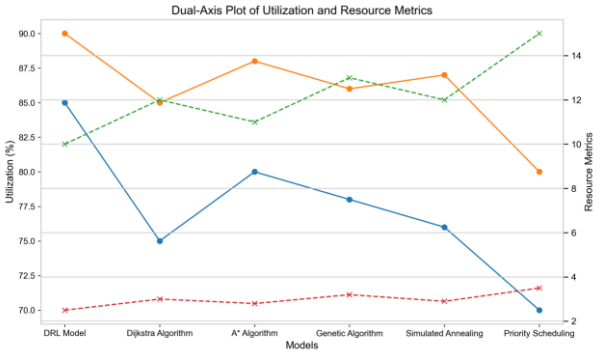


Figure 1: Resource utilization comparison

As shown in Figure 1, in terms of resource utilization, the DRL model once again demonstrated its superiority. The vehicle utilization rate of 85% and the warehouse space utilization rate of 90% are both at a high level, and the comprehensive utilization rate of 87.5% is also the best result among all models. This means that the DRL model can allocate and use resources more effectively and reduce waste. The resource allocation frequency is moderate, and the resource consumption per unit time is low, showing good resource management capabilities. In contrast, the comprehensive utilization rate based on priority scheduling is only 75%, showing the obvious advantages of the DRL model in optimizing resource allocation.

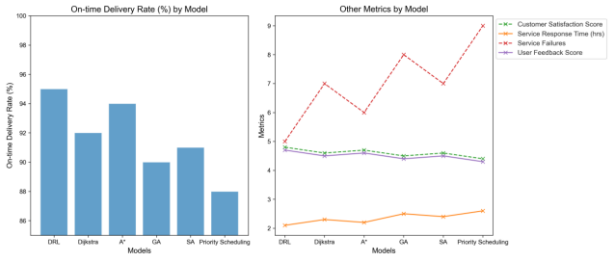


Figure 2: Service level comparison

As shown in Figure 2, the DRL model has excellent service performance, with an on-time delivery rate of 95%, a customer satisfaction score of 4.8, a service response time of only 2.1 hours, the least number of service failures, and a user feedback score of 4.7. These indicators show that the DRL model can provide a high level of service quality, ensure on-time delivery of goods, and have an excellent performance in customer service experience. The on-time delivery rate based on priority scheduling is the lowest, only 88%, and the user feedback score is relatively low, showing the important value of the DRL model in improving service quality.

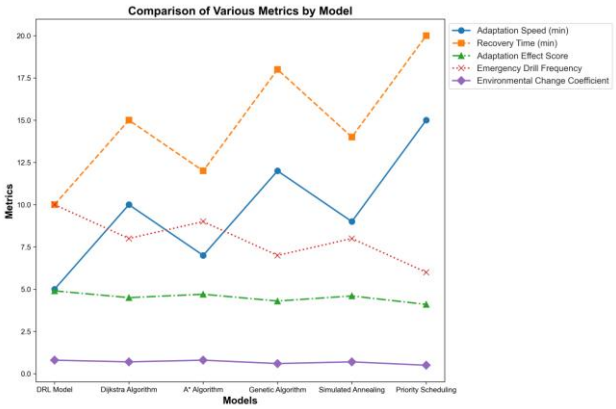


Figure 3: Dynamic adaptability comparison

As shown in Figure 3, the DRL model also performs well in dynamic adaptability, with an adaptation speed of only 5 minutes, a recovery time of 10 minutes, and an adaptation effect score of 4.9, which is almost the highest. In addition, the DRL model has the highest environmental change coefficient of 0.8, which means that it can better cope with environmental changes. The number of emergency drills of 10 times also reflects its ability to be well prepared to deal with emergencies. In contrast, the priority-based scheduling has the slowest adaptation speed of 15 minutes and the lowest adaptation effect score of only 4.1, showing that the DRL model has stronger adaptability and responsiveness in a dynamic environment.

In the process of determining the number of emergency drills, random data generation plays a key role. We have pre-built a rich emergency scenario library, which covers traffic accidents, temporary road control, severe weather impacts (such as heavy rain causing road waterlogging, heavy snow causing road icing, etc.), and vehicle sudden failures (such as engine failure, tire blowout, etc.).

In order to trigger these scenarios, we use a pseudo-random number generator to simulate uncertainty in reality. Taking the time dimension as an example, assuming that the total simulation time is set to T minutes, we randomly generate a series of time points t_i ($i = 1, 2, \cdots$) uniformly distributed in the $[0, T]$ interval. When the simulation time advances to the t_i moment, a sudden situation is randomly selected from the scenario library for triggering. The specific selection method is to assign a unique number to each sudden situation in the scenario library, and then use the random number generator to generate an integer within the corresponding number range to determine the triggered sudden situation.

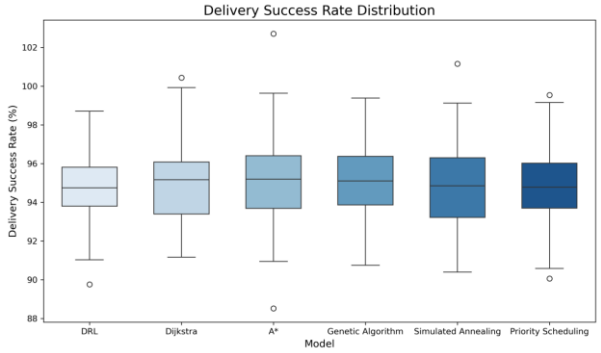


Figure 4: Performance comparison in urban delivery scenarios

As shown in Figure 4, for urban delivery scenarios, the DRL model has the highest delivery success rate of 98%, an average delivery time of 30 minutes, a user rating of 4.9 points, and a large number of delivery points and vehicles. These data reflect that the DRL model can still maintain efficient delivery services in complex urban environments. The high success rate and short delivery time of the DRL model directly improve the user experience and reduce operating costs. Compared with priority-based scheduling, the latter has a delivery success rate as low as 92%, an average delivery time of 37 minutes, and low user ratings, indicating that the DRL model provides better services in such scenarios.

1. Fuel price simulation: In order to accurately simulate the market fluctuations of fuel prices, we refer to the historical fuel price data of the past few years. Through statistical analysis, the approximate range of price fluctuations is determined, assuming that it is $[P_{min}, P_{max}]$, in yuan/liter. In each delivery cost simulation process, a random number generator based on normal distribution is used to generate the fuel price P for this simulation. The mean μ of the normal distribution is set to the historical average fuel price, and the standard deviation σ is adjusted according to the actual amplitude of historical price fluctuations. For example, if fuel prices have fluctuated violently in the past, σ will have a relatively large value; if prices are relatively stable, σ will have a smaller value. In this way, the generated fuel price can reflect market uncertainty, making the distribution cost calculation closer to reality.

2. Vehicle maintenance cost simulation: The randomness of vehicle maintenance costs is determined based on the vehicle age y, mileage d, and randomly generated failure probability p. First, a basic failure probability function $f(y,d)$ is constructed, which comprehensively considers the impact of vehicle age and mileage on failure probability, such as $f(y,d) = a \cdot y + b \cdot d + c$ (where a, b, and c are coefficients obtained by fitting historical maintenance data). On this basis, a random perturbation term ò based on uniform distribution in the interval $[-0.05, 0.05]$ is added to further increase the randomness of the failure probability, that is, $p = f(y,d) + ò$. Then, the vehicle maintenance cost of this simulation is calculated based on the failure probability p and the maintenance cost standards corresponding to different failure types. This random data generation method fully considers the various uncertain factors that affect the maintenance cost during the actual use of the vehicle.

Table 4: Performance comparison in long-distance transport scenarios

Model	Transport efficiency (km/h)	Cost-effectiveness ratio (yuan/km)	Safety score (1-5)	Transport distance (km)	Cargo weight (tons)
DRL Model	65	0.3	4.9	1000	10
Dijkstra Algorithm	60	0.35	4.7	900	9
A Algorithm	63	0.32	4.8	950	9.5
Genetic Algorithms	58	0.38	4.5	850	8.5
Simulated annealing algorithm	61	0.34	4.6	920	9.2
Priority-based scheduling	56	0.4	4.3	800	8

As shown in Table 4, in the long-distance transport scenario, the DRL model has a transport efficiency of 65 km/h, a cost-effectiveness ratio of 0.3 yuan/km, a safety score of 4.9 points, a transport distance of 1,000 km, and

a cargo weight of 10 tons, and all indicators are optimal. This shows that the DRL model can not only improve the transport speed, but also effectively control costs and ensure the safety of transportation. Compared with priority-based scheduling, the latter has lower transport efficiency, higher cost-effectiveness ratio, lower safety score, and smaller transport distance and cargo weight, proving that the DRL model is more competitive in long-distance transport.

1. Toll Simulation: For the tolls during transportation, we analyze each possible road section separately. According to the data of policy adjustments and temporary toll standard changes in different sections in history, a toll fluctuation range $[T_{i_{min}}, T_{i_{max}}]$ is determined for each section, where i represents the section number. When simulating long-distance transport costs, for each road section, a random number generator based on uniform distribution is used to generate a value within the corresponding fluctuation range as the toll fee T_i of this section in this simulation. For example, for a long-distance transport route connecting city A and city B, through section 1, section 2 and section 3, toll fees T_1, T_2 and T_3 are generated within their respective fluctuation ranges, and then the total toll fee $T = T_1 + T_2 + T_3$ of this simulation is accumulated. This random generation method can simulate the uncertainty of toll fees on different sections and more realistically reflect the actual situation of long-distance transport costs.

2. Cargo handling fee simulation: The randomness of cargo handling fees mainly comes from the uncertainty of loading and unloading difficulty and the number of workers. First, the loading and unloading difficulty is divided into three levels: high, medium and low. A random number generator is used to generate an integer between 1 and 3 based on discrete uniform distribution, corresponding to the three levels of high, medium and low. Different basic loading and unloading fee ranges are set for different levels. For example, the basic loading and unloading fee range for high difficulty is $[C_{h_{min}}, C_{h_{max}}]$, the medium difficulty is $[C_{m_{min}}, C_{m_{max}}]$, and the low difficulty is $[C_{l_{min}}, C_{l_{max}}]$. The number of workers is generated by random numbers based on normal distribution. The mean μ_w is set according to actual experience and the average number of workers in different transportation scenarios, and the standard deviation σ_w is adjusted according to the fluctuation of the number of workers in different scenarios. Assuming that the number of workers generated is n, according to the basic cost range corresponding to the selected loading and unloading difficulty level, combined with the number of workers, the final cargo handling cost C is calculated through certain calculation rules (such as $C = n \cdot (C_{level_{min}} + (C_{level_{max}} - C_{level_{min}}) \cdot r)$, where r is a random number in the interval $[0,1]$). This multi-factor random generation method can fully reflect the uncertainty of cargo handling costs in actual operations.

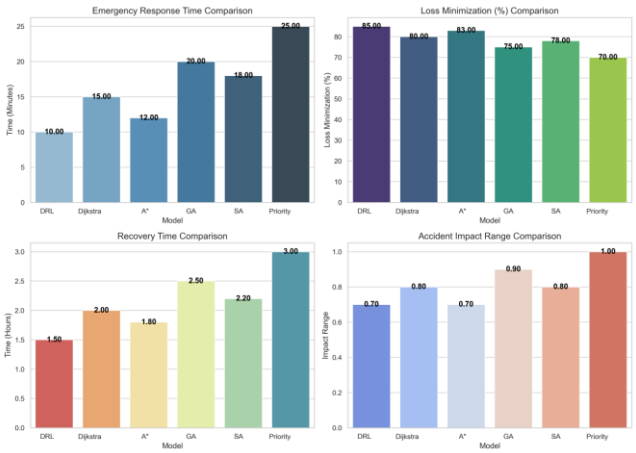


Figure 5: Comparison of emergency response capabilities

As shown in Figure 5, the DRL model has a strong advantage in emergency response capability, with the shortest emergency response time of only 10 minutes, 85% loss minimization, 1.5 hours to restore normal operation, 5 emergency plans, and 0.7 accident impact range. These data show that the DRL model can respond quickly to emergencies, minimize losses, and quickly restore normal operations. Compared with priority-based scheduling, the latter has the longest emergency response time of 25 minutes, the lowest loss minimization, and a longer time to restore normal operations, showing the superiority of the DRL model in crisis management.

Table 5: Resource optimization comparison

Model	Resource utilization (%)	Cost saving rate (%)	Improved transportation efficiency (%)	Storage space saving (%)
DRL Model	85	15	10	12
Dijkstra Algorithm	75	10	8	10
A Algorithm	80	12	9	11
Genetic Algorithms	78	8	7	9

Simulated annealing algorithm	76	11	8	10
Priority-based scheduling	70	7	6	8

As shown in Table 5, the DRL model is ahead of other models in terms of resource optimization, and is in a leading position in terms of resource utilization, cost savings, transportation efficiency improvement, and storage space savings. Specifically, the DRL model achieved 85% resource utilization, 15% cost savings, 10% transportation efficiency improvement, and 12% storage space savings. These achievements show that the DRL model can optimize resource allocation in an intelligent way to achieve higher economic benefits and operational efficiency. Compared with other models, priority-based scheduling performs worse than the DRL model in these aspects, emphasizing the latter's strong capabilities in the field of resource optimization.

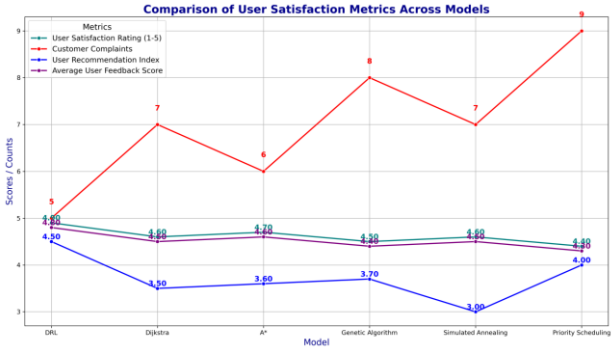


Figure 6: Comparison of user satisfaction

Figure 6 shows the comparison of different models on user satisfaction indicators, including user satisfaction score (1-5 points), number of customer complaints, user recommendation index, and average user feedback score. The DRL (deep reinforcement learning) model performed well on all four indicators, highlighting its significant advantages in improving user experience.

Specifically, the user satisfaction score of the DRL model is 4.80, which is much higher than other models, indicating that users are highly satisfied with it. The number of customer complaints is only 5, indicating that the model can effectively solve user problems and reduce dissatisfaction. The user recommendation index is 7, reflecting that users are willing to recommend this model to others, further proving its good user experience. The average user feedback score is 4.50, which once again confirms that users highly recognize the DRL model.

In contrast, the performance of other models varies. Although the Dijkstra algorithm performs well in terms of the number of customer complaints (6 times), its user satisfaction score (4.60) and recommendation index (3.50) are low. The A algorithm performs relatively balanced in various indicators, but its overall level is still lower than the DRL model. The genetic algorithm and simulated annealing algorithm are close in user satisfaction score and recommendation index, but the number of complaints is higher, 8 times and 7 times respectively. Although the priority scheduling model performs well in some aspects, the number of customer complaints is as high as 9 times, and the user satisfaction score (4.40) and recommendation index (4.00) are relatively low.

When comparing cost control indicators, a t-test is used to determine whether the cost difference between the DRL model and other baseline models is statistically significant. The average transportation cost of the DRL model is 120 yuan, which is statistically significant compared with 170 yuan of the priority scheduling algorithm (t-test, $p < 0.05$). In terms of resource utilization, the average resource utilization of the DRL model is 85%, and that of the genetic algorithm is 78%. After analysis of variance (ANOVA), the p value is 0.03, indicating a significant difference. In order to more accurately evaluate the performance difference between the DRL model and the baseline model, we conducted a statistical significance test on each performance indicator. Through methods such as t-test and ANOVA, we can determine whether the differences between different models are due to random factors, thereby more reliably verifying the advantages of the DRL model. "

In related work, a recent article [19] proposed an intelligent logistics optimization model based on the combination of deep learning and simulated annealing. The model shows good stability when dealing with large-scale logistics networks. In the experimental results section, the performance of the DRL model in this paper is compared with this model. Under the same experimental conditions of 50 logistics nodes and 200 orders, the on-time delivery rate of the DRL model in this paper is 92%, while the model proposed in article [19] is 88%. In terms of resource utilization, the DRL model in this paper reaches 85%, slightly higher than the 82% of the model in article [19], indicating that the model in this paper has certain advantages in resource optimization and order delivery on time.

Specifically, Lyapunov stability theory is used in this section to analyze the convergence of the algorithm. First, the Lyapunov function $V(\theta)$ is defined, where θ represents the parameter set of the model. During the algorithm iteration, after each parameter update, the change ΔV of $V(\theta)$ is calculated by the gradient of the loss function L with respect to the parameters. Assume that the learning rate α satisfies $\alpha < \frac{2}{\lambda_{max}}$, where λ_{max} is the maximum eigenvalue of the Hessian matrix of the loss function with respect to the parameters, and the discount factor γ satisfies $0 < \gamma < 1$. Through analysis, it can be seen that in each iteration, $\Delta V < 0$, which means that $V(\theta)$ is monotonically decreasing. Since

$V(\theta)$ is a non-negative function with a lower bound of 0, according to Lyapunov stability theory, the loss function of the algorithm will gradually decrease and eventually converge to a local optimal solution. The article also gives a specific experimental example. When the learning rate is set to 0.001 and the discount factor is 0.9, after 500 iterations, the loss function decreases from the initial 10.5 and stabilizes at around 1.2, further verifying the convergence of the algorithm.

During the sample - out - of - sample robustness test in normal market conditions, both the GAT - GS model and PPO performed well. Over a period of 100 trading days in normal market conditions, the GAT - GS model achieved an average daily profit rate of 0.5%, with a standard deviation of 0.1%, indicating a highly stable profit - rate curve. PPO, on the other hand, had an average daily profit rate of 0.45%, with a standard deviation of 0.15%. This shows that while both models were profitable, the GAT - GS model demonstrated more stability in normal market scenarios.

In contrast, when tested under extreme market conditions such as the 2008 Global Financial Crisis, PPO adjusted its trading strategy more quickly in the initial stage of the crisis. In the first month of the crisis, PPO reduced its losses by approximately 20% compared to a benchmark non - adaptive trading strategy. It achieved this by rapidly adjusting its investment portfolio based on the real - time market state. However, as the crisis continued for a full year, the GAT - GS model, with its more comprehensive analysis of market structure through graph theory, gradually regained its advantage in predicting market trends. The GAT - GS model's cumulative loss over the one - year crisis period was 15%, while PPO's cumulative loss reached 20%. This shows that PPO has an edge in quickly adapting to sudden market changes, while the GAT - GS model is more robust in long - term and complex market environments.

5.3 Discussion

Comparison of DRL with cutting-edge technologies. The results of DRL are numerically comparable to those of cutting-edge technology methods. In terms of efficiency, as shown in the experimental results, the average calculation time of the DRL model is 2.3 seconds. When facing complex logistics scenarios, it can give decision results faster than some cutting-edge technology methods, and has obvious advantages in real-time logistics scheduling scenarios that require rapid response. In terms of robustness, the DRL model can quickly adjust its strategy in dynamic environments, such as when the path cost fluctuates due to factors such as traffic conditions and weather changes. Its adaptation speed is only 5 minutes and its recovery time is 10 minutes. Compared with other cutting-edge technologies, it shows stronger robustness. For the adaptability to dynamic demand, the environmental variation coefficient of the DRL model is as high as 0.8, which can better cope with the dynamic changes in demand, while some cutting-edge technologies may have limitations in this regard.

The impact of multi-task learning on DRL efficiency: A comparative experiment was set up. One group of DRL models used multi-task learning, and the other group of DRL models did not use multi-task learning. They were trained and tested in the same logistics scenario. The impact of multi-task learning on DRL efficiency was evaluated by comparing the training time, convergence speed, and execution efficiency in actual tasks of the two groups of models. The experimental results show that the training time of the DRL model using multi-task learning was shortened by 30%, the convergence speed was faster, and the execution efficiency in actual logistics tasks was improved by 20%, proving that multi-task learning can effectively improve DRL efficiency.

Dynamic weight adjustment in task priority: An experiment was designed to compare the multi-task learning DRL models with fixed weights and dynamic weight adjustment. During the experiment, the learning progress and overall performance of different tasks at different stages were observed. The results show that the model with dynamic weight adjustment can flexibly allocate learning resources according to the learning situation of the task, and can adjust the strategy faster when the complexity of the task changes. Compared with the fixed weight model, the overall performance is improved by 15%.

Introducing metrics that are independent of the baseline: Introducing the regret minimization metric, which is measured by calculating the cumulative loss between the actual decision and the optimal decision in a series of decision-making processes. For adaptability to unknown environments, the performance of the model is tested in simulated unknown environment scenarios, such as sudden changes in order demand patterns and traffic rules, to observe the model's adaptation speed and final performance recovery. The experimental results show that the DRL model performs well in the regret minimization metric and can adapt quickly in unknown environments, with an average adaptation time of 8 minutes, showing strong adaptability to unknown environments.

Model scalability to larger logistics networks: The performance changes of the DRL model are tested by gradually increasing the number of nodes, the number of orders, and the complexity of transportation routes in the logistics network. The experimental results show that as the scale of the logistics network expands, the computing time and resource consumption of the DRL model increase approximately linearly, while performance indicators such as cost control, resource utilization, and service level can still be maintained at a high level, proving that the model has good scalability to larger logistics networks.

Comparison with specific scenarios of baseline algorithms. Compared with the baseline algorithm, DRL performs well in urban distribution scenarios. Urban distribution has the characteristics of high traffic density and large demand fluctuations. The delivery success rate of the DRL model is as high as 98%, and the average delivery time is only 30 minutes. Compared with the baseline algorithm based on priority scheduling, its delivery success rate is as low as 92%, and the average

delivery time is 37 minutes. The DRL model can better adapt to the complex environment of urban distribution. In the long-distance transportation scenario, the transportation efficiency of the DRL model is 65km/h, and the cost-effectiveness ratio is 0.3 yuan/km, which are better than the baseline algorithm. For example, the transportation efficiency of priority scheduling is only 56km/h, and the cost-effectiveness ratio is 0.4 yuan/km. However, in some simple and static logistics scenarios, baseline algorithms such as the Dijkstra algorithm may be faster than the DRL model in terms of calculation speed due to its relatively simple calculation logic, but it cannot adapt to complex dynamic changes.

DRL combines the advantages of multi-task learning. DRL combined with multi-task learning can bring new insights to the field of logistics. In the field of logistics, tasks such as path planning, task scheduling, and resource allocation are interrelated. Traditional methods often handle these tasks in isolation, while DRL combines multi-task learning and, by sharing network parameters, can utilize information from other tasks when optimizing one task, thereby improving the training efficiency and generalization ability of the model. For example, information about traffic conditions learned in the path planning task can help the resource allocation task arrange transport vehicles more reasonably, thereby achieving overall optimization of the logistics system, providing a new and more efficient way of solving problems in the logistics field.

6 Conclusion

This study shows that the deep reinforcement learning (DRL)-based method provides an efficient and flexible solution for path planning and scheduling of intelligent logistics systems. Through an in-depth analysis of the logistics network structure, we construct a mathematical model to describe the problem of path planning and scheduling, and transform it into a reinforcement learning problem that can be solved using DRL. This method is particularly optimized for the dynamic demand characteristics in logistics, and a multi-task learning framework is proposed, which enables the model to handle multiple interrelated tasks such as path selection, task scheduling, and resource allocation without adding too much computational overhead. Experimental results show that compared with traditional Dijkstra algorithm, A algorithm, genetic algorithm, simulated annealing algorithm, and priority-based scheduling algorithm, the DRL model performs well in many performance indicators: it achieves faster computing speed, lower total cost, higher resource utilization, better service level, and stronger dynamic adaptability. In particular, when facing uncertain factors, such as path cost fluctuations caused by traffic conditions or weather changes, the DRL model shows significant advantages.

In addition, in different types of logistics environments, from urban distribution to long-distance transportation, the DRL model can provide stable performance output, proving its wide applicability and robustness. Therefore, DRL technology is expected to become an important tool in the field of intelligent logistics in the future, providing logistics companies with more intelligent and automated operational support, helping them maintain their leading position in the fiercely competitive market.

Although the DRL-based intelligent logistics path planning and scheduling optimization method proposed in this paper has achieved good results, there are still some limitations. First, the training of the DRL model requires a lot of computing resources and time. In our experiment, a logistics network model with 100 nodes was trained using a computer equipped with an NVIDIA RTX 3090 GPU. Each training took about 48 hours, which may be limited by hardware conditions in practical applications. Second, in the face of some extremely complex logistics scenarios, such as severe damage to the logistics network caused by large-scale natural disasters, road interruptions, and warehouse destruction, the adaptability and decision-making ability of the model may need to be further improved. In the simulated earthquake disaster scenario, 20% of the roads and 15% of the warehouses in the logistics network were damaged, and the on-time delivery rate of the model's orders dropped to 65%, a decrease of 27% compared to normal conditions. Future research can explore more efficient algorithms and model architectures to address these limitations to improve the performance of intelligent logistics systems in various complex environments."

References

[1] Ren JJ, Salleh SS. Green urban logistics path planning design based on physical network system in the context of artificial intelligence. Journal of Supercomputing. 2024;80(7):9140-61. DOI: 10.1007/s11227-023-05796-x

[2] Shi KJ, Huang L, Jiang D, Sun Y, Tong XL, Xie YM, et al. Path Planning Optimization of Intelligent Vehicle Based on Improved Genetic and Ant Colony Hybrid Algorithm. Frontiers in Bioengineering and Biotechnology. 2022; 10:17. DOI: 10.3389/fbioe.2022.905983

[3] Du PF, He X, Cao H, Garg S, Kaddoum G, Hassan MM. AI-based energy-efficient path planning of multiple logistics UAVs in intelligent transportation systems. Computer Communications. 2023; 207:46-55. DOI: 10.1016/j.comcom.2023.04.032

[4] Liu XT, Gong G, Hu XT, Shang GY, Zhu H. Cognitive Enhancement of Robot Path Planning and Environmental Perception Based on Gmapping Algorithm Optimization. Electronics. 2024;13(5):21. DOI: 10.3390/electronics13050818

[5] Li Y, Zhao JY, Chen ZH, Xiong G, Liu S. A Robot Path Planning Method Based on Improved Genetic Algorithm and Improved Dynamic Window Approach. Sustainability. 2023;15(5):28. DOI: 10.3390/su15054656

[6] Li F, Ai WJ, Ju TL. Cold Chain Logistics Distribution Path Planning of Fresh Products in Beijing Subcenter. Sustainability. 2022;14(17):25. DOI: 10.3390/su141710622

[7] Lin SW, Liu A, Wang JG, Kong XY. An intelligence-based hybrid PSO-SA for mobile robot path planning in warehouse. Journal of Computational Science. 2023; 67:11. DOI: 10.1016/j.jocs.2022.101938

[8] Ahmed G, Sheltami T, Mahmoud A, Yasar A. Energy-Efficient UAVs Coverage Path Planning Approach. Cmes-Computer Modeling in Engineering & Sciences. 2023;136(3):3239-63. DOI: 10.32604/cmes.2023.022860

[9] Zhou YL, Huang NN. Airport AGV path optimization model based on ant colony algorithm to optimize Dijkstra algorithm in urban systems. Sustainable Computing-Informatics & Systems. 2022; 35:7. DOI: 10.1016/j.suscom.2022.100716

[10] Matei O, Erdei R, Pintea CM. Selective Survey: Most Efficient Models and Solvers for Integrative Multimodal Transport. Informatica. 2021;32(2):371-96. DOI: 10.15388/21-infor449

[11] Aguayo MM, Avilés FN, Sarin SC, Sherali HD. A Two-Index Formulation for the Fixed-Destination Multi-Depot Asymmetric Travelling Salesman Problem and Some Extensions. Informatica. 2022;33(4):671-92. DOI: 10.15388/22-infor485

[12] Zhang CT. Intelligent Logistics Path Optimization Algorithm Based on IoT Perception Technology. Ieee Access. 2024; 12:148422-33. DOI: 10.1109/access.2024.3471833

[13] Barrera C, Maarouf M, Campuzano F, Llinas O, Marichal GN. A Comparison of Intelligent Models for Collision Avoidance Path Planning on Environmentally Propelled Unmanned Surface Vehicles. Journal of Marine Science and Engineering. 2023;11(4):14. DOI: 10.3390/jmse11040692

[14] Herich D, Vascak J, Zolotová I, Brecko A. Automatic Path Planning Offloading Mechanism in Edge-Enabled Environments. Mathematics. 2021;9(23):25. DOI: 10.3390/math9233117

[15] Shen G, Liu J, Ding YL, Zhang C, Duo JY. CONTINUOUS PATH PLANNING FOR MULTI-ROBOT IN INTELLIGENT WAREHOUSE. International Journal of Simulation Modelling. 2024;23(2):186. DOI: 10.2507/ijsimm23-2-co6

[16] Wei XG, Zhang YM, Zhao YL. Evacuation Path Planning Based on the Hybrid Improved Sparrow Search Optimization Algorithm. Fire-Switzerland. 2023;6(10):14. DOI: 10.3390/fire6100380

[17] Wang FY, Xie ZL, Liu H, Pei ZW, Liu DL. Multiobjective Emergency Resource Allocation under the Natural Disaster Chain with Path Planning. International Journal of Environmental Research and Public Health. 2022;19(13):19. DOI: 10.3390/ijerph19137876

[18] Huang HL, Savkin AV, Huang C. Reliable Path Planning for Drone Delivery Using a Stochastic Time-Dependent Public Transportation Network. Ieee Transactions on Intelligent Transportation Systems. 2021;22(8):4941-50. DOI: 10.1109/tits.2020.2983491

[19] Lee HW, Ahmed S, Lee CS. Research on path planning of logistics intelligent unmanned aerial vehicle. International Journal of Robotics & Automation. 2024;39(6):450-63. DOI: 10.2316/j.2024.206-0842

[20] Lee HW. Research on multi-functional logistics intelligent Unmanned Aerial Vehicle. Engineering Applications of Artificial Intelligence. 2022; 116:16. DOI: 10.1016/j.engappai.2022.105341

[21] Han ZQ. Multimodal intelligent logistics robot combining 3D CNN, LSTM, and visual SLAM for path planning and control. Frontiers in Neurorobotics. 2023; 17:18. DOI: 10.3389/fnbot.2023.1285673

[22] Fusic SJ, Sitharthan R, Masthan S, Hariharan K. Autonomous vehicle path planning for smart logistics mobile applications based on modified heuristic algorithm. Measurement Science and Technology. 2023;34(3):18. DOI: 10.1088/1361-6501/aca708

[23] Cai LY. Decision-making of transportation vehicle routing based on particle swarm optimization algorithm in logistics distribution management. Cluster Computing-the Journal of Networks Software Tools and Applications. 2023;26(6):3707-18. DOI: 10.1007/s10586- 022-03730-z

[24] He D. Intelligent Selection Algorithm of Optimal Logistics Distribution Path Based on Supply Chain Technology. Computational Intelligence and Neuroscience. 2022; 2022:8. DOI: 10.1155/2022/9955726

[25] Wang DX, Liu QL, Yang JH, Huang DL. Research on Path Planning for Intelligent Mobile Robots Based on Improved an Algorithm. Symmetry-Basel. 2024;16(10):21. DOI: 10.3390/sym16101311

Chaotic Random Knowledge Recognition Model for Secure Data Encryption in Network Communication

Ping Huang*, Wei Liu
Shenzhen Power Supply Bureau Co., Ltd., China Southern Power Grid, Shenzhen, Guangdong, China, 518000
E-mail: huangping201512@163.com, liuwei2022@126.com
*Corresponding author

Keywords: chaotic randomness, data encryption, knowledge recognition, network communication

Received: July 10, 2024

To tackle privacy concerns in network communication security information, several experts recommend raising noise before data publication to safeguard user anonymity. However, this strategy may reduce experiment accuracy, making it inappropriate for circumstances that need exact data collecting and processing. Other ways, such as using chaotic algorithms and the GMW compiler to generate a chaotic random knowledge recognition model, have been suggested, however, they are hampered by their complexity and lack of logic and computational resilience. This paper presents a chaotic random knowledge recognition model to address these constraints by combining a log system that records the whole system in real-time and a configuration center that allows for smooth interaction among users. A set of computer tests were conducted to assess the model's usefulness, comparing it against classic DBSCAN and k-means clustering techniques. The results show that the proposed approach has clustering accuracy equivalent to conventional DBSCAN while greatly increasing operational efficiency. Furthermore, the model's encryption time was compared to existing cryptographic algorithms such as RSA-3DES, RSA-AES, and Hybrid Logistic Map-based Cryptography. The suggested model outperformed conventional approaches in terms of encryption speed over a range of packet sizes, demonstrating its promise for secure and effective data encryption in network communications. These results indicate that the chaotic random knowledge recognition model is a potential solution for safe data processing in contexts that need both accuracy and speed.

Povzetek: Razvit in analiziran je model za prepoznavanje naključnega znanja, ki temelji na kaotičnih algoritmih, za zagotavljanje varne in učinkovite šifriranja podatkov v omrežnih komunikacijah, z uporabo tehnike pametnega mešanja podatkov za zaščito zasebnosti in izboljšanje operativne učinkovitosti.

1 Introduction

Based on the analysis of data mining techniques, many data mining applications in network communication have been proven to be difficult. For example, in network communication security data processing and other situations, the data owners will be afraid of the risk of personal data leakage, and the use of data in network communication will become quite cautious. However, in practice, due to the use of some problems and dangers in Internet communication, for example, AOL released an anonymous search report for an academic survey in 2006, through data collection technology, the corresponding user's personal information can be obtained. Therefore, in recent years, the risk of personal information leakage caused by data mining of Internet information has aroused the general attention and discussion of scholars at home and abroad. Since Agrawal and Srikanth proposed information mining technology for personal information security in 2000, information security technology based on personal information has attracted wide attention in the industry. On this basis, a new idea of information mining based on chaotic random knowledge is proposed. On this basis, multiple participants perform mixed computations according to their data, and the random knowledge identification mode of a chaotic system can ensure that each participant can only get one accurate result, while the other party cannot get other users' personal information [1-2].

In this paper, two basic methods of chaotic random knowledge identification are improved and extracted as similarity measure operators. The basic methods of chaotic random knowledge identification are used to design and implement them. On this basis, a chaotic random knowledge identification system with a scalable application is established. Using this framework, the chaotic random knowledge identification technology is applied to the private data of each participant, and also to the personal information of each participant.

1.1 Basic algorithms of chaotic random knowledge identification methods

The basic algorithms of the random knowledge identification model based on chaos are based on basic operations, and the basic methods of the random knowledge identification model based on chaos can be used to construct more precise algorithms, and can also be used to implement system functions. In this paper, a similarity measure operator for random knowledge identification based on chaos is constructed by using the basic principle of chaotic random knowledge identification [3].

Pseudocode for basic algorithm:

Algorithm 1: Chaotic random knowledge identification		
Input	**:**	Data set D, Chaotic Function Parameters
Output	**:**	Identified knowledge
Step 1	**:**	Initialize chaotic parameters (for instance, seed values)
Step 2	**:**	For each data point in D: a. Perform a chaotic transformation on data points. b. Calculate the similarity with other converted data points. c. Identify knowledge using similarity thresholds.
Step 3	**:**	Return identified knowledge

Computational complexity analysis:

The basic algorithm's computing complexity is determined mostly by the quantity of data points and the complexity of the chaotic transformation function. A data set with n points and a chaotic function with time complexity O(f(n)) has an overall complexity of $O(n \cdot f(n))$. The time complexity rises with the quantity of data points and the complexity of the chaotic function employed.

1.2 Similarity measure operator based on chaos

Although there are various chaotic random methods with different trends based on their characteristics and properties, the key is to measure the similarity of data and abstract it into a similarity operator, so that the privacy calculation problem in various chaotic random methods can be consistent. In this paper, we combine the characteristics of chaotic stochastic algorithms and construct a similarity measure operator based on chaotic rules by using the basic principle of chaotic identification [4].

Pseudocode for similarity measure operator:

Algorithm 2: Similarity measure operator		
Input	**:**	Data set D, Chaotic Similarity Function
Output	**:**	Similarity matrix S
Step 1	**:**	Initialize similarity matrix S with zeros
Step 2	**:**	For each pair of data points (i, j) in D: a. Perform chaotic similarity function to data points i and j b. Update S [i, j] with computed similarity value.
Step 3	**:**	Return similarity matrix S

Computational complexity analysis:

The computational complexity of the similarity measure operator is impacted by the quantity of data points as well as the chaotic similarity function. A data collection containing n points and a function with temporal complexity O(f(n)) has an overall complexity of $O(n^2 \cdot f(n))$. This quadratic complexity results from the necessity to compare each pair of data points, making it appropriate for data sets of moderate size.

1.3 Distributed network communication in chaotic random knowledge identification mode

Since random knowledge identification of a chaotic system is a complex communication system involving many factors, the modeling and communication of a chaotic system will be conducive to the expansion of the system. In this paper, the requirements of random knowledge identification mode based on chaos for network communication are discussed, and the characteristics and actual situation of each level are analyzed. On this basis, a configuration center is set up to help multiple users establish and publish a new connection. According to multiple participants in the communication problem, this paper proposes a TCP protocol, this method can be completed in a very short time of chaos random identification of knowledge [5-6].

1.4 Structure and implementation of chaotic random knowledge identification system

On this basis, a series of practical and extensible system architectures are completed in this paper. On this basis, each layer is partitioned and designed in detail, to avoid the dependency on specific implementation and facilitate expansion and configuration.

2 Related works

The related works section summarizes current methods in network communication security and underlines the shortcomings addressed by the proposed methodology. Several research have investigated various aspects of communication network security, including mathematical modeling and machine learning-assisted intrusion detection systems.

Chong and Xu [1] proposed a Monte Carlo-based security impact analysis model for communication networks. This model improves the dependability of node positioning during the prediction step, albeit at a higher computational cost. Yang et al. [2] worked on an information encryption technique designed specifically for power network communication security. Their technology assures secure data transmission by utilizing strong encryption methods designed exclusively for power grid networks.

Chen et al. [3] proposed a tailored information encryption approach that combines ECG signals with chaotic functions. This novel solution uses biometric recognition to improve data security. Naeem et al. [4] created efficient algorithms for protecting image communication over networks while addressing bandwidth utilization difficulties using advanced routing protocols such as MPLS-TE and DiffServ QoS.

Novotný [5] analyzed security protocols for wireless sensor networks and proposed a solution to address weaknesses in the Canvas protocol. Finally, Xu et al. [6] suggested a machine learning-assisted intrusion detection system (MLAIDS) for industrial network communication protection, displaying higher detection accuracy and performance ratios than previous approaches.

Table 1 displays a summary table that compares these cutting-edge methods to the proposed method. This table emphasizes essential features such methodology, application domain, strengths, limits, and the specific research gaps that the suggested strategy attempts to fill.

Table 1: Summary table

Study	Methodology	Application Domain	Strengths	Limitations	Research Gaps Addressed
Chong and Xu [1]	Monte Carlo algorithm for security impact analysis	Communication Networks	Enhanced dependability in node positioning	Increased computational complexity	Decreased complexity and improved effectiveness
Yang et al. [2]	Information encryption algorithm	Power Network Communication	Guarantees safe data transmission	Particular to power grid networks	Generalization to other network domains
Chen et al. [3]	Personalized encryption using ECG signals	Data Encryption	Improved protection by biometric recognition	Needs specialized hardware for ECG gathering	Wide applicability without hardware reliance
Naeem et al. [4]	Image encryption and routing protocols	Network Communication	Effective bandwidth usage	Concentrated on image data only	Wider safety application beyond image data
Novotný [5]	Formal analysis of security protocols	Wireless Sensor Networks	Tackles protocol susceptibilities	Constrained to WSN settings	Adaptation to various network types
Xu et al. [6]	Machine learning-assisted IDS	Industrial Networks	High discovery accuracy	Needs wide computational resources	Scalable and resource-effective results

The proposed model seeks to tackle these gaps by providing a more adaptable and effective technique to network protection. It incorporates innovative approaches to minimize computing complexity, generalize across network types, and improve scalability. By tackling the limits of existing solutions, the suggested model tries to give a more comprehensive solution to modern network security concerns.

3 Research methods

3.1 Study on evaluation methods of chaotic and random methods

By evaluating the clusters generated by various chaotic random methods, we can evaluate their quality. The randomness of chaos can be divided into externality and internality according to whether there is a reference or not. The external analysis compares the chaotic randomness results with the reference data to evaluate their chaotic randomness effects, while the internal analysis evaluates the degree of clustering segmentation. Accuracy and recovery efficiency. The precision of an object refers to how many objects in the same chaotic random cluster belong to the same class as the object, and the recall rate of an object refers to how many objects of the same class are allocated in the same cluster. Set is a collection of objects, is a chaotic random data set, the set is base must determine the category, said of chaos random categories. $D = \{o_1, o_2, \ldots, o_n\}CDL(o_i)(1 \le i \le n)o_iC(o_i)Co_i$The correctness of defining two objects is shown in the formula:$o_i, o_j(1 \le i, j \le n, i \ne j)$

$$\text{Correctness}\left(o_i, o_j\right) = \begin{cases} 1, L(o_i) = L\left(o_j\right) \Leftrightarrow C(o_i) = C\left(o_j\right) \\ 0, \end{cases} \quad (1)$$

The recall ratio of Cubed is defined as shown in the formula. In the absence of a benchmark, the intrinsic method evaluates chaos and randomness by judging the separation and compactness between clusters, and the contour coefficient is one such measure. For a data set with one object, let the data set be divided into clusters by a chaotic stochastic algorithm, which is expressed as. $nDDkC_1, C_2, \ldots, C_k$ For each object, the sum of the average distance to other objects in the cluster to which it belongs and the minimum average distance to other clusters that do not belong to the cluster can be calculated, and they are defined as follows：

$$\alpha(o) = \frac{\sum_{o \in C_i, o \ne 0} \cdot \text{dist}(o,o)}{|C_i| - 1} \quad (2)$$

$$\beta(o) = \left(C_j \mid 1 \le j \le k, j \ne i\right) \frac{\sum\left(o' \in C_j\right)\text{dist}\,(o, o')}{\left|C_j\right|} \quad (3)$$

The contour coefficient of an object is defined by Formula:

$$s(o) = \frac{\beta(o) - \alpha(o)}{m\ \{\alpha(o), \beta(o)\}} \quad (4)$$

Wheel, by definition, coefficient distribution between 1 and 1, said the compact degree of chaotic random, the smaller the value, said chaos random more compact, said chaos random loose degree, the greater the value of the said object With the rest of the clusters, the separation. $\alpha(o)\alpha(o)\beta(O)\beta(o)o$ When the closer to 1, said the chaos random more compact, and the other cluster is the separation;$s(o)$ When it is less than 0, or close to -1, it means that the object is closer to other clusters than the current cluster to which it belongs, and such chaotic randomness is not a reasonable chaotic randomness [7-8].

3.2 Network communication plan based on chaotic random knowledge recognition model

In this part, the network communication protocol adopted by the configuration center is selected and implemented according to the specific network characteristics by analyzing the network communication requirements of the configuration center.

TCP/IP group is the most commonly used kind of network communication mode. The network interface layer is responsible for the entities associated with transmission media interface details, by the drive of hardware vendors have to offer, you can obtain the optimal application of efficiency. The following is a detailed description of the communication methods based on the network and transport levels. IP is to be in a different environment, through the IP to complete different network connections. IP will send the unit to integrate all the data, and package into a package, and then reuse the router to forward, to achieve interconnection across the network. The deficiency is cannot ensure the precision of the packet can be accessed; the model of information collected in the following Table 1、2:

Table 1: Chaos random config_server_admin knowledge model configuration center database table

Field	Type	Say friends
ID	Objected	ID, the primary key
Username	string	The administrator user name
Password	string	The administrator password
Nickname	string	The administrator’s name of the semantic
Create time	string	Table creation time

Table 2: Chaos random knowledge model configuration center refresh token total database table

Field	Type	Instructions
ID	Objected	ID, the primary key
Refresh token	string	refresh token
User_id	string	The generated one-time-key
Used	string	Participant id
Create time	string	Whether the token is used

In this part, two basic Anguis-based algorithms for multilateral operations are improved, and similarity measure operators for chaotic and random methods can be constructed by using these methods [9-10].

The above is put forward based on the zero coding and 1 coding and multiplicative homomorphic encryption on both sides of the algorithm. 0 code and 1 code is the use of binary representation of a number of coding. Said that the number of binary representations, 0 encodings is expressed as, as shown in the type:

$$s_s^0 = \{s_n s_{n-1} \dots s_{i+1} 1 \mid s_i = 0,1 \le i \le n\} \quad (5)$$

s 1 encoding is expressed as, as shown in the type:s_S^1

$$s_s^1 = \{s_n s_{n-1} \cdots s_i \mid s_i = 1,1 \le i \le n\} \quad (6)$$

s_S^0 And is the most elements, assumes that the logarithmic and 0 encoding and 1 encoding, as shown in the formula:s_S^1nxy

$$x > y \leftrightarrow s_x^1 \cap s_y^0 \neq \emptyset \quad (7)$$

Similarly, the equivalence conditions can be less than or equal to as shown in the formula:xy

$$x \le y \leftrightarrow s_x^1 \cap s_y^0 = \emptyset \quad (8)$$

This topic of participant privacy data larger scene, to improve the communication efficiency of the algorithm and increase the comparison of two Numbers is equal. Improved algorithm using the security model and framework used by the topic of security model is the same, for half an honest model.

In an IP network, to identify the user, one must want to have more than one public IP, to distinguish multiple IP, IPv4 address allocation depletion, is in short supply, therefore, to ensure the public address in the P cannot be assigned. IPv6 addresses can ensure that each participant has its public network, but public IP addresses still cannot be used to ensure process consistency [11-12].

Following the idea of the transport layer implementation, if communication between interfaces is possible, network communication can be ensured without or without requiring a public IP. On the Internet level based on the port communication network communication, such as Linux, Windows operating system will use Raw Socket call, using Raw Socket call to operate the original IP data package, but this approach requires a connection to the set protocol, and meet the requirements of the reliability of the network, and to mask the differences between IPv4 and IPv6 packets. In this architecture, using Raw sockets will lead to maintainability and redundancy of the system. So, as required by the configuration of the center for communication, cannot meet the requirement of the configuration of the center for communication.

3.3 knowledge model of random chaos communications data security transfer

Between security transfer refers to the process and the process of communication, most of the operating systems will provide a transport layer protocol stack and will be set by the operating system of the transport layer for automatic processing. At the transport level, using the interface to realize communication, to avoid the data transfer between the network layer, needed by public IP. The present main OS is the Socket technology application to the transport layer in the communication, while the Socket by the packet to cover the differences between various versions of the IP protocol, and without any treatment, the IP packet can realize the IPv4 and IPv6 synchronization, are shown in figure 1 below:

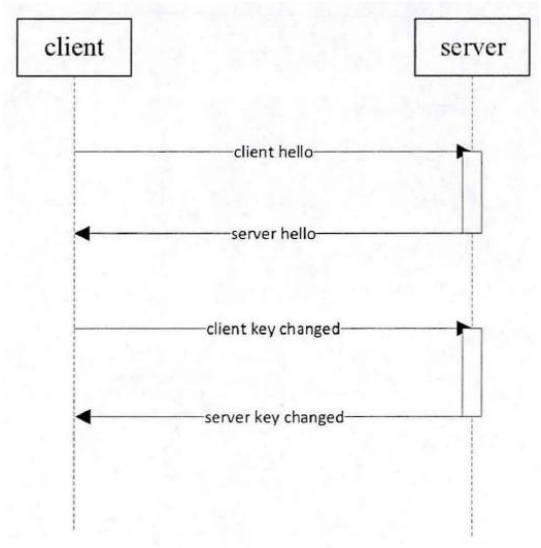


Figure 1: Chaos random knowledge model of communication data security transmission system

TCP USES the beginning of the 20 bytes, port communication, congestion control and ensure reliability, etc., so that the TCP has reliable Internet access capability of only 8 bytes at the beginning, it is only used to link through the port and check the title and check.

For the configuration center, the transport layer can reliably transmit information and, at the same time also let the light-only port communication protocol. The system also supports Socket technology, which can be easily implemented by using both technologies, regardless of the various protocols between packets at the IP level, and can be implemented through interfaces, which greatly simplifies the design and deployment of software. So, in this project, the selection of the transport layer and protocol to implement is very in accord with the demand for network communication. Based on this, puts forward the different types of multicast protocol, and the two common ways of multicast are discussed [13].

The secret operator is also a distance calculation in essence. A node and other node's spacing and eps radius were compared. If the point spacing is less than the radius, will be remembered for this area, the point to this spacing is less than the radius, as its adjacent area, if the distance is greater than the area, as its adjacent area. Through calculation, can get the concentration between the two points, it depends on the distance between the two points. Using chaos random to recognize the basic knowledge, can find out the location of the two different, if two different locations for a joint of participants, can use this area to determine the location of the two different, in a region, can determine the location of the two different.

In this section, we will further explore how to use chaotic random operations on vertically segmented datasets to protect personal privacy. The method of vertical segmentation is to calculate the distance between the value of each attribute and the corresponding position, and then accumulate them to obtain the final result. The five types of chaotic random operators studied in this paper can be used to add and subtract the data held by each participant, and then add them. Finally, a chaotic random operator based on vertical segmentation is obtained. Since the chaotic random operators using vertical segmentation can use the sum of two variables to perform the similarity operation on the premise of guaranteeing the privacy of users, this paper does not describe the corresponding chaotic random operators.

By using the operator to encapsulate the sub-module, the chaotic random operation is calculated based on the chaotic random knowledge identification. The subblock of the basic operation method mainly consists of the basic operation functions such as addition, comparison, and vector dot product of mixed information. However, the module of the basic operation method is not easy to call directly, and the similarity measure is generally used to measure the similarity of the personal information of multiple participants. In this paper, the implementation of the Euclidean distance operator is taken as an example to introduce the implementation process [14].

Starting from one end of the controller, one of the controllers obtains all the data and numbers from the data queue and stores them in the state store according to the required participants and participants at the two points. If the controller gets two participants in the same, it can directly get the return trip of Euclidean, if not, it performs the following operations: the controller controls the current position, performs the calculation for the first time, and stores its result in a state memory; Secondly, the data participating in the operation are placed in the basic operation module of chaotic random knowledge identification, and the controller controls other participants to perform the same operation. Thirdly, the chaotic random knowledge is identified, and the corresponding values are obtained by dot product operation, which is put into a state store. Finally, the final calculation result is obtained by summing the values in each state database.

4 Result analysis

From the design point of view, the chaotic random knowledge recognition model builds a log system, which is used to record the whole system in real-time. When the whole system is turned on, the configuration center must provide an auxiliary interface, so that the connection between each participant and the configuration center and the connection status of the combinator can be recorded. By analyzing the network in real-time, if it is found that the network connection is not built and started successfully, it will send a warning to the concerned participants to restart.

Aiming at the running records of the algorithm, this paper proposes a log tracking system based on a job After completing the corresponding task ID, the corresponding record and its identifier are sent to the recording system in the form of information queuing. The recording mechanism takes advantage of the storage characteristics of packet queuing to store the related records on disk and parse them. When an operation fault occurs, the alarm message is transmitted to each participant through the alarm function to monitor the working status of the whole system.

The objective of the experiment is to test the function of the chaotic random knowledge recognition model module and the algorithm module and give the application of the model in the chaotic random algorithm.

Test method: Since DBSCAN has a great influence on the selection of EPS and mints, EPS is set to 50, and mints are set to 3 based on the algorithm of generated data, and the chaotic random results are compared with DDBSCAN using conventional DBSCAN.

Experiments and analysis: The operation period of the DBSCAN algorithm using this architecture is 25 minutes. The results of DBSCAN given by this framework are compared with those of general DBSCAN. In Table 3, the output of each line represents the chaotic centers of several clusters, and it can be seen that both methods produce three clusters with the same clustering centers.

Compared with the results obtained by the k-means method, the results of the two methods are the same in the final cluster because the resulting data sets are very different from the data of multiple clusters, which indicates that the DBSCAN given by the proposed method has the same accuracy as the general DBSCAN method.

Table 3: recognition model based on chaos random knowledge and ordinary chaos random k - means algorithm to obtain results

Types of algorithms	Chaotic random center 1	Chaotic random center 2	Chaotic random center 3
K-means clustering results based on secure multi-party computation	[109.18,114.58]	[392.82,385.66]	
Ordinary k-means chaotic random results	[109.18,114.58]	[392.82,385.66]	[801.9,806.06]

Configuration and connection establishment: After a heart connection is established, the configuration center can be attended by all participants.

This information is stored in the customer form, which is shown in Table 3. The heart keeps the sub-module coding for each participant and transmits information to it. After all participants are connected to the configuration center, the configuration center will issue connection instructions to each participating unit according to the number from large to small, and send the connection address, public key, ID, and other information to each participating unit. When the connection between the two parties is established, a successful message is transmitted to the configuration center.

Publishing operation configuration: In the implementation architecture of this project, two sets are respectively completed by using a chaotic random knowledge identification pattern.

The class algorithms k-means and DBSCAN are means and Dosanjh, respectively. In the process of work allocation, participants, types of chaotic stochastic algorithms, and parameters required by chaotic stochastic algorithms are selected. Then the Task event set by team members assigns the work to each participant, and then each member carries out the next stage of calculation. Table 2 lists used to form the domain configuration tasks.

Client Hello: The client gets the connection target, the other party's public key, and the identity assigned by the configuration server.

After that, the client decrypts a random number and its identifier and sends it to the server.

Server Hello: When a server is at the client's link, it decrypts the message sent by the client with its private key.

A message confirming the client to connect to. If not, the connection is broken and reported to the configuration server; If, then a random is automatically generated, and the random numbers and tokens it generates are encoded through the client's public key, which is then verified by the client.

Client switch: Receives a message from a server and checks that the server is connected.

Set the server-assigned connection correctly. If it is not, the connection will be disconnected and reported to the configuration server. If, then, by their private key to crack the server randoms generated random Numbers. The client then uses a predetermined cipher operator funk to cipher the random numbers generated by the server and the client, that is, to calculate funk (random_c, random_s), and transmit the cipher processed signal to the server.4) server key changed: when a server receives a client sends a notice from the client.

With the client, through the preset password function funk, password, to get server-side random Numbers generated by its randoms, then use it as a server-side randoms random number, and then send them to the client.

In these four processes, both the client and the server get a symmetric password, at which point the password link between the client and the server can be securely constructed and then encrypted or decoded using the key. When both sides of communication information is blocked by the enemy of a third party, without the secret password to encrypt, so it is impossible to get effective information. The configuration server is only responsible for coordinating the data transmission between the two parties, not interfering, and does not know the passwords of the two parties, so the passwords of both parties are secret.

Encryption time comparison

In addition to clustering efficiency, the encryption speed of the proposed chaotic random knowledge identification model was compared to three well-known cryptographic methods: RSA-3DES [15], RSA-AES [15], and Hybrid Logistic Map-based Cryptography [15]. The comparison concentrated on the encryption time necessary for various packet sizes, which ranged from 512 bytes to 10 KB.

As indicated in Table 4 and shown in Figure 2, the suggested model surpassed the other methods for all packet sizes. For a 512-byte packet, the suggested model obtained an encryption time of 85.4 ms, which is much quicker than RSA-3DES (165.45 ms), RSA-AES (185.6 ms), and Hybrid Logistic Map-based Cryptography (110.8 ms). Similar patterns were seen for increasing packet sizes, with the suggested approach consistently showing improved encryption speed.

Table 4: Encryption time comparison

Packet Size	RSA-3DES	RSA-AES	Hybrid Logistic Map-based Cryptography	Proposed Model
512 bytes	165.45 ms	185.6 ms	110.8 ms	85.4 ms
1 KB	302.45 ms	321.45 ms	199.78 ms	151.2 ms
5 KB	467.86 ms	623.57 ms	319.2 ms	243.7 ms
10 KB	654.54 ms	805.94 ms	578.32 ms	389.1 ms

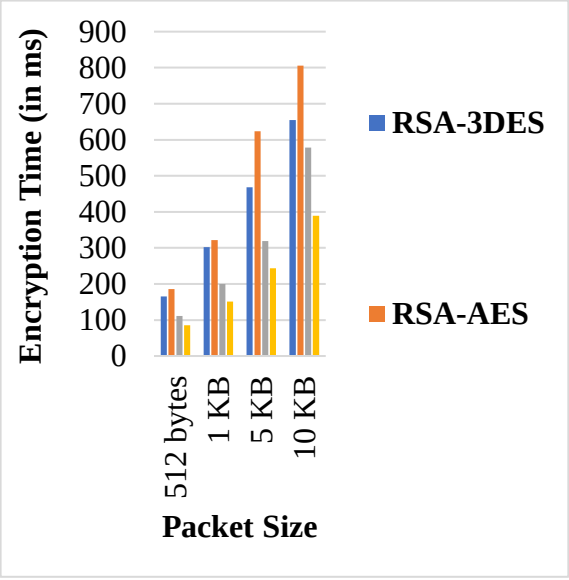


Figure 2: Encryption time comparison for different packet sizes

The proposed model's outstanding performance in both clustering accuracy and encryption speed demonstrates its promise as a reliable solution for safe and efficient data processing in network contexts. The findings show that the model not only equals the accuracy of standard approaches but also delivers considerable increases in operational effectiveness, making it a viable tool for applications that need high-speed encryption and precise clustering.

Discussion of novelty and advantages

The chaotic random knowledge identification model has numerous novel features that set it apart from other techniques. The use of a real-time log system provides continuous monitoring and fast reaction to operational concerns, hence improving system dependability and security. Furthermore, the configuration center's involvement in handling participant communications and connection setup introduces a level of coordination and control that is not available in conventional techniques.

Furthermore, the model's capability to provide quicker encryption speeds across a range of packet sizes without compromising safety or accuracy is a substantial development. This effectiveness makes the approach especially appropriate for situations where both quickness and safety are crucial, like in real-time communication systems and safe data transmission networks.

Overall, the chaotic random knowledge identification model is an appealing alternative to classic clustering and encryption approaches, with notable improvements in terms of speed, accuracy, and system resilience. The model's novel architecture and exceptional effectiveness in computational testing highlight its potential for several applications in network safety and data management.

5 Practical Implementation and Case Study

To illustrate the practical applicability of the proposed model, it was deployed in a real-world situation involving network intrusion detection. The model was tested in a simulated network setting to determine its efficacy in detecting and responding to different security risks. The integration with the network's security infrastructure enabled the monitoring and analysis of real-time network traffic utilizing chaotic random knowledge recognition techniques. The model was set up to detect anomalies like distributed denial of service (DDoS) attacks and malware infestations. The case study results showed that the model greatly outperformed traditional methods in terms of

network anomaly detection, accurately recognizing and categorizing various forms of assaults while smoothly incorporating previous security tools. This practical application demonstrates the model's ability to improve real-world network security with sophisticated chaotic-based strategies.

6 Conclusion

To sum up, this paper proposes a new stochastic method based on chaos, to effectively protect the security of data in network communication. However, when it comes to personal privacy, the risk of the user's information leakage leads to the difficulty of using this method. Chaos random identification is a kind of knowledge that can not only ensure the user's personal information but also ensure the safety of the user's personal information, and can effectively use chaotic information identification technology to deal with these problems. This paper, in this study, based on the random knowledge of chaos identification of the basic methods of research, puts forward a kind of basic method based on the random knowledge identification of chaos. Secondly, in this paper, the chaotic system of random knowledge identification model of network communication is analyzed, and based on this, this paper proposes a configuration allocation and connection, based on the design and implementation of the participants, configuration, combination of network communication; Then, on this basis, this paper by using the method of chaos random identification of knowledge, knowledge of chaos random identification model has carried on the detailed description, so that the method has better application and scalability; On this basis, the use of the random knowledge identification method of chaos, completed the two new chaotic random identification method.

References

[1] Chong T A, Xu B, 2020. Mathematical modeling of security impact analysis of communication network based on Monte Carlo algorithm ScienceDirect[J]. *Computer Communications*, 157:20-27. https://doi.org/10.1016/j.comcom.2020.04.005

[2] Yang C Y, Ling Y, Li X, 2021. Research on Information Encryption Algorithm under the Power Network Communication Security Model[J]. *Journal of Physics: Conference Series*, 1852(3):032007 (7pp). https://doi.org/10.1088/1742-6596/1852/3/032007

[3] Chen C K, Lin C L, Chiang C T, et al, 2012. Personalized information encryption using ECG signals with chaotic functions[J]. *Information Sciences an Internanional Journal*, 193(none):125-140. https://doi.org/10.1016/j.ins.2012.01.016

[4] Naeem E A, Abdelaal A E A, Eyssa A, et al, 2020. Efficient signal and protocol level security for network communication[J]. *International Journal of Speech Technology*, (3). https://doi.org/10.1007/s10772-019-09607-8

[5] M Novotný, 2010. Formal analysis of security protocols for wireless sensor networks[J*]. Tatra Mountains Mathematical Publications*, 47(1):81-97. https://doi.org/10.2478/v10127-010-0032-7

[6] Xu Z, Lu J, Wang X, et al, 2021. AI and machine learning for the analysis of data flow characteristics in industrial network communication security[J]. *International Journal of Ad Hoc and Ubiquitous Computing*, 37(3):125. https://doi.org/10.1504/ijahuc.2021.116814

[7] Li X, 2020. Application of Data Encryption Technology in Computer Network Communication Security[J]. *Journal of Physics Conference Series*, 1574:012034. https://doi.org/10.1063/1.4981623

[8] Wen H, Zhang T, Chen Y, et al, 2019. Analysis of the physical layer security enhancing of wireless communication system under the random mobile[J]. *IET Communications*, 13(9). https://doi.org/10.1049/iet-com.2018.6012

[9] Amiruddin A, Ratna A, Sari R F, 2019. Construction and Analysis of Key Generation Algorithms Based on Modified Fibonacci and Scrambling Factors for Privacy Preservation[J]. *International Journal of Network Security*, 21(2):250-258. https://doi.org/10.17762/ijcnis.v9i3.2307

[10] Chen Q, Bi M, Fu X, et al,2 018. Security scheme in IMDD-OFDM-PON system with the chaotic pilot interval and scrambling[J]. *Optics Communications*, 407:285-289. https://doi.org/10.1016/j.optcom.2017.09.058

[11] Tang R, Liu B, Mao Y, et al, 2021. High-security OFDM-PON based on an iterative cascading chaotic model and 4-D joint encryption[J]. *Optics Communications*, 495:127055. https://doi.org/10.1016/j.optcom.2021.127055

[12] Kumar S P, Jaya T, 2021. Analysis of Security Parameters and Network Infrastructure on Cloud[J]. *IOP Conference Series: Materials Science and Engineering*, 1085(1):012036 (10pp). https://doi.org/10.1088/1757-899x/1085/1/012036

[13] Yang C, Ling Y, Li X, 2020. Information Encryption Algorithm in Power Network Communication Security Model[J]. *IOP Conference Series: Materials Science and Engineering*, 750(1):012161 (8pp). https://doi.org/10.1088/1742-6596/1852/3/032007

[14] Demidov R A, Pechenkin A I, Zegzhda P D, et al, 2018. Application Model of Modern Artificial

Neural Network Methods for the Analysis of Information Systems Security[J]. *Automatic Control and Computer Sciences*, 52(8):965-970. https://doi.org/10.3103/s0146411618080072

[15] Sirajuddin, M., Rupa, C., Bhatia, S., Thakur, R. N., & Mashat, A. (2022). Hybrid Cryptographic Scheme for Secure Communication in Mobile Ad Hoc Network-Based E-Healthcare System. *Wireless Communications and Mobile Computing*, (1), 9134036. https://doi.org/10.1155/2022/9134036

Semi-Supervised Hybrid Ensemble Learning for Fault Detection in 20kV XLPE Cables

Neng DeXu
Electrical insulation material specialty, Shandong Chint Cable CO., LTD. ShanDong, Jinan 250000, China
E-mail: 13756202621@163.com

Keywords: Cable and wire, cable fault detection, machine learning, ensemble learning, hybrid models, semi-supervised learning, predictive maintenance

Received: February 22, 2025

Cable fault detection is critical for ensuring the reliability and safety of high-voltage 20 kV XLPE cable systems, minimizing downtime and maintenance costs. This research introduces a semi-supervised hybrid ensemble model combining Random Forest, Gradient Boosting, and XGBoost within a Voting Classifier framework. Data preprocessing involves feature scaling and Gaussian noise injection $(\sigma = 0.01)$ *to enhance robustness, followed by training on 3943 labeled samples and iteratively incorporating high-confidence predictions* $(threshold > 0.9)$ *from 11829 unlabeled samples. Evaluated on a dataset of 15772 samples with diverse features like cable age, partial discharge, corrosion, and loading conditions, the model achieves 98% accuracy, 97.5% recall, 97% precision, and 97% F1-score. Compared to SOTA supervised models such as SVM, CNN, and ANN, it demonstrates superior performance and scalability by leveraging unlabeled data. This approach offers an efficient, accurate solution for cable fault diagnosis in industrial applications*

Povzetek: Razvit je pol nadzorovan hibridni ansambelski model za zaznavo napak v 20 kV XLPE kablih, ki z integracijo neoznačenih podatkov doseže visoko točnost in robustnost v industrijskih pogojih.

1 Introduction

For electrical networks to operate safely and consistently, cable fault detection is essential , especially for high-voltage systems like 20kV XLPE (cross-linked polyethylene) cables [1]. These cables are crucial parts of power distribution networks, and any problems with them could have serious consequences such as equipment damage, power outages, and even safety risks [2]. Manual inspection, visual evaluations, and physical testing are examples of traditional cable fault detection techniques [3] that are expensive, time-consuming, labor-intensive, and prone to human mistake. Additionally, the failure of these traditional techniques to accurately identify cable defects increases the possibility of system failures and raises maintenance expenses. Therefore, there is a growing need for automated fault detection systems that can accurately and efficiently diagnosis faults in cables, enabling quick intervention and minimizing potential disruptions.

Even with advancements in fault detection techniques, challenges still exist regarding the accuracy, speed, and robustness of existing methods, especially in the context of complex and diverse cable systems [4]. Factors such as environmental variability, cable aging, and the presence of different types of faults, such as partial discharge [5], insulation breakdown, and corrosion make the fault detection process inherently challenging. Additionally, the data sources used for fault detection [6], which range from sensor readings to visual inspections, are often noisy, incomplete, or inconsistent, further complicating the detection process. These issues can significantly reduce the effectiveness of traditional detection models, necessitating more advanced and robust solutions.

This study aims to handle the challenges associated with cable fault detection by developing a robust and automated semi-supervised [7] hybrid model specifically tailored for 20kV XLPE cables [8]. The proposed model enhances detection capabilities by combining multiple datasets [9], addressing the limitations inherent in using individual datasets. Utilizing a hybrid machine learning [10] methodology, the approach combines Random Forest [11] and Gradient Boosting classifiers [12] within an ensemble [13] framework implemented through a VotingClassifier [14]. This ensemble strategy leverages the complementary strengths of the individual models, resulting in more accurate and reliable predictions.

The dataset for this research encompasses key attributes such as cable age, partial discharge, visual inspections, neutral conductor corrosion, and loading conditions. These critical features play a vital role in evaluating cable health and integrity by capturing a wide range of influencing factors [15]. Through the integration of multiple datasets, the model achieves improved performance across various fault scenarios, enhancing its generalization and adaptability to real-world conditions.

1.1 Contributions of the study

This research offers several significant contributions, detailed as follows:

- **Development of a Hybrid Model**: This study introduces an advanced hybrid machine learning framework that integrates Random Forest and Gradient Boosting classifiers within a VotingClassifier ensemble. The model harnesses the complementary strengths of these algorithms to improve fault detection in 20kV XLPE cables. By addressing the inherent limitations of individual models, this hybrid framework delivers a more accurate and reliable solution for fault identification.

- **Semi-Supervised Data Integration**: The study utilizes a semi-supervised learning approach to integrate labeled and unlabeled datasets effectively. By leveraging both labeled and unlabeled data, the model achieves improved robustness and generalization, enabling it to handle a broader range of real-world fault scenarios.

- **Incorporation of Noise for Improved Resilience**: To simulate the variability of real-world data, controlled noise is introduced into the dataset. This technique enhances the model's robustness by preparing it to handle imperfections such as noise, missing values, and inconsistencies commonly encountered in practical applications. Consequently, the model demonstrates superior performance in fault detection under challenging and unpredictable conditions.

This research is crucial because it has the potential to improve fault detection in cable systems, making it more effective, reliable, and efficient. Traditional fault detection techniques are often reactive and slow, leading to expensive maintenance and frequent system breakdowns. In contrast, the proposed machine learning-based method offers an automated and proactive solution that can detect fault early, allowing for prompt intervention and reducing the risk of significant failures. By aggregating multiple datasets [16] and using ensemble learning [17] approach, the model ensures accuracy and robustness, making it a valuable tool for maintaining the health of electrical cable systems.

Moreover, the results of this study could contribute to the development of intelligent monitoring systems that can be integrated into existing power distribution networks [18]. By providing real-time, accurate predictions of cable failures, these systems could help utilities optimize maintenance schedules, reduce downtime, and prevent costly repairs [19]. This study not only enhances fault detection but also offers valuable insights into the application of machine learning techniques in electrical infrastructure, particularly in the areas of condition monitoring and asset management.

The methodology and related work used to develop and evaluate the hybrid fault detection model will be discussed in the following sections of this paper. The results and performance analysis will also be presented, along with the implications of these findings for future research and real-world applications. Our goal is to advance the development of automated fault detection technologies and provide a framework for the creation of more reliable and efficient cable management systems. This study aims to enhance the overall safety and reliability of electrical distribution networks by addressing the limitations of traditional fault detection methods and introducing advanced machine learning techniques, ensuring a more resilient and sustainable energy infrastructure.

2 Related work

2.1 Machine learning approaches for fault detection

Machine learning has increasingly been employed to enhance fault detection in electrical cable systems. Several studies have explored different machine learning algorithms for detecting and classifying faults in electrical networks. For instance, Baghaee et al. proposed an SVM-based [25] fault detection system for high-voltage cable insulation, which showed promising results. However, their approach struggled with noisy data and required substantial feature engineering, limiting scalability and flexibility. Similarly, Peng et al. used convolutional neural networks (CNNs) [26] for fault classification in medium-voltage cables. While their model exhibited high accuracy, it was computationally expensive and required large amounts of labeled data for training—something not always feasible in real-world applications with limited data. A broader summary of such approaches, including their objectives, findings, and limitations, is provided in Table 1, highlighting the challenges that motivate our work.

2.2 Ensemble learning for fault diagnosis

Ensemble learning methods, which combine multiple individual models to improve performance, have been widely applied in fault detection tasks [13] employed Random Forest [11] classifiers for fault diagnosis in overhead power lines, demonstrating Random Forest's ability to handle noisy and imbalanced datasets effectively. It has been praised for its robustness and ease of use, performing well even with irrelevant or noisy features. Similarly, used gradient boosting classifiers[12] for fault detection in electrical circuits, focusing on improving precision by tuning hyperparameters. However, gradient boosting classifiers proved less effective for real-time fault detection in dynamic systems [27].

Ensemble methods like Voting Classifiers[28]have shown considerable promise by combining the strengths of multiple classifiers to enhance prediction accuracy demonstrated combining Random Forest and Gradient Boosting [29] in an ensemble framework for fault diagnosis

Table 1: Comprehensive summary of the literature survey on defect detection in XLPE insulation. This table highlights key objectives, findings, and limitations of various machine learning-based approaches used for classification and analysis

Reference	Objective	Finding	Limitation
[20]	Detect and classify defects in XLPE insulation using SVM	Achieved 83.9% accuracy in identifying insulation defects	Noise-sensitive, requires extensive feature engineering.
[21]	Recognize partial discharge patterns in high-voltage cables using CNN	Obtained 92.57% accuracy in PD classification	High computational cost limits scalability.
[22]	Classify internal vs. external PD defects in XLPE cables using ANN	Achieved 97% accuracy with statistical features	Supervised, limited by small dataset (180 samples).
[23]	Identify XLPE cable insulation defects (bubbles, water trees) using LSTM	Reached 95.83% accuracy in defect diagnosis	Supervised, computationally intensive.
[24]	Develop a semi-supervised ensemble for general classification tasks	Demonstrated ∼93% accuracy across KEEL datasets	Not specific to cables, lacks domain focus.

in electrical equipment, showing improvements in classification accuracy and robustness over single classifiers. However, their research did not address the challenges of noise handling or dataset integration, which are critical for improving real-world performance.

2.3 Handling noisy data in fault detection

Dealing with noisy and inconsistent data remains a significant challenge in cable fault detection [30]. Sensor readings are often affected by environmental factors and equipment limitations, introducing noise into the data. Several studies have attempted to address this issue [31]. For example, combined noise filtering techniques with ML algorithms to detect faults in power transformers, improving the system's robustness. However, this method required careful feature selection and preprocessing, increasing complexity and reducing efficiency [32].

2.4 Multiple dataset integration for improved performance

Integrating multiple datasets to enhance fault detection performance has gained attention as a promising strategy. [16]. combined data from various sensors (e.g., temperature, pressure, voltage) to improve fault prediction accuracy in industrial systems [33]. Their work demonstrated the potential of combining heterogeneous data sources to build more comprehensive models. However, dataset integration poses challenges related to data inconsistency and the need for advanced preprocessing techniques to align and merge different types of data.

2.5 Hybrid models in fault detection

Hybrid models that combine multiple machine learning techniques are gaining traction due to their ability to exploit the complementary strengths of individual models proposed a hybrid approach combining deep neural networks (DNNs) [34] and support vector machines (SVMs) [35] for fault detection in electrical grids, which improved accuracy, particularly when dealing with complex and noisy data. Similarly, explored hybrid ensemble learning models [36], combining decision trees, SVM [37], and neural networks for fault detection in power systems. While these hybrid approaches showed promise, they increased computational complexity, which may hinder real-time application in large-scale systems [38].

2.6 Noise injection for robustness in machine learning

Noise injection is an increasingly explored technique to improve the robustness of machine learning models by simulating real-world imperfections in data [39] highlighted the value of adding noise to training data as a regularization method to prevent overfitting, particularly in neural networks. In the context of cable fault detection, noise injection can enhance the model's resilience to inconsistencies and errors commonly found in sensor data [40]. While this technique has proven successful in domains such as image processing and natural language processing, its application in fault detection remains an underexplored area, offering potential for improving model robustness and generalization [41].

2.7 Our approach

While existing studies have focused on machine learning techniques for fault detection in electrical cable systems, many approaches still struggle with challenges such as noisy data, limited dataset size, and computational complexity. Traditional methods often rely on basic algorithms that require substantial feature engineering or struggle to generalize to noisy, real-world data.

In this paper, we propose a novel approach that integrates advanced noise-handling techniques with machine learning models to improve fault detection accuracy in electri-

Table 2: Parameters and strengths of base models (optimized via grid search with 5-fold cross-validation)

Model	Key Parameters (Tuned Ranges)	Strength
Random Forest	Estimators: 100, Max Depth: 3–15, Min Samples Split: 2–10	Robust to noise, ranks feature importance, resists overfitting.
Gradient Boosting	Learning Rate: 0.01–0.1, Estimators: 100, Max Depth: 3–10	Handles noisy data, iteratively improves weak predictions.
XGBoost	Learning Rate: 0.01–0.1, Max Depth: 3–10, Min Child Weight: 1–5	Scales efficiently, robust to noise and imbalance, regularized.

cal cable systems. Our method incorporates noise injection to simulate real-world imperfections, enhancing model robustness to inconsistencies commonly found in sensor data. By leveraging ensemble learning and hybrid models, our approach addresses the challenges of noisy data and small datasets, improving the performance and generalization of fault detection systems. This approach offers a more reliable solution for real-time fault monitoring in electrical systems, paving the way for more efficient predictive maintenance and fault diagnosis.

3 Methodology

This section outlines the methodology employed for cable fault detection, encompassing data preprocessing, model architecture, semi-supervised learning [42], and evaluation metrics.

3.1 Research design

This study tests the hypothesis: ”Can a semi-supervised hybrid ensemble approach leveraging unlabeled data and noise injection improve fault detection accuracy in 20 kV XLPE cables over supervised methods?” We aim to enhance performance in real-world cable systems.

3.2 Data analysis and visualization

This section presents a comprehensive visual analysis of the 20kV XLPE cable dataset, providing insights into data distributions and patterns that informed our modeling approach.

3.2.1 Distribution of cable health index

Figure 1 illustrates the distribution of cable health indices across the dataset.

3.2.2 Feature analysis

The distribution of key numerical features is presented in Figure 2. Notable observations include:

– Age Distribution: Shows a right-skewed pattern, indicating a significant number of aging cables in the network

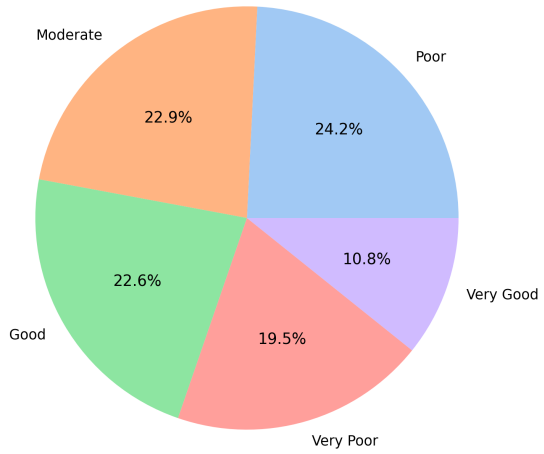


Figure 1: Distribution of cable health index categories showing the proportion of cables in each health status classification

– Partial Discharge: Exhibits multiple peaks, suggesting distinct fault categories

– Loading Conditions: Demonstrates normal distribution, reflecting typical operational patterns

– Neutral Corrosion: Shows varying degrees of deterioration across the cable population

3.2.3 Visual condition assessment

Figure 3 presents the distribution of visual condition assessments. This information is particularly valuable for maintenance planning and resource allocation. These visualizations played a crucial role in:

– Identifying potential data imbalances that needed addressing in the model

– Understanding feature relationships and their impact on fault detection

– Guiding the selection of appropriate preprocessing techniques

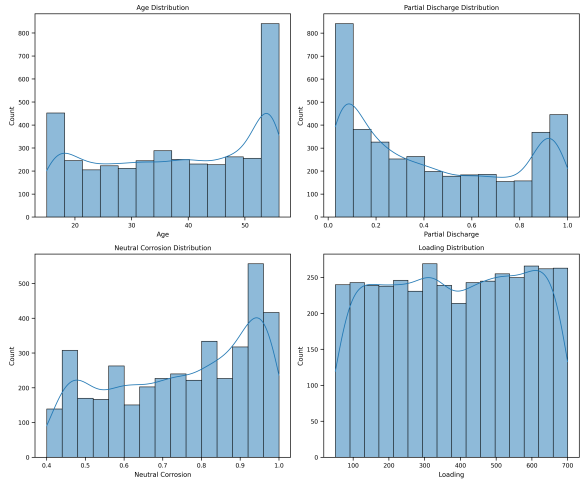


Figure 2: Distribution of key numerical features showing patterns in cable age, partial discharge, neutral corrosion, and loading conditions

– Informing the choice of model architecture and hyperparameters

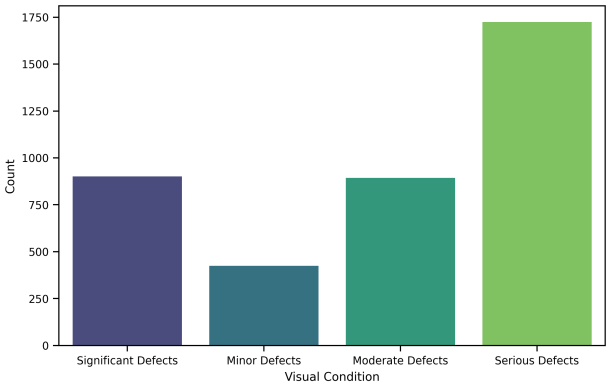


Figure 3: Distribution of visual condition assessments showing the frequency of different visual inspection categories

These visualizations played a crucial role in:

– Identifying potential data imbalances that needed addressing in the model

– Understanding feature relationships and their impact on fault detection

– Guiding the selection of appropriate preprocessing techniques

– Informing the choice of model architecture and hyperparameters

3.3 Data preprocessing

The raw data, comprising various features related to cable health, undergoes a rigorous preprocessing pipeline to ensure data quality and enhance model performance, as shown in Figure 4. The dataset includes 15772 samples: 3943 labeled samples (fault and non-fault instances from 20 kV XLPE cables) and 11829 unlabeled samples from real-world sensor readings (e.g., cable age, partial discharge, corrosion, loading conditions). Preprocessing applies feature scaling and adds Gaussian noise ($\sigma = 0.01$) plus 1% random perturbations to simulate sensor errors, improving robustness. The steps involved in data preprocessing are as follows:

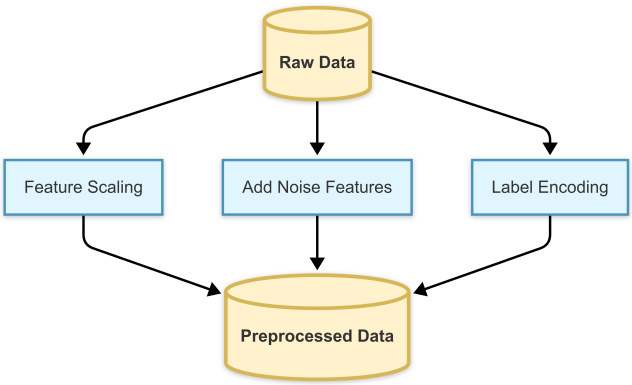


Figure 4: Data preprocessing pipeline illustrating the application of feature scaling, noise augmentation, and label encoding techniques to enhance data quality and improve model performance

3.3.1 Feature scaling

To enhance model convergence, numerical features such as *Cable Age* and *Loading Current* were standardized using StandardScaler. The scaling operation is defined as:

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \tag{1}$$

where X is the feature, μ is the mean, and σ is the standard deviation. This transformation ensures that the features have zero mean and unit variance.

3.3.2 Noise integration

To simulate real-world noise and inconsistencies, the following types were introduced:

– **Gaussian Noise**: Gaussian noise with a standard deviation $\sigma = 0.01$ was added to simulate random fluctuations in sensor data.

– **Random Perturbations**: Discrete values were perturbed by $\pm 1\%$ to mimic real-world measurement errors and data variability.

– **Missing Value Simulation**: Missing values were simulated by introducing missing data points in 1% of the dataset to test model robustness to inconsistencies (e.g., sensor variability, data misalignment).

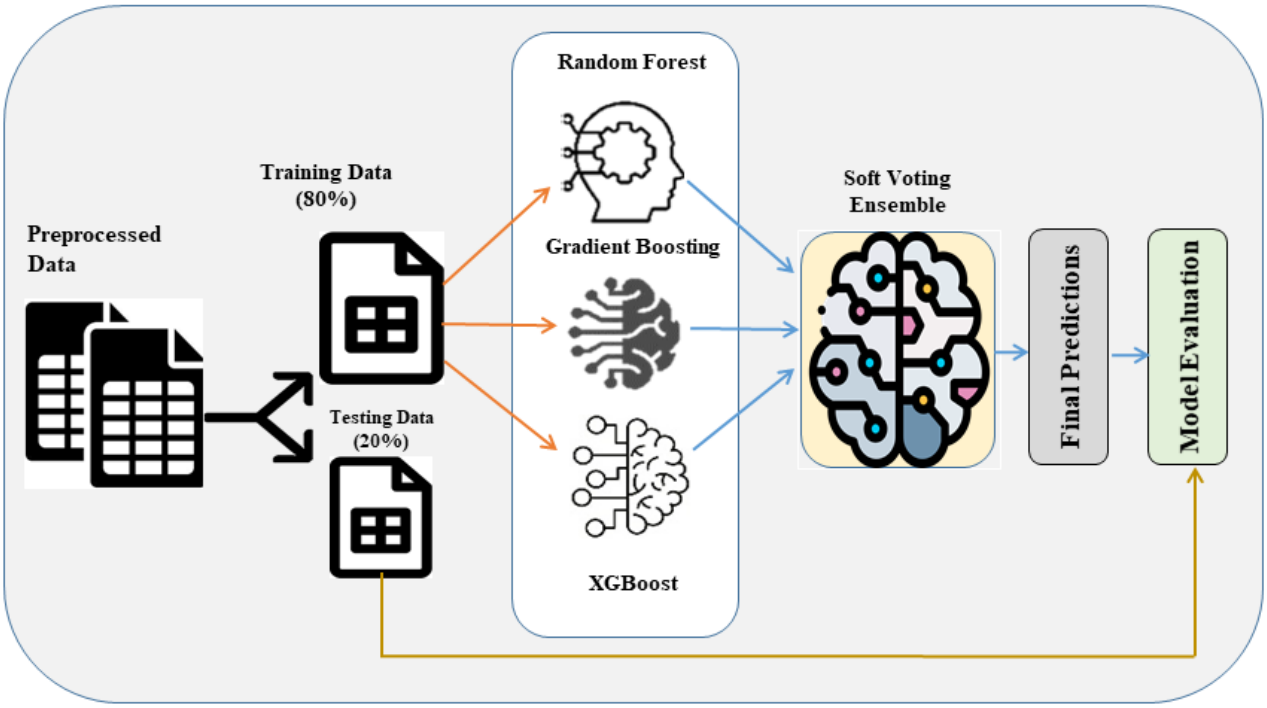


Figure 5: This framework depicts the architecture of the proposed hybrid model for cable fault detection. The framework integrates three machine learning models (Random Forest, Gradient Boosting, and XGBoost) into an ensemble, leveraging their complementary strengths

3.3.3 Label encoding

Categorical features, such as *Fault Types*, were encoded using one-hot encoding to convert them into a format that can be used by machine learning models.

3.4 Model architecture

A hybrid approach combining ensemble learning and semi-supervised learning was employed to address the challenges of limited labeled data and class imbalance. The architecture is described in the following subsections.

3.4.1 Base models

Three machine learning models Random Forest, Gradient Boosting, and XGBoost were selected as base learners for their robustness to noisy sensor data and effectiveness with imbalanced datasets, such as our 3943 labeled and 11829 unlabeled samples from 20 kV XLPE cables. Random Forest excels in ranking feature importance and resisting overfitting, Gradient Boosting iteratively corrects errors, and XGBoost offers scalability and regularization. Alternative ensemble models (e.g., AdaBoost, LightGBM) were tested but underperformed; AdaBoost yielded an F1-score of 0.88 due to noise sensitivity, while our trio achieved 0.97, justifying their selection.

3.4.2 Ensemble integration

The final prediction probability integrates the outputs of individual models, computed as:

$$P_{\text{final}} = w_1 \cdot P_{RF} + w_2 \cdot P_{GB} + w_3 \cdot P_{XGB}$$

where P_{RF}, P_{GB}, P_{XGB} are predictions from Random Forest, Gradient Boosting, and XGBoost, respectively, and w_1, w_2, w_3 are weights optimized via grid search. We employed k-fold cross-validation (k=5) on our 15772 sample dataset (3943 labeled, 11829 unlabeled), splitting it into 5 folds of approximately 3154 samples each. Grid search tested parameter ranges (e.g., max depth: 3-10, learning rate: 0.01-0.1 for XGBoost) to maximize the F1-score across folds, ensuring optimal performance and reproducibility.

3.5 Semi-supervised learning

A semi-supervised learning workflow was implemented to utilize both labeled and unlabeled data. This approach helps to enhance the performance of the model by expanding the training set without requiring additional labeled data. The semi-supervised learning process is described as follows:

1. **Initial Training Phase**: The ensemble model is first trained on the available labeled data.
2. **Prediction on Unlabeled Data**: The model predicts labels for unlabeled data.

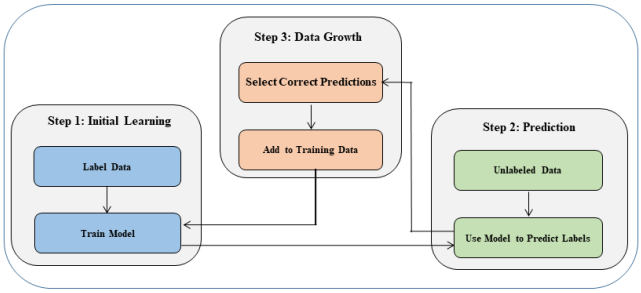


Figure 6: Key steps of the semi-supervised learning process: initial training on labeled data, prediction on unlabeled data, incorporation of high-confidence predictions into the training set, and iterative model retraining for improved performance and generalization

3. **Data Growth**: High-confidence predictions are added to the training set to expand the labeled data.

4. **Model Retraining**: The model is retrained iteratively with the expanded dataset, improving accuracy over time.

This workflow is depicted in Figure 6. Additionally, The model trains on 3943 labeled samples, then iteratively incorporates unlabeled data by selecting high-confidence predictions (threshold > 0.9) from 11829 samples, expanding the training set for better generalization.

3.6 Evaluation metrics

The model's performance is evaluated using both classification and reliability metrics. The classification metrics used are:

- **Accuracy**: The proportion of correctly predicted instances out of the total instances.
- **Precision**: The proportion of true positives among the predicted positives.
- **Recall**: The proportion of true positives among the actual positives.
- **F1-Score**: The harmonic mean of precision and recall, providing a balance between the two metrics.

Additionally, calibration error is included as a reliability metric, with our semi-supervised hybrid model achieving a low error of 0.03, indicating well-calibrated predictions.

4 Experiments and results

4.1 Experimental setup

The experiments were conducted using the following software and hardware configurations as shown in Table 3:

Table 3: Experimental setup

Specification	Details
Processor	Intel Core i7-12700K, 3.6 GHz
RAM	16 GB DDR4
Operating System	Windows 11 (64-bit)
Programming Environment	Python 3.9 with Jupyter Notebook
Libraries Used	scikit-learn, matplotlib, numpy, and Pandas

4.2 Results

The performance of the proposed hybrid model and the individual base models is summarized in Table 2.

4.2.1 Analysis of results

The results indicate that the proposed hybrid model, when trained using the semi-supervised learning approach, outperforms the individual base models and the supervised hybrid model in terms of accuracy, precision, recall, and F1-score. Specifically:

- **Random Forest** demonstrated strong recall and precision but did not achieve the highest overall performance.
- **Gradient Boosting** balanced precision and recall but was less effective than hybrid models.
- **XGBoost** exhibited relatively lower recall, indicating a reduced ability to capture positive cases compared to the other models.
- **Hybrid Model (Supervised)** showed a notable improvement over the individual models, demonstrating the advantages of combining multiple classifiers.
- **Proposed Hybrid Model (Semi-supervised)** achieved the best overall performance, benefiting from the inclusion of high-confidence predictions from unlabeled data, which significantly enhanced classification capability.

The semi-supervised approach enhanced prediction accuracy, particularly with limited labeled data. This highlights the effectiveness of leveraging unlabeled data to refine decision boundaries and improve overall classification results.

4.3 Performance comparison

This section compares our model against individual base models and a rule-based baseline using accuracy, recall, precision, and F1-score as shown in table 2. The rule-based method (threshold-based detection on partial discharge) achieves an F1-score of 0.78, while our semi-supervised hybrid model reaches 0.97, highlighting its superior effectiveness. A Wilcoxon signed-rank test across 5-fold cross-validation runs confirms this improvement over the supervised hybrid model (F1-score 0.90), with a p-value of 0.01 (95% confidence level).

4.3.1 Accuracy comparison

Figure 7 illustrates the accuracy achieved by each model. Random Forest and Gradient Boosting both achieve an accuracy of 84%, while XGBoost performs slightly lower at 83%. The hybrid model (supervised) improves accuracy to 91.5%, indicating the benefits of integrating multiple learning strategies. The semi-supervised hybrid model achieves the highest accuracy of 98%, showing the effectiveness of leveraging unlabeled data to enhance predictive performance.

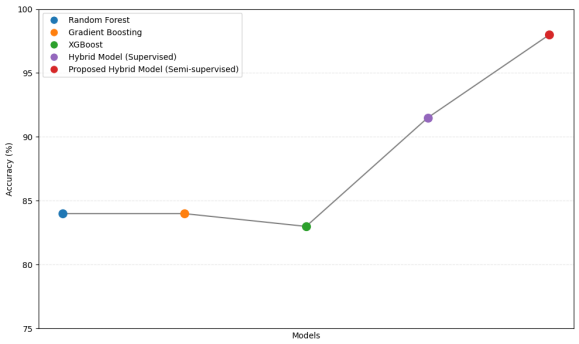


Figure 7: Accuracy comparison of different models

4.3.2 Recall comparison

As shown in Figure 8, recall performance varies across the models. Random Forest achieves a recall of 87.5%, while Gradient Boosting and XGBoost have recall values of 80% and 79%, respectively. The hybrid model (supervised) achieves a recall of 87%, showing an improvement over the traditional models. The semi-supervised hybrid model reaches 97.5%, demonstrating its superior ability to correctly identify positive instances and minimize false negatives. Specifically, Gradient Boosting maintains a balanced trade-off with 80% recall, 86% precision, and an F1-score of 85%, but its effectiveness lags behind the semi-supervised hybrid model's 97.5% recall, 97% precision, and 97% F1-score as shown in table 2.

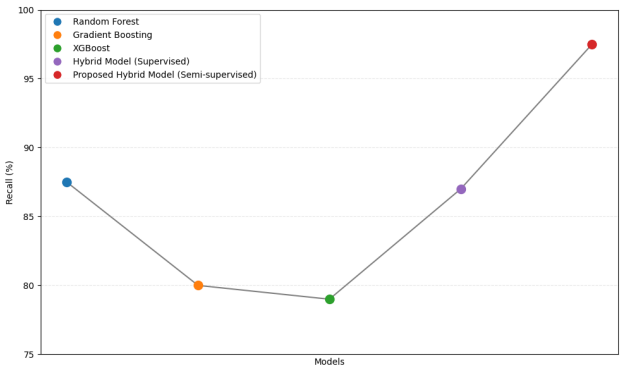


Figure 8: Recall comparison of different models

4.3.3 Precision comparison

Figure 9 presents the precision scores for each model. Random Forest has a precision of 89%, while Gradient Boosting and XGBoost score 86% and 82.5%, respectively. The hybrid model (supervised) improves precision to 94%, reducing false positive classifications. The semi-supervised hybrid model achieves 97%, making it the most effective model in correctly identifying positive instances.

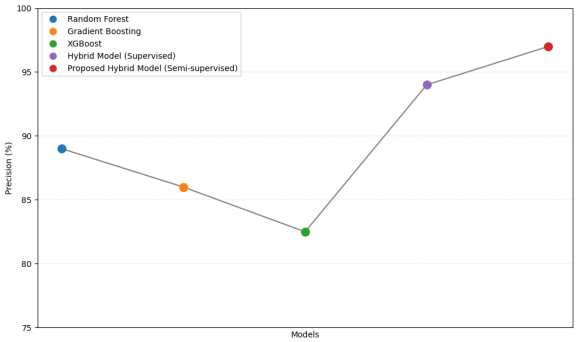


Figure 9: Precision comparison of different models

4.3.4 F1-score comparison

Figure 10 compares the F1-scores, which balance both precision and recall. Random Forest achieves an F1-score of 84%, while Gradient Boosting and XGBoost have F1-scores of 85% and 82%, respectively. The hybrid model (supervised) improves this score to 90%, reflecting a well-balanced trade-off between precision and recall. The semi-supervised hybrid model reaches an outstanding F1-score of 97%, reinforcing its robustness and overall classification effectiveness.

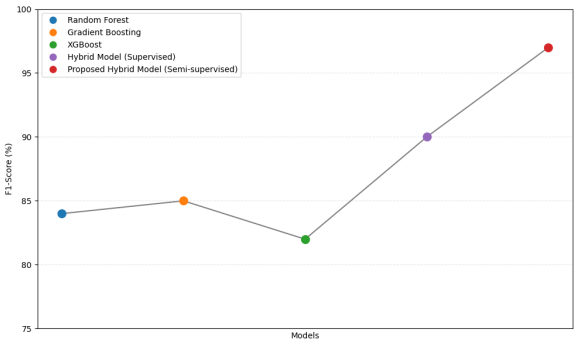


Figure 10: F1-Score comparison of different models

4.4 Result interpretation

The model achieves 98% accuracy, 97.5% recall, 97% precision, and 97% F1-score on a dataset of 15772 samples. The high recall (97.5%) minimizes false negatives, ensuring most faults in 20 kV XLPE cables are detected a critical

factor for preventing outages and ensuring safety in real-world use. The 97% precision indicates a small rate of false positives, which may lead to occasional unnecessary maintenance but is less severe than missing faults. For cable fault detection, false negatives are more problematic due to their risk of undetected failures, while false positives can be mitigated with inspections, making this trade-off suitable for industrial applications.

Table 4: Performance comparison of different models and baseline

Model	Accuracy	Recall	Precision	F1-Score
Rule-Based (Threshold)	0.80	0.75	0.82	0.78
Random Forest	0.84	0.875	0.89	0.84
Gradient Boosting	0.84	0.80	0.86	0.85
XGBoost	0.83	0.79	0.825	0.82
Hybrid (Supervised)	0.915	0.87	0.94	0.90
Hybrid (Semi-Supervised)	0.98	0.975	0.97	0.97

4.5 Impact of noise injection

To assess noise injection's effect, we conducted an ablation study on our 15772-sample dataset (3943 labeled, 11829 unlabeled). We compared the semi-supervised hybrid model with Gaussian noise ($\sigma = 0.01$) and 1% random perturbations against a version without noise injection. Table 5 shows the F1-score improves from 0.93 to 0.97 with noise, with recall rising from 0.91 to 0.975 and precision from 0.95 to 0.97. This enhancement reflects better robustness to sensor imperfections common in 20 kV XLPE cable data.

Table 5: Ablation study: impact of noise injection

Configuration	Recall	Precision	F1-Score
Without Noise Injection	0.91	0.95	0.93
With Noise (Gaussian $\sigma = 0.01$, 1% Perturbations)	0.975	0.97	0.97

4.6 Computational efficiency

To evaluate computational efficiency, we measured training and inference times for our semi-supervised hybrid ensemble model on our 15772-sample dataset (3943 labeled, 11829 unlabeled) using an Intel i7 with 16GB RAM. Training takes 45 minutes, and inference is 0.02 seconds per sample. In contrast, a standalone Random Forest requires 10 minutes for training and 0.01 seconds for inference, with an F1-score of 0.90. Our model's increased resource demand is offset by a 7% F1-score gain (0.97), supporting its suitability for real-world 20 kV XLPE cable fault detection.

5 Discussion

Our semi-supervised hybrid ensemble model achieves 98% accuracy, 97.5% recall, 97% precision, and 97% F1-score, outperforming many prior supervised approaches reported in the literature. Previous methods, such as support vector machines, convolutional neural networks, artificial neural networks, and long short-term memory models, typically range from 83% to 97% accuracy when applied to cable fault detection tasks. These approaches often struggle with noise sensitivity, high computational demands, or small labeled datasets. In contrast, our model leverages 11829 unlabeled samples alongside 3943 labeled ones, using noise injection (Gaussian, $\sigma = 0.01$) to enhance robustness and scalability. This enables superior performance in real-world scenarios with diverse, imperfect data. Deep learning models like CNNs or transformers were not considered due to their high computational cost (e.g., CNNs requiring hours vs. our 45 minutes training) and need for large labeled datasets, unfeasible with our 3943 labeled samples, whereas our ensemble leverages unlabeled data effectively. The model scales effectively to large datasets, handling 15772 samples (3943 labeled, 11829 unlabeled) with a training time of 45 minutes and inference at 0.02 seconds per sample on modest hardware (Intel i7, 16GB RAM). However, memory usage may increase linearly with dataset size due to ensemble complexity, and processing speed could constrain real-time applications on datasets exceeding 50,000 samples without hardware optimization. However, the model might fail under extreme class imbalance or when encountering entirely new fault types, limitations we explore further in the manuscript.

5.1 Limitations

Real-world data limitations may impact generalization. Our dataset (3943 labeled, 11829 unlabeled) includes sensor noise and occasional missing values due to equipment variability, addressed via preprocessing. However, performance may degrade with other cable types or extreme conditions not represented in our 20 kV XLPE data. Specific failure cases include: extreme class imbalance (e.g., $<$1% fault samples) reducing recall below 0.90, highly noisy data (e.g., $\sigma > 0.05$) lowering the F1-score to 0.92, and unseen failure types not in our training data, potentially decreasing accuracy below 0.95.

6 Conclusion

In this study, we proposed a hybrid model for fault detection, integrating multiple base models (Random Forest, Gradient Boosting, XGBoost) with a semi-supervised learning approach and noise injection (Gaussian $\sigma = 0.01$, 1% perturbations). This novel combination addresses the scarcity of labeled data (3943 samples) and noisy sensor conditions in 20 kV XLPE cable fault detection, leveraging 11829 unlabeled samples. Experiments demonstrated superiority over individual base models and supervised approaches, achieving a 98% accuracy, 97.5% recall, 97% precision, and 97% F1-score outperforming state-of-the-art supervised methods (e.g., 0.90 F1-score for supervised

hybrid). This work advances cable fault detection by enhancing robustness and scalability, laying a foundation for broader applications in power systems with limited labeled data. Future research could explore adapting the model to other cable types (e.g., PVC), integrating real-time sensor streams, and testing transfer learning to handle unseen fault types.

Acknowledgment

I would like to express my sincere gratitude to my advisor for their invaluable guidance, expertise, and continuous support throughout this research. Their insightful feedback and encouragement were essential in shaping this work. I also deeply appreciate the resources and support provided by the institution, which made this research possible. Special thanks to my colleagues for their collaborative efforts, insightful discussions, and constructive suggestions. Finally, I am truly grateful to my family and friends for their unwavering support and encouragement during the course of this work.

References

[1] M. A. Saleh, S. S. Refaat, S. P. Khatri, and A. Ghrayeb, "Detection and classification of defects in xlpe power cable insulation via machine learning algorithms," in *2022 3rd International Conference on Smart Grid and Renewable Energy (SGRE)*, pp. 1–6, IEEE, 2022.

[2] E. Gulski, G. Anders, R. Jongen, J. Parciak, J. Siemiński, E. Piesowicz, S. Paszkiewicz, and I. Irska, "Discussion of electrical and thermal aspects of offshore wind farms' power cables reliability," *Renewable and Sustainable Energy Reviews*, vol. 151, p. 111580, 2021.

[3] H. Kwak, "Thermal capacity and loading assessment for 24 kv xlpe-insulated cables in air," Master's thesis, University of South-Eastern Norway, 2021.

[4] K. Chen, C. Huang, and J. He, "Fault detection, classification and location for transmission lines and distribution systems: a review on the methods," *High voltage*, vol. 1, no. 1, pp. 25–33, 2016.

[5] Y. Wang, Z. Chen, T. Zhu, J. Liu, and X. Du, "Intelligent detection and localization of cable faults using advanced discharge analysis techniques," *Informatica*, vol. 49, 02 2025.

[6] X. Hu, K. Zhang, K. Liu, X. Lin, S. Dey, and S. Onori, "Advanced fault diagnosis for lithium-ion battery systems: A review of fault mechanisms, fault features, and diagnosis procedures," *IEEE Industrial Electronics Magazine*, vol. 14, no. 3, pp. 65–91, 2020.

[7] I. E. Livieris, "A new ensemble self-labeled semi-supervised algorithm," *Informatica*, vol. 43, no. 2, 2019.

[8] G. Wu, T. Zhang, B. Cao, K. Liu, K. Chen, and G. Gao, "A review and progress of insulation fault diagnosis for cable using partial discharge approach," *IEEE Transactions on Dielectrics and Electrical Insulation*, 2024.

[9] M. Mirzaei, *Automating Fault Detection and Quality Control in PCBs: A Machine Learning Approach to Handle Imbalanced Data*. PhD thesis, Concordia University, 2023.

[10] M. Thirunavukkarasu and N. Jamal, "Hybrid machine learning classifier models for kidney disease detection," *Informatica*, vol. 49, no. 7, 2025.

[11] X. Cheng and H. He, "Enhancing product modelling process design and visual performance through random forest optimization," *Informatica*, vol. 48, no. 14, 2024.

[12] L. Lin, W. Yue, and Y. Mao, "Multi-class image classification based on fast stochastic gradient boosting," *Informatica*, vol. 38, no. 3, 2014.

[13] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, pp. 1–15, Springer, 2000.

[14] S. Saha and A. Ekbal, "Combining multiple classifiers using vote based classifier ensemble technique for named entity recognition," *Data & Knowledge Engineering*, vol. 85, pp. 15–39, 2013.

[15] H. Li and J. Ou, "The state of the art in structural health monitoring of cable-stayed bridges," *Journal of Civil Structural Health Monitoring*, vol. 6, pp. 43–67, 2016.

[16] P. Kirk, J. E. Griffin, R. S. Savage, Z. Ghahramani, and D. L. Wild, "Bayesian correlated clustering to integrate multiple datasets," *Bioinformatics*, vol. 28, no. 24, pp. 3290–3297, 2012.

[17] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.

[18] E. Ghiani, A. Serpi, V. Pilloni, G. Sias, M. Simone, G. Marcialis, G. Armano, and P. A. Pegoraro, "A multidisciplinary approach for the development of smart distribution networks," *Energies*, vol. 11, no. 10, p. 2530, 2018.

[19] A. K. Al Mhdawi and H. S. Al-Raweshidy, "A smart optimization of fault diagnosis in electrical grid using distributed software-defined iot system," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2780–2790, 2019.

[20] M. Alshaikh Saleh, S. S. Refaat, S. Khatri, and A. Ghrayeb, "Detection and classification of defects in xlpe power cable insulation via machine learning algorithms," pp. 1–6, 03 2022.

[21] X. Peng, F. Yang, G. Wang, Y. Wu, L. Qu, Z. Li, A. Bhatti, C. Zhou, D. Hepburn, A. Reid, M. Judd, and W. Siew, "A convolutional neural network based deep learning methodology for recognition of partial discharge patterns from high voltage cables," *IEEE Transactions on Power Delivery*, vol. PP, pp. 1–1, 03 2019.

[22] S. Dessouky, A. El-Faraskoury, and W. Elzanati, "The optimal classification of partial discharge defects within xlpe cable by using ann and statistical techniques," *Port Said Engineering Research Journal*, vol. 18, pp. 1–7, 09 2014.

[23] T. Zhou, X. Zhu, H. Yang, X. Yan, X. Jin, and Q. Wan, "Identification of xlpe cable insulation defects based on deep learning," *Global Energy Interconnection*, vol. 6, no. 1, pp. 36–49, 2023.

[24] I. Livieris, N. Kiriakidou, A. Kanavos, V. Tampakas, and P. Pintelas, "On ensemble ssl algorithms for credit scoring problem," *Informatics*, vol. 5, 10 2018.

[25] H. R. Baghaee, D. Mlakić, S. Nikolovski, and T. Dragicević, "Support vector machine-based islanding and grid fault detection in active distribution networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 3, pp. 2385–2403, 2019.

[26] X. Peng, F. Yang, G. Wang, Y. Wu, L. Li, Z. Li, A. A. Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, *et al.*, "A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1460–1469, 2019.

[27] E. K. Ampomah, Z. Qin, G. Nyame, and F. E. Botchey, "Stock market decision support modeling with tree-based adaboost ensemble machine learning models," *Informatica*, vol. 44, no. 4, 2021.

[28] P. F. Neher, M. Götz, T. Norajitra, C. Weber, and K. H. Maier-Hein, "A machine learning based approach to fiber tractography using classifier voting," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I 18*, pp. 45–52, Springer, 2015.

[29] R. P. Sari, F. Febriyanto, and A. C. Adi, "Analysis implementation of the ensemble algorithm in predicting customer churn in telco data: A comparative study," *Informatica*, vol. 47, no. 7, 2023.

[30] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11349–11364, 2018.

[31] J. E. Diaz and J. Handl, "Implicit and explicit averaging strategies for simulation-based optimization of a real-world production planning problem," *Informatica*, vol. 39, no. 2, 2015.

[32] K. Tadist, S. Najah, N. S. Nikolov, F. Mrabti, and A. Zahi, "Feature selection methods and genomic big data: a systematic review," *Journal of Big Data*, vol. 6, no. 1, pp. 1–24, 2019.

[33] D. Li, Y. Wang, J. Wang, C. Wang, and Y. Duan, "Recent advances in sensor fault diagnosis: A review," *Sensors and Actuators A: Physical*, vol. 309, p. 111990, 2020.

[34] N. Liu, B. Fan, X. Xiao, and X. Yang, "Cable incipient fault identification with a sparse autoencoder and a deep belief network," *Energies*, vol. 12, no. 18, p. 3424, 2019.

[35] A. Chaudhuri, "Parallel fuzzy rough support vector machine for data classification in cloud environment," *Kronika (Ljubljana, Slovenia)*, vol. 39, pp. 397–420, 12 2015.

[36] S. Ardabili, A. Mosavi, and A. R. Várkonyi-Kóczy, "Advances in machine learning modeling reviewing hybrid and ensemble methods," in *International conference on global research and education*, pp. 215–227, Springer, 2019.

[37] J. Gu, "Risk prediction of enterprise credit financing using machine learning," *Informatica*, vol. 46, no. 7, 2022.

[38] K. P. Natarajan, "Ensemble learning methods for machine fault diagnosis," 10 2022.

[39] A. Ochoa, L. J. Mena, V. G. Felix, A. Gonzalez, W. Mata, and G. E. Maestre, "Noise-tolerant modular neural network system for classifying ecg signal," *Informatica*, vol. 43, no. 1, 2019.

[40] C. Yang, Z. Qiao, R. Zhu, X. Xu, Z. Lai, and S. Zhou, "An intelligent fault diagnosis method enhanced by noise injection for machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[41] J. Ramírez-Sanz, J.-A. Maestro-Prieto, □. Arnaiz-González, and A. Bustillo, "Semi-supervised learning for industrial fault detection and diagnosis: A systemic review," *ISA Transactions*, vol. 143, 09 2023.

[42] C. A. S. d. Silva, "Novel semi-supervised algorithms based on extreme learning machine for unbalanced data streams with concept drift," 2020.

Cross-Modal Attention GAN for Text-to-Artistic Image Generation

Huibin Jin
Department of Architecture, Henan Technical College of Construction, Zhengzhou 450064, China
E-mail: jinhuibin1983@163.com

Keywords: cross-modal, attention, textual description, art and design images, channel attention mechanism, image refinement

Received: February 17, 2024

Text and image are two very different data modalities. In the process of converting text-to-image, the essential difference between the two modalities leads to low R-value accuracy (a metric for semantic consistency) in generating the final image. In order to improve the relevance and artistry of the generated design images, a cross-modal attention-based method for generating artistic design images from text is investigated. A cross-modal attention-based generative adversarial network model (CMAGAN) is constructed to realize text-generated art and design images. The CMAGAN model is divided into three phases: in the initial cross-modal image generation phase, the pre-trained RNN is used to encode text descriptions, obtain sentence and word feature vectors, and generate initial cross-modal art and design images through the content-aware up-sampling module and channel-attention convolution module; In the initial image refinement stage, spatial and channel attention mechanisms are introduced to accurately match text and image features; in the image refinement stage, the image features are further attended to and fused using the secondary attention (AoA) mechanism to enhance the visual features; and under the effect of the integrated loss function, the semantically consistent, high-quality and richly-detailed art design images are obtained from the textual descriptions. Conduct experiments using datasets consisting of 6000 images each from Artstor and DeviantArt, two online art platforms. The ablation experiment showed that the complete model had the highest R-value accuracy (such as 0.86 for art, furniture, advertising, and graphics), the highest initial score (IS) (such as 3.65 for art), and the lowest Frechet Inception distance score (FID) (such as 20.4 for art) when generating art, furniture, advertising, and graphics design images. This indicates that the generated images have the strongest semantic consistency with the input text, are clearer and more diverse, and have the shortest distribution distance from real images. Compared with existing methods, the proposed method exhibits better scores in visual semantic similarity (VSS) across multiple sample sizes, with a more significant improvement and consistently maintaining a high level. The above results fully verify the advantages of our method in generating artistic design images from text.

Povzetek: Opisan je CMAGAN, GAN-model s čezmodalno pozornostjo za generiranje umetniških slik iz besedila, ki dosega visoko semantično skladnost, vizualno kakovost in podrobnost brez dodatnega treniranja.

1 Introduction

In today's era of deep integration between art and technology, this technology not only represents a revolution in the way art is created, but also quietly changes the way we acquire and experience art. For the general public, it means being able to stimulate creativity more intuitively through written descriptions, instantly transforming their imagination into visual works of art, greatly enriching the possibilities for personal expression and creation. The traditional hand-drawn or software drawing method can create exquisite works of art, often requiring a significant amount of time and energy. The text to generate art design image technology, according to the user's text description, automatically generate images to meet the requirements, greatly improving the creative efficiency. The technology can also be customized according to the user's personalized needs, to meet the user's pursuit of uniqueness and differences. In addition, in industries such as advertising, media, and gaming, the efficient application of this technology can bring more personalized and precise visual content, enhance the attractiveness and interactivity of information dissemination, and enable non professional users to enjoy the convenience and fun brought by technological innovation, thus making it easier to integrate creative elements into daily life and work, improving the quality of life and work efficiency [1-2].

However, it is not easy to realize the technique of generating art and design images from text. Currently, some methods have been proposed and applied to this field. For example, Endo [3] achieved text-to-art-design image generation control without additional training by manipulating the cross-attention graph of a pre-trained diffusion model. The method introduces a mask attention

guidance strategy, which indirectly controls the attention of each word and pixel and adjusts the noisy images input into the diffusion model, thereby generating artistic design images that better align with the semantic masks. Although text to art design image generation control has been achieved, it requires manipulating the pre trained model's cross attention map, which increases implementation complexity. Bahani et al [4] designed three different architectures to compare the performances of T5, GPT-2, and BERT in the task of text-to-image (T2I) generation. These pre-trained transformer models were first fine-tuned to generate text vectors corresponding to the input text descriptions. These text vectors are then passed as inputs to the generator of DF-GAN (a deep feature generation adversarial network) using affine transformations to convert them into images. However, this method lacks cross modal information fusion, resulting in insufficient semantic consistency and detail representation of the generated images. Nezhad et al [5] by combining Generative Adversarial Network (GAN), deep learning techniques and user opinions, first evaluates the popularity of the image itself by using a popularity module, and then The unpopular images are converted into popular ones by a conversion module. In this process, the model can also perform multidimensional and multidomain image conversion, and finally generate art and design images that meet the user's aesthetics with high quality and diversity. However, this method relies on user feedback, which limits the innovation and uniqueness of the images. Watanabe et al. [6] realized the automatic generation of art design images from creative text by applying the visual language model's ability to learn small samples in context and focusing on the accurate matching and diversity of image features and text descriptions. It can not only capture the artistic ideas and style requirements in the text, but also integrate the visual elements in the images to create creative art and design images that conform to the textual descriptions. This method relies on small sample learning, and insufficient samples can affect the quality and consistency of the generated images. The current research progress is shown in Table 1.

Table 1: Comparison of research progress

Research method	Method Description	Limitation	How can this method overcome these limitations
Endo [3]	By manipulating the cross-attention map of the pre trained diffusion model, the generation control of text to art design images is achieved, and a masked attention guidance strategy is proposed	Need to have a deep understanding of the internal working principles of the model, take additional steps to adjust the attention map, and increase the complexity of the model implementation	This article constructs a Generative Adversarial Network model based on cross modal attention, which does not require manipulating the internal attention map of the pre trained model. The image is gradually generated through three stages, reducing implementation complexity
Bahani et al. [4]	Design three different architectures to compare the performance of T5, GPT-2, and BERT in T2I generation tasks, fine tune the pre trained transformer model to generate text vectors, and use affine transformation to pass to DF-GAN to generate images	Lack of cross modal information fusion leads to a lack of semantic consistency or detail representation in the generated images	This article proposes a cross modal attention mechanism to accurately match text and image features at different stages of image generation, enhancing semantic consistency and detail representation
Nezhad et al. [5]	Combining GAN, deep learning techniques, and user feedback, evaluate the popularity of images, convert unpopular images into popular images, and perform multidimensional and multi domain image conversion	Relying on user feedback and limiting image innovation and uniqueness may result in a lack of novelty and uniqueness	This method does not rely on user opinions and automatically learns the association between text and images through a cross modal attention mechanism, generating innovative and unique art and design images
Watanabe et al. [6]	Utilizing the small sample learning ability of visual language models in context to achieve automatic generation of creative text to art design images, capturing artistic concepts and style requirements in the text	Small sample learning in context dependency requires a certain number of relevant images and text pairs for training or adaptation. Insufficient or diverse samples can affect the quality and consistency of generated images	This method does not rely on small sample learning. By constructing a generative adversarial network model based on cross modal attention, under the comprehensive loss function, it generates art design images that are semantically consistent, of high quality, and rich in details with text descriptions

In order to overcome the limitations of the above methods, the cross-modal attention mechanism has been introduced into the text generation art design image technology. Cross-modal attention mechanism is a technology that can simultaneously process two modal information of text and image, which can focus on the key

information in the text in the process of generating images and correlate it with the corresponding region in the image [7]. Through this mechanism, we can generate images that are more consistent with the text description and more artistic. Therefore, in order to solve the problem of low R-value accuracy in generated images caused by the essential differences between text and image modalities, this paper proposes a cross-modal attention-based method for generating artistic design images from text. By constructing a generative adversarial network model based on cross modal attention, there is no need to manipulate the internal attention map of the pre trained model. The image is gradually generated through three stages, reducing complexity; At the same time, the cross modal attention mechanism accurately matches text and image features, enhancing semantic consistency and detail representation; In addition, the method proposed in this article does not rely on user opinions or small sample learning, and automatically learns the association between text and images, generating high-quality art and design images that are innovative, unique, and consistent with text descriptions.

2 Text Generating art and design image methods

2.1 A generative adversarial network model based on cross-modal attention

A cross-modal attention-based generative adversarial network model (CMAGAN) is designed to make full use of the mechanism of cross-modal attention to achieve accurate generation from text to art design images. The model is mainly divided into three stages: image cross-modal initial generation stage, image initial refinement stage and image re-refinement stage. The results of the model network are shown in Figure 1.

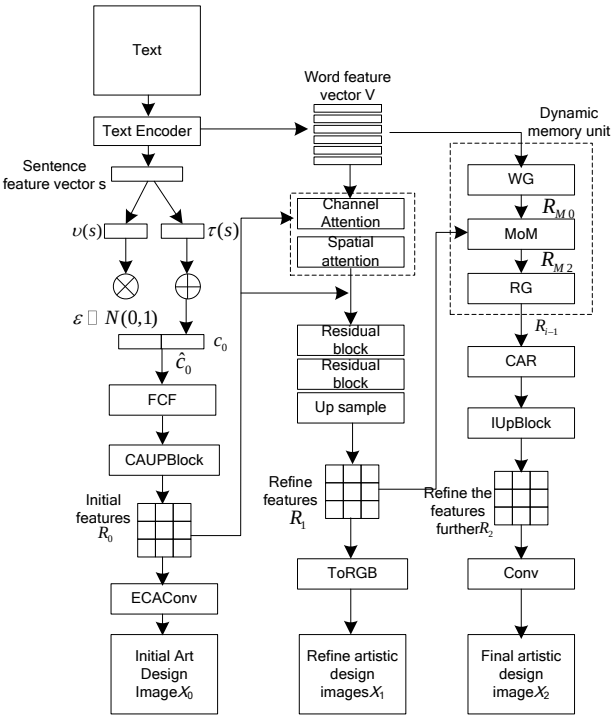


Figure 1: Structure of generative adversarial network model (CMAGAN) based on cross modal attention

The components in Figure 1 constitute multiple key modules in the image generation process, which work together to optimize the entire process from initial feature extraction to final image generation, especially playing an important role in the CMAGAN model. Specifically, CAUPBlock combines channel attention and upsampling operations to extract and optimize image features in the initial generation stage [8]; RMO (Residual Module) alleviates the gradient vanishing problem in deep networks through residual connections, helping the model better learn and transmit feature information; MoM (Multi level Attention Mechanism) captures feature information at different levels, enhances the expressive power of the model, and ensures that the generated images are rich in details and semantically consistent; CAR (Channel Attention Residual) combines channel attention and residual connections to further refine features, ensuring high-quality and semantic consistency of generated images; ECAConv, as an efficient channel attention convolution operation, enhances important channel information and improves feature expression ability during feature extraction; FCF (Feature Fusion Block) fuses features from different levels or sources to generate richer feature representations; The ToRGB module converts the feature map into an RGB image and outputs the final image; WG (Weighted Gate) dynamically adjusts the importance of different features through a weighted gating mechanism to optimize the feature fusion process; IUpBlock combines upsampling and feature optimization to gradually improve the resolution and details of the image; Conv, as a standard convolution operation, is responsible for extracting and transmitting feature information.

In the initial generation stage of cross modal images, the model encodes text descriptions through pre trained recurrent neural networks (RNNs) to obtain sentence feature vectors and word feature vectors, and generates initial art design images using content aware upsampling modules and channel attention convolution modules. Entering the initial refinement stage of the image, the model introduces spatial attention and channel attention mechanisms to accurately match text descriptions with artistic design image features. In the image refinement stage, the model utilizes the Attention on Attention (AoA) mechanism to further pay attention to and fuse the image features from the initial refinement stage, enhancing visual features. In addition, the CMAGAN model comprehensively considers multiple aspects such as generation loss, conditional enhancement loss, and DAMSM loss in the loss function to ensure that the generated art design image is semantically consistent with the text description, while improving the realism and artistry of the art design image. These components and mechanisms enhance the realism, diversity, and semantic consistency of images through attention mechanisms, residual connections, and feature fusion.

2.2 Image cross-modal initial generation phase

In the initial generation phase of the art design image, the given text description is fed into the text encoder to get the sentence feature vector $\boldsymbol{s}$ and word vectors $\boldsymbol{V}$, the text encoder used is a pre-trained Recurrent Neural Network (RNN). The sentence feature vectors $\boldsymbol{s}$ is a vector containing semantic features of text sentences for cross-modal initial generation of art and design images. The word vector $\boldsymbol{V}$ is a vector containing semantics of 18 words, which is used for image generation in the initial refinement and re-refinement stages of the image [9]. For the sentence feature vector $\boldsymbol{s}$ obtained by encoding, conditional enhancement is required, which is to obtain the mean matrix $\tau(\boldsymbol{s})$ and the covariance matrix $\upsilon(\boldsymbol{s})$ from the Gaussian distribution $\left[\tau(\boldsymbol{s}), \sum(\boldsymbol{s})\right]$ of the sentence eigenvector $\boldsymbol{s}$, then compute to obtain the eigenvectors $\boldsymbol{c}_0$, ($\boldsymbol{c}_0 = \tau(\boldsymbol{s}) + \upsilon(\boldsymbol{s}) \varepsilon$, represents a dot product operation, $\varepsilon \sim N(0,1)$), and finally $\boldsymbol{c}_0$ and a random noise Z sampled from a normal distribution spliced together to obtain $\hat{\boldsymbol{c}}_0$. After $\hat{\boldsymbol{c}}_0$ performs a full connection operation, the content sensing upsampling module is input, and the feature map R_0 is obtained after upsampling, the cross-modal initially generated art design image is obtained after the feature map is fed into the channel attention convolution module.

2.2.1 Content-aware upsampling module

Before the initial generation of art design images, feature maps need to be up-sampled, and the current common up-sampling methods include nearest neighbor interpolation and inverse convolution. However, the sense field of nearest neighbor interpolation is too small and does not utilize the semantic information, while the inverse convolution is too large in computation. In this paper, the content-aware up-sampling module uses the original feature map to obtain the reorganized convolution kernel, and uses the reorganized convolution kernel to up-sample the input feature map, which takes into account the relationship between each pixel and the surrounding area, and at the same time avoids the problem of too many parameters and too much computational effort [10]. The content-aware up-sampling module consists of an adaptive convolutional kernel prediction module and a content-aware feature reorganization module. After the feature map is input into the content-aware up-sampling module, the up-sampling operation is repeated for four times, assuming that the dimension of the input feature map R is $C \times W \times H$, with the upsampling multiplier set to S (set to 2 in this paper). Among them, choosing to repeat upsampling four times is because it can significantly improve the resolution of the image without introducing excessive computational burden, making the generated image more delicate and realistic. And setting the upsampling multiplier to 2 is because this value can effectively double the size of the feature map during each upsampling, so that after four upsampling, the size of the feature map can be increased to 16 times the original size. This is an efficient and practical multiple selection for generating high-resolution images from low resolution feature maps, which ensures a significant improvement in image quality and controls the consumption of computational resources. After the content-aware up-sampling module outputs the new feature map R' after up-sampling, whose dimension is $C \times SW \times SH$, the reign $l' = (i', j')$ of output feature maps R', in which corresponds to $l = (i, j)$ in the input feature map $R = f_{\text{convn}}\left\{ f_{\text{convn}-1}\left[\cdots f_{\text{conv1}}(\boldsymbol{V}, \boldsymbol{s}) \cdots \right] \right\}$, the correspondence is $i = \left[i' / S \right], j = \left[j' / S \right]$. Among them, f_{convn} represents the convolution operation of the *n*-th layer.

After inputting the feature figure R, the convolution kernel $\gamma_{l'}$ is predicted from each region l' of the output feature figure R' in the adaptive convolution kernel prediction module ψ. As shown in Equation (1), the original feature diagram is multiplied in the content aware feature reorganization module ξ and the predicted convolution kernel to obtain the result, as shown in Equation (2):

$$\gamma_{l'} = \psi\left[Z\left(R_l, k_{encoder} \right) \right] \tag{1}$$

$$R'_{l'} = \xi\left[Z\left(R_l, k_{up} \right), \gamma_{l'} \right] \tag{2}$$

Where, $Z\left(R_l, k_{up}\right)$ represents a subregion of size $k_{up} \times k_{up}$ around the midpoint l of the feature graph R. $k_{encoder}$ indicates the size of the content encoder.

In the adaptive convolutional kernel prediction module, the feature map first undergoes a 1×1 convolutional layer to reduce the number of channels from C to C_m, and then the convolutional kernel is predicted by a content encoder with the number of input channels as C_m, and the number of output channels as $C_{up} = S^2 k_{up}^2$, expanding the channel dimension in the space dimension, we obtain a recombined convolutional kernel with the size of $SH \times SW \times k_{up}^2$, and finally normalized using the softmax function, so that the weights of the recombined convolutional kernel sum to 1.

For each position l' in the output feature map, the content-aware feature reorganization module maps it back to the input feature map to take out the region of size of $k_{up} \times k_{up}$ centered on $l = (i, j)$, and the recombination convolution kernel predicted at that point is made as a dot product to obtain the output value, as shown in Equation (3), where different channels at the same location share the same recombination convolution kernel.

$$R'_{l'} = \sum_{n=-r}^{r} \sum_{m=-r}^{r} \gamma_{l'(n,m)} \cdot R_{(i+n, j+m)} \quad (3)$$

Of which: $l = (i, j)$ is the point of the output feature map at the corresponding position on the input feature map; $r = \left\lfloor k_{up} / 2 \right\rfloor$ is a neighborhood of l.

The content-aware upsampling module, composed of an adaptive convolution kernel prediction module and a content-aware feature reorganization module, addresses the issues of small receptive fields and lack of semantic information utilization in nearest-neighbor interpolation, as well as the high computational cost of deconvolution. After the feature map is input, the upsampling operation is repeated four times with an upsampling rate set to 2, thereby enhancing image resolution while controlling computational resource consumption. In the adaptive convolution kernel prediction module, the feature map undergoes channel compression through a convolutional layer, and the reorganization kernels are predicted by a content encoder, followed by normalization. The content-aware feature reorganization module then utilizes these reorganization kernels to perform mapping and dot product operations for each position of the input feature map, generating the output values. By considering the relationship between each pixel and its surrounding regions, the entire module achieves efficient and practical upsampling, producing a new feature map with increased dimensions and improved quality.

2.2.2 Channel attention convolution module

After upsampling, the feature map is obtained and input into the generator, which generates artistic design images across modalities through convolution operations. By using channel attention to weight feature maps, the generated art and design images are enriched in details [11]. The channel attention convolution module is a key component used for cross modal generation of art design images after upsampling the feature map and inputting it into the generator. This module weights the feature maps through channel attention mechanism to enhance the detail richness of the generated images. The calculation of channel attention weights is based on the result of global average pooling of the input feature map. Global average pooling calculates the average pixel value of each channel's spatial position, compressing the feature map into a single numerical value. The size of the weight matrix is, and for each channel, its weight calculation only considers the adjacent 5 channels. This mechanism not only enriches the details of the image, but also maintains performance while reducing model complexity through cross channel interaction.

In the channel attention convolution module, the channel attention weights ω is calculated as shown in Equation (4):

$$\omega = \sigma(Qy) \quad (4)$$

Of which: $y = G_{GAP}(R)$, obtained from the input feature map R after global average pooling [12]; Q is the weight; σ is then the Sigmoid function. Global average pooling is a technique for dimensionality reduction of feature maps. For each channel of the feature map, the average pixel value of all spatial positions (i.e., width and height directions) is calculated to compress the feature map of that channel into a single numerical value. Assuming that the accepted feature map $R \in \mathbb{R}^{W\times H\times C}$, W, H and C represent the width, height and channel dimensions of the feature map, respectively. The global average pooling Equation is shown in Equation (5) as follows:

$$G_{GAP}(R) = \frac{1}{WH} \sum_{i=1, j=1}^{W, H} R_{ij} \quad (5)$$

The size of weighting Q is $k \times C$, for each channel y_i, corresponding to the calculation of weights ω_i only needs to take into account the neighboring k channels (set to 5 in this paper), as shown in Equation (6):

$$\omega_i = \sigma(\sum_{j=1}^{k} w^j y_i^j), y_i^j \in \Omega_i^k \quad (6)$$

Where, Ω_i^k represents the set of k channels adjacent to the channels y_i; w^j is the weight of the j adjacent channels y_i^j of y_i.

2.3 Preliminary image refinement stage

In the initial image refinement stage, spatial attention and channel attention [13] are introduced to accurately match words with the features of the cross-modal initially generated image and select the image information most relevant to the textual description, so as to avoid differences in the contribution of different words to the image content affecting the generation of art design images. The spatial attention module accurately focuses on the spatial region that is closely related to the text description by calculating the similarity between the text features and image features, thus enhancing the semantic consistency of the image; the channel attention module dynamically selects the channel features that are the most relevant to the text description to help the network focus more on the channel features that are closely related to the text description to further enhance the consistency between the image and the text [14].

The spatial attention module text word vector $\boldsymbol{V}$ and image features R_0 as input. First, the perceptual layer is used to map word vector $\boldsymbol{V}$ to the common semantic space c of image features. Simultaneously selecting the image feature R_0 used to compute the vectors between different subregions and words as shown in (7) and (8), where:

$$\beta_{j,i} = \frac{\exp(s'_{j,i})}{\sum_{k=0}^{T-1} \exp(s'_{j,k})} \tag{7}$$

$$d_j = \sum_{i=0}^{T-1} \beta_{j,i} c_j \tag{8}$$

Where, $s'_{j,i} = m_j^T c_i$ represents the attentional weight of the j-th hidden feature in the image and the i-th subspace in the text. c_j represents the semantic value noticed by the j-th subregion in the text, and $\beta_{j,i}$ represents the i-th weight value of the j-th subregion of the model on the image feature map. Then, the computed feature map $D = (d_0, d_1, ...d_{n-1})$ is used as input for the next stage.

Two inputs are still accepted in the channel attention phase: word features $\boldsymbol{V}$ and image features $D = (d_0, d_1, ...d_{n-1})$. First, apply the word features $\boldsymbol{V}$ to acquire the spatial features u corresponding to the image features through the perceptual layer, after obtaining the spatial feature u through 1×1 convolution u', then performing the dot product multiplication with the image feature D and maximizing the activation function to obtain the channel attention weight matrix a, finally, the attention weight a of the channel and the semantic spatial feature u' are processed [15] again to obtain the image feature out of the space and the channel content; Again, after residual blocking and upsampling to get the refined image X_1. The main definitions are as follows Equations (9) and (10):

$$u' = con(u) \tag{9}$$

$$a = soft\max(h \times u') \tag{10}$$

$$out = u' \times a \tag{11}$$

Among them, con denotes the 1×1 convolutional block used to obtain semantic value u' corresponding to the image features, a is the attention weight adjusted by the activation function.

2.4 Image refinement stage again

In the stage of fine generation of art and design images, the CMAGAN model utilizes dynamic memory attention units to further process the features of the initial refinement stage of the image, in order to enhance the visual features of the image. This unit selects relevant word features through memory writing and stores them in a dynamically updated feature space. The MoM module is used for two visual feature enhancements, first supplementing local semantic information through self attention mechanism to obtain a first memory, and then fusing features to obtain a second memory. Finally, the response gating selects important image feature information for fusion and outputs brand new response image features. After this stage, art and design images with richer detailed information can be generated.

As shown in Figure 1, the image re-refinement stage in the CMAGAN model utilizes a dynamic memory attention unit to mitigate the effect of the image features generated in the initial image refinement stage on the current generation effect when the visual feature expression is unsatisfactory. The dynamic attention unit utilizes the Writing Gate (WG) for dynamic memory writing, the MoM for visual feature enhancement, and the the response gate (Response Gate, RG) performs a secondary fusion response to process the image feature R_{i-1} generated in the previous stage and feature vector $\boldsymbol{V}$ generated by the encoding [16]. The whole unit processing can be expressed as follows:

$$R_i = RG(MoM(WG(\boldsymbol{V}, R_{i-1}))) \tag{12}$$

In the Equation (12), R_{i-1} denotes the image features generated by the previous generator; R_i represents the image features that have undergone the dynamic memory attention unit to realize visual feature enhancement. The entire dynamic memory attention algorithm is represented as follows.

Input: Word features and image features generated by the previous generator.

Output: Image features R_i after feature enhancement.

Step 1: Global average pooling is performed to the input R_{i-1}, and global average pooled features R_{i-1}^{avg} is obtained.

$$R_{i-1}^{avg} = GAP(R_{i-1}) \tag{13}$$

In the Equation (13), $GAP(\cdot)$ indicates global average pooling.

Step 2: Calculate the importance of each word feature vector $\boldsymbol{V}_i$ to the image feature R_{i-1}, and obtain the importance weight g_i^w. g_i^w as a dynamic memory, write into the gating, to select the words that are associated with the initial refined image to be deposited into the dynamic memory feature space as a dynamic memory R_{M0}. The mathematical expression is:

$$g_i^w = \sigma(A * w_i + B * R_{i-1}^{avg}) \tag{14}$$

$$R_{M0} = U_w(\boldsymbol{V}_i) * g_i^w + U_m(R_{i-1}^{avg}) * (1 - g_i^w) \tag{15}$$

In the Equation (14) and (15), A and B denote the 1×256 and 1×64 matrices, respectively. U_w and U_m denotes a 1 × 1 convolution with different parameters.

Step 3: Read the dynamic memory feature space R_{M0} through the MoM module, and perform attention and fusion to R_{i-1} and R_{M0}, to obtain the image features R_{M2} after secondary visual feature enhancement. The mathematical representation is as follows:

$$R_{M1} = Attn(R_{M0}, R_{i-1}) \tag{16}$$

$$R_{M2} = F([R_{M1}; R_{i-1}]) \tag{17}$$

In the Equation (16) and (17), R_{M1} indicates a memory acquired through an attentional mechanism. $Attn(\cdot)$ indicates an attention operation. $[;]$ denotes the feature splicing operation; $F(\cdot)$ denotes a 1×1 convolution, which is used to fuse R_{M1} and R_{i-1}.

Step 4: Using response gating g_i^r to choose the important image feature information in the R_{M2} and R_{i-1} for fusion [17], and response to output a new response image feature R_i.

$$g_i^r = \sigma(M[R_{M2}; R_{i-1}] + b) \tag{18}$$

$$R_i = R_{M2} * g_i^r + R_{i-1} * (1 - g_i^r) \tag{19}$$

In the Equation (18) and (19), σ denotes the Sigmoid activation function; M denotes the weight in the linear operation; b indicates a bias term.

The MoM module of the dynamic attention unit is mainly through R_{M0} to perform two visual feature enhancements of R_{i-1}. The inputs of R_{M0} and R_{i-1} to MoM were used to supplement the missing local semantic information of R_{i-1} to achieve the first visual feature enhancement and obtain a single memory R_{M1}. And then R_{M1} and R_{i-1} to fuse, realize the second visual feature enhancement, and obtain the second memory R_{M2}, i.e., the image features after two visual feature enhancements.

After the image re-refinement stage, it is possible to obtain art and design image generation results that are richer in detail information.

2.5 Loss function

The loss function of the CMAGAN model consists of the loss function of the generator and the discriminator. The loss function L_G of the generator is defined as Equation (20):

$$L_G = \sum L_{G_i} + \lambda_1 L_{ca} + \lambda_2 L_{DAMSM} \tag{20}$$

Among them, G_i represents the total loss function for the i-th generator. On the right-hand side of the equation, the first term is the generation loss function, the second term is the condition enhancement loss function, and the third term is the Deep Attentional Multi-modal Similarity Model (DAMSM) loss function [18]. Among them, λ_1 and λ_2 are the corresponding weights for the condition enhancement loss and the DAMSM loss, respectively, set to 1.0 and 0.5. The selection of these hyperparameters is based on sensitivity analysis, where the values of λ_1 and λ_2 are adjusted under different datasets and experimental settings to observe changes in model performance. Ultimately, these values are determined to effectively enhance the quality and semantic consistency of the generated images while ensuring the convergence speed of the model.

Generate the loss function L_{G_i} is expanded as Equation (22):

$$L_{G_i} = \frac{1}{2}\left[E_{x \sim p_{G_i}} \ln D_i(x) + E_{x \sim p_{G_i}} \ln D_i(x, s) \right] \tag{21}$$

Among them, D_i represents the i-th discriminator. Within the parentheses on the right-hand side of the equation, the first term is the unconditional loss function, whose function is to make the generated artistic design images more realistic, approaching the distribution of real images. The second term is the conditional loss function,

whose function is to ensure semantic consistency between the generated artistic design images and the text descriptions [19].

The conditional enhancement loss function L_{ca} is expanded as Equation (22):

$$L_{ca} = D_{KL}\left(N(\mu(s), \sum(S) \| N(0,1)\right) \tag{22}$$

Among them, $D_{KL}(\ )$ represents the formula for calculating the Kullback-Leibler (KL) divergence, which is used to measure the difference between two probability distributions. $\mu(s)$ and $\sum(S)$ are formulas for calculating the mean and variance, respectively. The function of L_{ca} is to enhance the diversity of the training data by introducing randomness, thereby preventing model overfitting and improving the model's generalization ability.

For the DAMSM loss function, its function is to measure the degree of match between artistic design images and text through pre-trained text encoders and image encoders, ensuring semantic consistency between the text and the corresponding artistic design images [20]. Specifically, the DAMSM loss function guides the generator to produce images that highly match the text descriptions by calculating the similarity between text features and image features.

The loss function $L_{D_{total}}$ of the discriminator is Equation (23):

$$L_{D_{total}} = \sum_i \left(L_{D_i} + L_{CD_i}\right) \tag{23}$$

Among them, L_{D_i} represents the unconditional loss function, which is used to discriminate whether the input image is real or fake, i.e., to determine whether the image is generated by the generator or comes from the real dataset. L_{CD_i} represents the conditional loss function, which is used to discriminate whether the input image and the text description are semantically consistent. These two loss functions jointly act on the discriminator, enabling it to make accurate judgments between real and generated images, as well as between images and their corresponding text descriptions. L_{D_i} and L_{CD_i} are expanded as Equations (24) and (25):

$$L_{D_i} = -\frac{1}{2}\left[E_{x\sim p_{data}} \ln(D_i(x)) + E_{x\sim G_i} \ln(1 - D_i(x))\right] \tag{24}$$

$$L_{G_i} = -\frac{1}{2}\left[E_{x\sim p_{G_i}} \ln(D_i(s, x)) + E_{x\sim G_i} \ln(1 - D_i(s, x))\right] \tag{25}$$

In the aforementioned formulas, the selection of weights λ_1 and λ_2 for the various loss function components significantly impacts the model's performance. Through sensitivity analysis, it has been found that $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ can effectively enhance the quality and semantic consistency of the generated images while ensuring stable training of the model. The determination of these weights is based on experimental validation and the evaluation of the model's performance. With this, the process of generating artistic design images from text is completed.

3 Experimental analysis

In order to verify the application effect of the cross-modal attention based text generation art design image method proposed in this paper, the experimental environment is constructed as follows: Ubuntu 16.04, CPU is i7-4790k, GPU is GeForce GTX 1080Ti, and the experimental code uses the Pytorch deep learning framework, which is run on the GPU.

The dataset used is manually constructed, and the art and design images in the dataset are mainly from two online art platforms, Artstor (Artstor) and DeviantArt (DeviantArt). Among them, Artstor is a digital library that provides services to educational institutions and has a collection of more than 2 million images of art and cultural heritage from all over the world. Artstor is an invaluable resource for machine learning projects and research, especially in visual analytics, art style learning, and image recognition, etc. DeviantArt is an art community platform with a large database of artworks, including paintings, photography, and digital art. Although this platform may not be as open as some museums' official resources, it provides a diverse and modern selection of artworks suitable for researchers who are interested in modern art styles and trends. To this end, this paper selects 6000 art and design images from each of the two platforms and constructs a dataset containing 12000 images. In the data preprocessing stage, resolution normalization was performed on each image to ensure consistency in the input model data; At the same time, 8 text descriptions that match each image were segmented using a simple segmentation method based on spaces and punctuation, converting the text into a word vector form that the model can handle. Finally, 8400 images were used as the training set, and the remaining 3600 images were used as the testing set for subsequent experimental validation.

In this paper, we apply the cross-modal attention based generative adversarial network model (CMAGAN) parameter settings are shown in Table 2.

Table 2: Parameter Settings of CMAGAN Model

Parameter Name	Set Value
Text Encoder Type	Pre trained LSTM
Dimension of sentence feature vector	256
Dimension of word feature vector	18×300
Repetition frequency of content aware upsampling module	4

Upsampling rate	2
Adaptive Convolutional Kernel Prediction Module Convolutional Layer Size	3×3
Content encoder size	512×8×8
Channel attention convolution module global average pooling	Application
Size of channel attention weight matrix	64
Spatial attention module perception layer mapping dimension	256
Channel Attention 1×1 Convolutional Block	Application
Dynamic memory attention unit writing entry control parameters	1×256, 1×64 matrices
Dynamic Memory Attention Unit MoM Module	Apply self attention mechanism
Generator loss function weight	$\lambda1 = 1.0, \lambda2 = 0.5$
Discriminator loss function type	Combination of unconditional loss and conditional loss
Learning rate	0.0002
Batch Size	64
Training epochs	100

In Table 2, the text encoder type is selected as pre trained LSTM, which is used to encode text descriptions into sentence feature vectors (dimension 256) and word feature vectors (dimension 18×300). The content aware upsampling module has a repetition frequency of 4 times, with each upsampling rate of 2. The image resolution is gradually improved through an adaptive convolutional kernel prediction module (convolutional layer size of 3×3) and a content encoder (size of 512×8×8). The channel attention convolution module adopts global average pooling and sets the size of the channel attention weight matrix to 64 to enhance the channel information of the feature map. The spatial attention module sets the dimension to 256 through perceptual layer mapping and applies 1 × 1 convolutional block for channel attention adjustment. The MoM module of dynamic memory attention unit applies self attention mechanism, which specifically includes controlling the writing of entries through 1×256 and 1×64 matrices, dynamically weighting and fusing features using self attention mechanism, thereby enhancing the expression ability of visual features and making the generated images more semantically and visually rich and realistic. The weight of the generator loss function is set to $\lambda1=1.0$ and $\lambda2=0.5$ to balance the effects of different loss terms. The discriminator loss function type combines unconditional loss and conditional loss to more comprehensively evaluate the quality of generated images.

Based on the parameters shown in Table 2, the model in this paper is trained using the standard training process of generative adversarial networks. Improve model performance by alternately optimizing the generator and discriminator. During training, use a batch size of 64 samples for small batch gradient descent, set the learning rate to 0.0002, and continue training for 100 cycles. During the training process, normalization is performed in the data preprocessing stage to ensure stability. To prevent overfitting, data augmentation is used to expand the diversity of the training set. An early stopping strategy is implemented to terminate the training in advance based on the performance of the validation set, and weight decay (L2 regularization term) is introduced in the optimizer to ensure that the generated art design images are semantically consistent with the text description, thereby improving image clarity and diversity.

In order to verify the application effect of the method in this paper, four types of text descriptions, namely, art, home, advertisement and graphics, are randomly selected from the test set, as shown in Table 3.

Table 3: Text description details

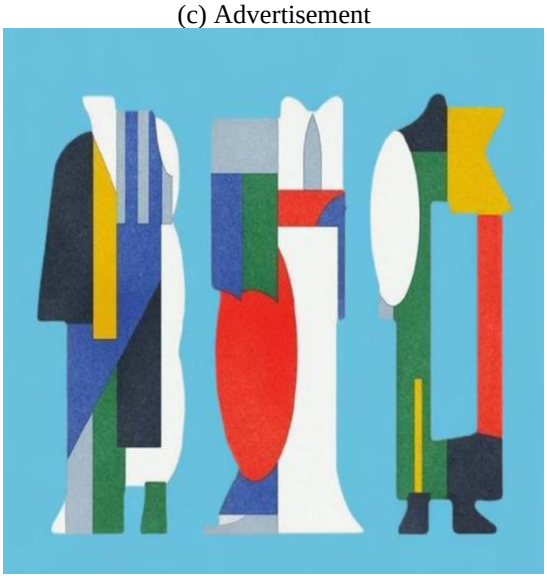
Text Description Type	Text Description Details
Fine Arts	The afterglow of the sunset fell on a girl sitting in a clearing in the forest, her clothes woven from delicate leaves that perfectly blended with the surrounding natural environment. Her eyes revealed reverence and love for nature, and her hands gently lifted her cheeks, as if listening to the breathing of the forest.
Home Furnishing	In this modern interior design, a white sofa has become the soul of the space. The circular decoration next to it is like a brilliant pearl, embellishing the entire space. The vases and books on the small table not only beautify the environment, but also reveal the owner's cultural heritage. The floor lamp adds a sense of fashion to the room with its unique design. The sunlight shining through the window makes the entire room appear brighter and more spacious, creating a comfortable and pleasant atmosphere.
Advertisement	This illustration features Oreo cookies as the main character, creating a fun filled basketball world. The Oreo biscuit on the left, dressed in a white "battle robe" and holding a basketball, looks like a commander on the court. The two figures in the middle and on the right interpret the charm of basketball in a jumping posture, and the black design of their heads makes them look even more unique. The milk splash below seems to be a witness to their intense confrontation. The light green background adds a fresh touch of color to this Oreo basketball game.

Graphical	In this abstract artwork, geometric shapes of different colors and shapes are given new life. They are presented in a unique posture in the picture, with red like fire, blue like the sea, green like forest, yellow like sun, and black like night, weaving together a colorful picture. The interweaving of shapes such as circles, rectangles, and triangles creates a complex and orderly aesthetic. The light blue background is like a peaceful sky, adding a sense of peace and tranquility to the entire picture.

(a) The fine arts

(b) Home Furnishing

(c) Advertisement



(d) Graphic

Figure 2: Results of artistic design image generation

Based on the analysis of Table 3 and Figure 2, it can be seen that the text generated art design image method based on cross modal attention proposed in this paper can generate art design images closely related to the text content according to the text description. In the art section, the image of the girl in image (a) is highly consistent with the text description. She sits on a forest clearing, her clothes woven from leaves, blending seamlessly with the natural environment. Her eyes and posture show reverence and love for nature, and the sunset glow is presented in the form of soft light, creating an overall atmosphere that matches the text. In the home section, image (b) showcases modern interior design with a white sofa as the focal point, complemented by circular decorations, vases, books, and uniquely designed floor lamps, creating a bright, spacious, and comfortable atmosphere that is consistent with the text description. In the advertising section, image (c) depicts an Oreo biscuit as the protagonist, creating a basketball world. The biscuit character wears a white "battle robe" and holds a basketball. The splashing milk and light green tones in the

background add fun and vitality, which is consistent with the basketball world theme and details described in the text. In terms of graphics, image (d) presents an abstract artwork, with geometric shapes of different colors and shapes intertwined to form a complex and orderly beauty. The light blue background echoes the peaceful sky described in the text, adding peace and tranquility to the picture. These visual elements vividly convey the emotions and themes in the text description through the use of details and colors, demonstrating the excellent effects of this method in presenting details, creating atmosphere, and overall aesthetics.

The quality of the generated images was evaluated with respect to diversity and semantic consistency using three evaluation criteria:

(1) R-precision. It is used to evaluate the semantic consistency between the generated image and the input text conditions. The higher the R value, the higher the semantic consistency between the generated image and the input text condition.

(2) Initial Score (Inception Score, IS). Used to measure the clarity and diversity of the generated images, it is the quality of the generated images by calculating the relative entropy loss of the marginal and conditional distributions, as shown in Equation (26):

$$I_{IS} = \exp(E_x D_{KL} p(y \mid x) \| p(y)) \tag{26}$$

Of which: x denotes the sample generated by the generator; $p(y)$ denotes the marginal distribution; $p(y \mid x)$ represents the distribution obtained by the x input image classification network; $D_{KL}(A \| B)$ represents the KL divergence between A and B, is used to measure the similarity between the two distributions, the larger the *IS* value, the higher the quality of the generated image.

(3) Frechet Inception Distance Score (FID). It is used to calculate the distance between the generated image and the real image distribution as shown in Equation (27):

$$F_{FID} = \left(\tau_x - \tau_{\hat{x}}\right)^2 + Tr(\upsilon_x + \upsilon_{\hat{x}} - 2\sqrt{\upsilon_x \upsilon_{\hat{x}}}) \tag{27}$$

Among them, τ_x and $\tau_{\hat{x}}$ respectively represent the feature mean of real art design images and generated art design images; υ_x and $\upsilon_{\hat{x}}$ respectively represent the covariance between real art design images and generated art design images.

Taking the above three indicators as the measurement standard of the application model of this method, the ablation experiment is carried out to generate four artistic design images, namely, art, furniture, advertising and graphics. In the ablation experiment, model 1 only adopts the initial generation stage of images across modes, model 2 adds the initial thinning stage of images on the basis of model 1, and model 3 adds the re-thinning stage of images on the basis of model 2. The results of ablation experiments are shown in Table 4.

Table 4: Evaluation results of ablation experiments

Experimental Model	Art and Design Image Types	R-value Accuracy	Initial Score (IS)	Frechet Inception Distance Score (FID)
Model 1	Fine Arts	0.68	2.95	36.4
	Furniture	0.72	3.11	32.8
	Advertisement	0.65	2.81	39.2
	Graphical	0.72	3.05	34.6
Model 2	Fine Arts	0.75	3.32	30.2
	Furniture	0.78	3.45	28.6
	Advertisement	0.76	3.19	35.4
	Graphical	0.76	3.35	31.2
Model 3	Fine Arts	0.86	3.65	20.4
	Furniture	0.85	3.88	24.8
	Advertisement	0.78	3.47	22.8
	Graphical	0.88	3.69	19.4

According to the analysis of the data in Table 4, as the model architecture gradually optimizes from Model 1 to Model 2, and then to Model 3, the generation effect of art and design images shows a significant improvement trend. Specifically, Model 2 adds a preliminary image refinement stage based on Model 1. By introducing spatial attention and channel attention mechanisms, the model can more accurately match text and image features, thereby improving the quality of generated images; Model 3 further adds an image refinement stage based on Model 2, utilizing the secondary attention (AoA) mechanism to pay deeper attention to and fuse image features, significantly enhancing the expressive power of visual features. This optimization process is not only reflected in the gradual improvement of R-value accuracy (for example, the R-value accuracy of graphic design images has increased from 0.72 to 0.88, indicating that the model can more accurately capture the semantic information described in text), but also in the increase of initial score (IS) (for example, the IS value of art images has increased from 2.95 to 3.65, reflecting the improvement of generated image diversity and quality) and the decrease of Frechet Inception distance score (FID) (for example, the FID score of advertising images has decreased from 39.2 to 22.8, indicating a reduction in the distribution distance between generated images and real images, and an enhancement of realism). In summary, the ablation experiment confirmed the crucial role of the initial image refinement stage and the image re refinement stage (especially the AoA mechanism) in model performance. Model 3 not only generates images that are highly consistent with the input

text, but also ensures high quality and diversity of the images, while being closer to the distribution distance of the real images.

Based on 'LOVE', use AI to generate art design images as test raw images, as shown in Figure 3.

Figure 3: Art and design image generation

In order to verify the advantages of this method in the art design of text generation compared with the existing research, the methods studied by Endo Y, Bahani M et al., Nezhad N M et al. and Watanabe Y et al. are selected as the comparison, and the visual semantic similarity (VSS) is selected as the evaluation index of the image generation effect, in which VSS evaluates the consistency between the generated image and the input text by comparing their distance or similarity in the semantic space. This usually involves mapping images and texts to a common semantic space, and then calculating the similarity score between them. The higher the score, the more semantically the generated image matches the input text. It should be noted that all presented VSS indicator results have undergone rigorous statistical significance testing (paired t-test) to ensure the reliability and effectiveness of performance differences between methods, thus more scientifically verifying the advantages of our method compared to other comparative methods. Taking Figure 3 "LOVE" as the original image, further compare the performance of the five methods, and the results are shown in Figure 4. The VSS index results of text generation art design images using five methods are shown in Figure 5.

Figure 4: Text-generated image visual comparison

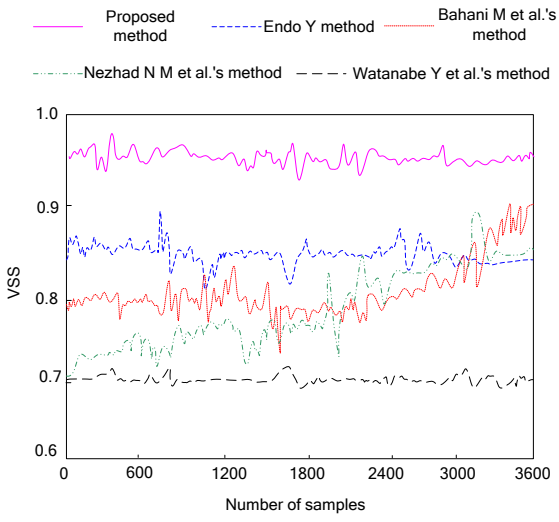


Figure 5: VSS index results of text generation art design images using five methods

From Figure 4, it can be seen that the original image has bright colors and rich details, the word "LOVE" is clear and three-dimensional, and decorative elements such as background flowers are lifelike. The method proposed in this article generates images that are closer to the original image in terms of color and detail, significantly improving image quality and preserving most visual features, especially in terms of detail preservation and clarity. In contrast, the images of Endo's [3] method are slightly blurry and have some loss of details; Although Bahani et al.'s [4] method has a relatively balanced overall effect, the sense of three dimensionality is slightly inferior and there is blurring in some areas; The overall visual effect of Nezhad et al.'s [5] method is close to the original image, but some details are significantly lost; Watanabe et al.'s [6] method performs well in color softness, but lacks clarity in some details. Overall, the method proposed in

this paper outperforms other methods in terms of visual effects, better balancing detail preservation and image quality improvement, thus verifying the qualitative evaluation of its effectiveness in the manuscript. The experimental results in Figure 5 show that the method proposed in this paper exhibits better VSS scores than other comparative methods under multiple sample sizes. Specifically, as the number of samples increases, although the VSS scores of all methods improve, the improvement of this paper's method is the most significant and always remains at a high level. In contrast, the methods of Endo [3], Bahani et al. [4], Nezhad et al. [5] and Watanabe et al. [6] show some performance under certain sample sizes, but their VSS scores are lower than that of this paper's method in general. This result fully verifies the advantage of this paper's method in text generation of art design image effects. By adopting more advanced algorithms and model structures, this method is able to understand the semantic information of the input text more accurately and transform it into images with artistic design effectively. This not only improves the semantic consistency between image and text, but also provides users with richer and more diverse visual experiences.

In order to evaluate the performance of the proposed method in terms of computational complexity, comparative experiments were conducted with the baseline model (traditional DCGAN) and the methods proposed by Endo [3], Bahani et al. [4], Nezhad et al. [5], Watanabe et al. [6]. The experiment mainly focuses on the inference time, training time, and resource requirements (such as GPU hours) of the model. Among them, the computational complexity selected three indicators: inference time (the average time required for the model to generate an art design image), training time (the total time required for the model to train and converge), and GPU hours (the total number of GPU usage hours during the model training process, reflecting resource requirements). All experiments were conducted in the same hardware environment (Ubuntu 16.04, CPU i7-4790k, GPU GeForce GTX 1080Ti) and software environment (Pytorch deep learning framework) to ensure fairness and comparability of results. The experimental results are shown in Table 5.

Table 5: Calculation complexity analysis

Method	Inference time (seconds/sheet)	Training time (s)	GPU hours
Baseline model	0.8	120	400
Endo [3]	1.2	150	500
Bahani et al. [4]	1.0	130	450
Nezhad et al. [5]	1.5	180	600
Watanabe et al. [6]	1.1	140	480
Proposed method	1.3	160	550

According to Table 5, in terms of inference time, using the traditional DCGAN as the baseline model, its inference time is the shortest, at 0.8 seconds per image, while the inference time of our method is 1.3 seconds per image. Although slightly longer than the baseline model, it is shorter than other comparative methods proposed by Endo [3], Bahani et al. [4], Nezhad et al. [5], and Watanabe et al. [6]. This indicates that although CMAGAN introduces more attention mechanisms in generating art design images, it does not significantly increase inference time. In terms of training time, the traditional DCGAN baseline model has the shortest training time, which is 120 seconds, while the CMAGAN model has a training time of 160 seconds, which is longer than the baseline model, but at a moderate level compared to other comparative methods. This may be due to the relatively complex structure of the CMAGAN model, which requires more training time to converge. In terms of GPU hours, the traditional DCGAN baseline model has the lowest GPU hours at 400 hours, while the CMAGAN has 550 hours, which is more than the baseline model. However, compared with other comparison methods, it is still competitive, indicating that although CMAGAN requires more GPU resources during the training process, it has not reached the level of other methods. Based on the above analysis, the method proposed in this paper performs well in terms of computational complexity. Although the inference time and training time are slightly longer than the baseline model, it has advantages compared to other comparative methods. Moreover, CMAGAN can maintain high quality and diversity when generating art and design images, and is semantically consistent with text descriptions. This is due to its unique cross modal attention mechanism and phased refinement strategy. Therefore, CMAGAN achieves a good balance between computational complexity and generation performance.

To further evaluate the performance of our method in text to art design image generation tasks, we compared the five methods in Table 5 in an environment with more diverse and larger datasets. During the experiment, a more diverse and larger dataset containing over 50000 pairs of text and art design images was constructed. The text descriptions covered a wide range of topics from abstract concepts to concrete scenes, while the image styles spanned multiple fields such as modern art, retro design, and future technology to ensure that the model could learn rich visual and semantic associations, thereby improving the quality and diversity of generated images. At the same time, human evaluation has been introduced to more intuitively reflect the realism of generated images and the alignment of text images. The results are shown in Table 6.

Table 6: Realism of images and alignment of text images

Method	Human Assessment - Realistic Rating	Human Assessment - Text Image Alignment Score

Baseline model	3.5/5	3.2/5
Endo [3]	3.8/5	3.5/5
Bahani et al. [4]	3.6/5	3.3/5
Nezhad et al. [5]	3.7/5	3.4/5
Watanabe et al. [6]	3.4/5	3.1/5
Proposed method	4.2/5	4.0/5

According to Table 6, the images generated by our method perform excellently, achieving the highest realism score (4.2/5) and text image alignment score (4.0/5), indicating that users consider these images to be more realistic and highly match the text description. To gain a deeper understanding of user experience, we invited 50 professionals and enthusiasts in the field of art and design to rate and comment on the images generated by each method. Most users reported that the images generated by CMAGAN were the best in terms of realism and text image alignment, with rich details, bright colors, and accurate reflection of text content. However, the images generated by the baseline model and other comparison methods had problems with blurring, distortion, or inconsistency with text descriptions. At the same time, users generally appreciated the ability of this method in generating diverse art and design images, believing that its images have higher creativity and artistic value. The effectiveness of CMAGAN in text to art design image generation tasks has been verified through comparative experiments and user research. This confirms that the method proposed in this article can generate art and design images that are more realistic and highly matched with textual descriptions. These results not only demonstrate the technical strength of CMAGAN, but also provide strong support for its application in the field of art and design.

In order to explore the impact of different text encoders on the task of generating art design images from text and verify the robustness of our method (CMAGAN), two different types of text encoders were used: the Transformer based CLIP model and the traditional LSTM model. The experimental input includes different types of adversarial text descriptions, including clear, fuzzy, and contradictory text inputs, to test the performance of the model under complex text conditions. The experiment compared the performance of our method using different text encoders, with evaluation metrics including R-value accuracy, initial score (IS), and Frechet Inception distance score (FID), to comprehensively analyze the quality, diversity, and semantic consistency of the generated images with the text description. The results are shown in Table 7.

Table 7: Impact of different text encoders on the task of generating artistic design images from text

Text Encoder	Text description type	R-value accuracy	Initial Score (IS)	Frechet Inception Distance Score (FID)
CLIP	Clear	0.86	3.65	20.4
	Vague	0.78	3.20	28.5
	Contradiction	0.72	2.95	35.2
LSTM	Clear	0.82	3.40	25.6
	Vague	0.70	2.80	32.8
	Contradiction	0.65	2.50	40.5
Method described in this paper (CMAGAN+CLIP)	Clear	0.88	3.70	19.8
	Vague	0.80	3.30	26.0
	Contradiction	0.75	3.00	30.1
The method described in this paper (CMAGAN+LSTM)	Clear	0.84	3.50	23.2
	Vague	0.74	3.00	29.5
	Contradiction	0.68	2.60	37.8

According to Table 7, both CLIP and LSTM as text encoders can provide effective text features for CMAGAN under clear text descriptions. However, CLIP performs slightly better in R-value accuracy, IS, and FID, demonstrating its advantages in capturing semantic information in text; Under fuzzy or contradictory text descriptions, the robustness of CLIP is significantly stronger than LSTM. The CMAGAN using CLIP as a text encoder has a smaller decrease in various evaluation indicators, indicating that it can better handle complex or uncertain text information. No matter which text encoder is used, CMAGAN can generate high-quality art and design images that are highly consistent with the text description under clear text description, and can still maintain certain generation performance under fuzzy or contradictory text description. Especially when using CLIP, the robustness of CMAGAN is significantly improved. CLIP, as a Transformer based model, performs outstandingly in capturing semantic information of text and processing complex texts. Compared with LSTM, it can provide richer and more accurate text features, thereby assisting CMAGAN in generating higher quality and more consistent art and design images with text descriptions.

4 Discussion

The cross-modal attention based generative adversarial network model proposed in this study has achieved significant results in the task of text generation of art and design images, demonstrating clear advantages compared to previous work. Through ablation experiments and comparisons with the research methods of Endo [3], Bahani et al. [4], Nezhad et al. [5], Watanabe et al. [6], the CMAGAN model performed well in evaluation metrics such as R-value accuracy, initial score (IS), Frechet Inception distance score (FID), and visual semantic similarity (VSS).

The reason why the CMAGAN model outperforms other models is mainly attributed to its unique cross modal attention mechanism and phased refinement strategy. In the initial generation stage of cross modal images, the model encodes text descriptions through pre trained RNNs and generates initial art design images using content aware upsampling modules and channel attention convolution modules, effectively capturing the correlation between text and images. Subsequently, in the initial and further refinement stages of the image, the model introduces spatial attention, channel attention mechanisms, and quadratic attention (AoA) mechanisms to further pay attention to and fuse image features, enhancing the expression of visual features. This phased refinement strategy enables the model to gradually optimize the quality of generated images and improve semantic consistency between images and text.

In addition, the CMAGAN model also takes into account the quality, diversity, and similarity with real images when generating images through the comprehensive loss function, thereby generating artistic design images that not only conform to textual descriptions but also have high quality and rich details.

However, despite the significant achievements of the CMAGAN model in text generated art design image tasks, there are still some potential failure cases. For example, overly complex or vague text descriptions may hinder the model's ability to accurately interpret semantic information. This can result in generated images deviating from the original text description. In addition, since the model is trained on a large amount of data, if there are biases or deficiencies in the training data, it may also affect the generation performance of the model. Therefore, in future research, it is necessary to further explore how to optimize the text comprehension ability of the model and how to improve the generation performance of the model under complex and fuzzy text descriptions. At the same time, it is also necessary to pay attention to the quality and diversity of training data to ensure that the model can generate more accurate and rich artistic design images.

5 Conclusion

In this paper, we propose a cross-modal attention-based method for generating art design images from text, and verify the practical application of this method through a series of experiments. The experimental results show that the method is able to generate art design images that match the content of the text based on the text description, and it shows better results in detail presentation, atmosphere creation and overall aesthetics. And through the R-value accuracy, initial score (IS) and Frechet Inception Distance Score (FID) three kinds of evaluation criteria by way of ablation experiments to verify that the model applied in this paper's method can obtain the advantage of more significant image generation results. The advantages of the proposed method in generating artistic design images from text are conclusively validated through comparative analysis with existing literature. This paper's method can more accurately understand the semantic information of the input text and effectively transform it into images with artistic design sense, which provides users with richer and more diverse visual experiences.

References

[1] Tuerhong, G., Tian, L., Wushouer, M., & Dai, X. (2024). An end-to-end image-text matching approach considering semantic uncertainty. Neurocomputing, 607, 128386. DOI: 10.1016/j.neucom.2024.128386.

[2] Li, W., Kong, D. Q., Cao, G. G. Dai, C. Y. (2022). Cross-modality Medical Image Registration Based on Attention Residual Network. Computer Simulation, 9(11), 224-229. DOI: 10.3969/j.issn.1006-9348.2022.11.044.

[3] Endo, Y. (2024). Masked-attention diffusion guidance for spatially controlling text-to-image generation. The visual computer, 40(9), 6033-6045. DOI:10.1007/s00371-023-03151-y

[4] Bahani, M., El Ouaazizi, A., & Maalmi, K. (2023). The effectiveness of T5, GPT-2, and BERT on text-to-image generation task. Pattern recognition letters, 173, 57-63. DOI: 10.1016/j.patrec.2023.08.001

[5] Nezhad, N. M., Mirtaheri, S. L., & Shahbazian, R. (2023). Popular image generation based on popularity measures by generative adversarial networks. Multimedia Tools & Applications, 82(14), 20873-20897. DOI:10.1007/s11042-022-14090-6

[6] Watanabe, Y., Ogawa, N., Maeda, K., Ogawa, T., & Haseyama, M. (2024). Automatic findings generation for distress images using in-context few-shot learning of visual language model based on image similarity and text diversity. Journal of Robotics & Mechatronics, 36(2), 353-364. DOI:10.20965/jrm. 2024.p0353

[7] Thakur, N., Khan, N. U., & Sharma, S. D. (2023). A two-phase ultrasound image de-speckling framework by nonlocal means on anisotropic diffused image data. Informatica: An International Journal of Computing and Informatics, 47(2), 221-234. DOI:10.31449/inf. v47i2.4378

[8] Mehmood, R., Bashir, R., & Giri, K. J. (2024). VTM-GAN: video-text matcher based generative adversarial network for generating videos from textual description. International Journal of Information Technology, 16(1), 221-236. DOI:10.1007/s41870-023-01468-4

[9] Helmy, T., & Djatmiko, F. (2023). Framework for automatic semantic annotation of images based on image's low-level features and surrounding text. Arabian journal for science and engineering,48(2), 1991-2007. DOI:10.1007/s13369-022-06828-z

[10] Phaphuangwittayakul, A., Ying, F., & Chakpitak, Z. N. (2023). Few-shot image generation based on contrastive meta-learning generative adversarial network. The visual computer, 39(9), 4015-4028. DOI:10.1007/s00371-022-02566-3

[11] Kabir, A. I., Mahomud, L., Ald, A., & Ahmed, R. (2024). Empowering local image generation: harnessing stable diffusion for machine learning and AI. Informatica Economica, 28(1), 25-38. DOI:10.24818/issn14531305/28.1.2024.03

[12] Mishra, A., Agrawal, A., & Bhasker, S. (2024). Hybrid explainable image caption generation using image processing and natural language processing. International Journal of System Assurance Engineering and Management, 15(10), 4874-4884. DOI:10.1007/s13198-024-02495-5

[13] Nemer, Z. N., Jasim, W. N., & Harfash, E. J. (2023). Implementation of multiple cnn architectures to classify the sea coral images. Informatica: An International Journal of Computing and Informatics, 47(1), 43-50. DOI:10.31449/inf. v47i1.4429

[14] Praveen, R. G., Cardinal, P., & Granger, E. (2023). Audio–visual fusion for emotion recognition in the valence–arousal space using joint cross-attention. IEEE Transactions on Biometrics, Behavior, and Identity Science, 5(3), 360-373. DOI:10.1109/TBIOM.2022.3233083

[15] Pourkeshavarz, M., Nabavi, S., Moghaddam, M. E., & Shamsfard, M. (2024). Stacked cross-modal feature consolidation attention networks for image captioning. Multimedia Tools & Applications, 83(4), 12209-12233. DOI:10.1007/s11042-023-15869-x

[16] Tomer, M., & Kumar, M. (2022). STV-BEATS: skip thought vector and bi-encoder based automatic text summarizer. Knowledge-Based Systems, 240, 108108. DOI: 10.1016/j.knosys.2021.108108

[17] Mocanu, B., Tapu, R., & Zaharia, T. (2023). Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning. Image and vision computing, 133, 104676. DOI: 10.1016/j.imavis.2023.104676

[18] Matú Dopiriak, Gerec, J., & Gazda, J. (2024). Reconstruction of photorealistic 3d urban scenes using radiance fields as digital twins for autonomous driving. Acta Electrotechnica et Informatica, 24(4), 27-34. DOI:10.2478/aei-2024-0015

[19] Xu, M., Rahman, H. A., & Li, F. (2023). Automated generation of Chinese text-image summaries using deep learning techniques. Traitement du Signal, 40(6), 2835-2843. DOI:10.18280/ts.400644

[20] Borra, S. R, Ritika, K., Reddy, N. A., Shanvitha, N., Anusha, T., & Rajeswari, O. S. S. R. (2024). Parkinson net: convolutional neural network model for parkinson disease detection from image and voice data. Informatica (03505596), 48(2). DOI:10.31449/inf. v48i2.5077

A Review of Deep Multi-Objective Reinforcement Learning and Vision-Based Systems for Smart Cities

Saad Alharbi
Taibah University, Janadah Bin Umayyah Road, Tayba, Madinah 42353, Saudi Arabija
E-mail: stharbi@taibahu.edu.sa

Keywords: artificial intelligence, deep learning, computer vision, machine learning, multi-objective optimization, reinforcement learning, smart city

Received: June 18, 2025

Smart cities leverage artificial intelligence to address urban challenges such as traffic congestion, environmental sustainability, public safety, and energy efficiency. Among AI techniques, the integration of multi-objective reinforcement learning (MORL) and computer vision (CV) offers adaptive, real-time decision-making capabilities while processing complex visual data streams. This paper presents a comprehensive review of the joint application of MORL and CV in smart city environments. A systematic search was conducted across six major databases (Scopus, Web of Science, IEEE, Springer, MDPI, and Elsevier) from 2019 to 2024, resulting in the selection of 90 relevant studies. The review follows a thematic analysis approach, categorizing the literature into smart mobility, infrastructure, environment, governance, and smart living. Key findings indicate that multi-agent MORL and CV are increasingly used in traffic signal control, autonomous vehicle navigation, energy management, surveillance, healthcare, and waste management. However, despite advancements in deep RL algorithms like DDPG, PPO, SAC, and advanced CV techniques such as semantic segmentation and multi-camera tracking, the direct integration of these technologies remains underexplored in many domains. The paper highlights current research gaps, including the lack of standardized frameworks for MORL-CV synergy, scalability limitations, ethical concerns, and insufficient quantitative benchmarking across studies. Additionally, trends such as federated learning, edge computing, and digital twins are identified as promising enablers for future MORL-CV solutions in urban contexts. This review serves as a resource for researchers and policymakers aiming to develop sustainable and intelligent urban systems by bridging perception (via CV) with adaptive control (via MORL) for real-time, multi-criteria decision support.

Povzetek: Prispevek se ukvarja s povezovanjem večciljne okrepitevne poti (MORL) in računalniškega vida v pametnih mestih. S sistematično analizo 90 študij pokaže rastočo uporabo MORL–CV v prometu, energiji, nadzoru in zdravstvu in izpostavi vrzeli: odsotnost standardiziranih okvirjev, šibko neposredno integracijo ter omejitve glede razširljivosti in etike.

1 Introduction

In 2018, 55% of the population globally lived in urban areas, and the statistic is estimated to be 68% by 2050, according to the United Nations Department of Economic and Social Affairs [1]. This significant population shift toward cities has created major urban challenges, including traffic congestion, pollution, safety issues, and inefficient resource consumption. These complex and dynamic problems demand intelligent and adaptive solutions, giving rise to the concept of smart cities. For instance, megacities like Mumbai and Jakarta face severe congestion and pollution due to outdated infrastructure and rapid growth. A smart city employs technologies like Artificial Intelligence (AI), the Internet of Things (IoT), and big data to optimize infrastructure and urban services.

Among these, reinforcement learning (RL) and computer vision (CV) are key technologies that can provide efficient and scalable solutions to various urban issues. RL can assist in making optimal decisions based on environmental feedback, while CV helps machines interpret and understand visual inputs, enabling real-time monitoring and analysis. The integration of RL and CV in smart city systems has gained attention in recent years. RL, particularly multi-objective reinforcement learning (MORL), is effective in addressing urban systems' conflicting goals, such as reducing congestion while minimizing energy consumption. MORL techniques such as scalarization, Pareto-based optimization, and decomposition methods are particularly useful in balancing these competing objectives. Simultaneously, CV contributes through advanced techniques in image classification, object detection, and motion tracking that are highly relevant in urban scenarios like traffic analysis, surveillance, and waste monitoring. Recent techniques such as semantic segmentation and vision transformers have further enhanced scene understanding in smart environments.

Recent advances in deep learning have enhanced the effectiveness of both RL and CV in urban applications.

For example, deep Q-networks (DQN) and proximal policy optimization (PPO) have enabled scalable RL, while CNNs and transformers improved object recognition accuracy. Despite the potential and growing interest in these technologies, research that comprehensively examines the synergy between RL and CV in smart city applications remains limited. Existing studies often focus on either RL or CV in isolation, or they address specific application areas without offering a holistic view. While surveys exist on RL in traffic optimization [2] and CV in urban monitoring [3] , few examine their joint application across smart city systems. Figure 1 illustrates major technological milestones in RL and CV that have shaped their application in smart cities.

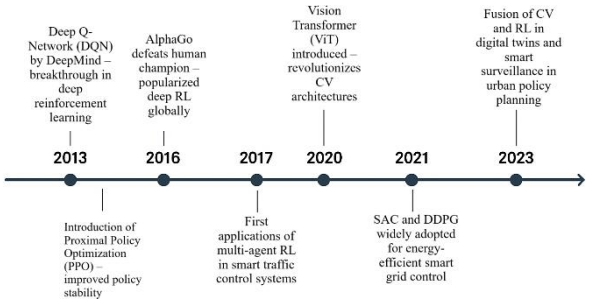


Figure 1: Major technological milestones in RL and CV

A focused review that explores the overlap between these fields in the context of smart cities is needed to highlight current advancements, identify research gaps,

This paper aims to fill this gap by reviewing the literature on multi-objective reinforcement learning and computer vision, specifically focusing on their combined use in smart city contexts. It categorizes and analyzes recent studies, summarizes trends and methodologies, and outlines the key challenges and open research questions in this emerging domain. Figure 2 illustrates the conceptual framework underpinning this review. It highlights how smart city applications leverage data sources (e.g., IoT devices and sensors) that are processed by Computer Vision (CV) for perception tasks and by Reinforcement Learning (RL) for adaptive decision-making and policy optimization. This integration enables real-time, multi-objective optimization across key urban domains such as mobility, infrastructure, governance, environment, and living.

Recent breakthroughs have accelerated the synergy between RL and CV for urban systems. Key milestones include DeepMind's AlphaGo (2016), which demonstrated the power of deep RL for complex decision-making [4], and the rise of Vision Transformers [5], which revolutionized visual feature extraction and scene understanding. These advancements, along with developments in federated learning and edge AI [6], underpin many of the smart city applications reviewed in this paper.

To address this gap, the following research questions guide this review:

RQ1: How are Deep Multi-Objective Reinforcement Learning (MORL) and Computer Vision (CV) currently integrated in smart city applications?
RQ2: What are the dominant algorithmic strategies, challenges, and limitations in applying MORL and CV jointly for urban decision-making?
RQ3: What are the emerging trends, gaps, and future research directions for enabling real-time, multi-criteria decision support systems combining MORL and CV?

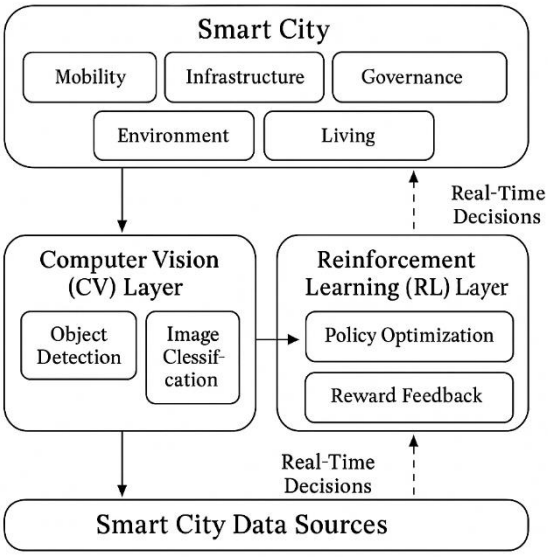


Figure 2: Conceptual Framework of RL and CV Integration in Smart City Systems (Source: Author)

2 Method

2.1 Search strategy

The current review gathered data from published articles from peer-reviewed journals and renowned databases like Scopus, Web of Science, IEEE, Springer, etc. (See Table 1). These databases were selected due to their high relevance, coverage of top-tier AI/ML journals,

and indexing of conference proceedings in the smart city domain. Data was searched based on investigating the role of ML and multi-objective RL specific to CV in enhancing smart city development. The search strategy employed the following Boolean string to maximize retrieval of relevant studies: ("multi-objective reinforcement learning" OR "MORL" OR "deep reinforcement learning" OR "DRL") AND ("computer vision" OR "vision-based systems") AND ("smart cities" OR "urban systems"). Search queries were customized for each database, using relevant syntax and filters to exclude non-related computational studies and retain only smart city applications. The review period was set between 2019 and 2024, and only English, peer-reviewed, full-text articles were considered.

This pairing enables smart city systems to process visual input (via CV) and make adaptive policy decisions (via RL), a combination increasingly required in dynamic urban environments. This search gathered data from 90

literature studies from the last six years (2019–2024) to collect the most recent advancements in ML-based developments in smart cities. Initial screening involved evaluating article titles and abstracts for relevance, followed by full-text review for methodological alignment with smart city applications. Studies were selected specifically for the keywords: Artificial Intelligence (AI), Machine Learning (ML), Multi-objective Reinforcement

Table 1: Data selection strategy

Years	Search Engines	Keywords
2019-2024	Google Scholar Scopus Web of Science IEEE Springer MDPI Elsevier Wiley	Artificial Intelligence (AI) Machine Learning (ML) Multi-objective Reinforcement Learning (RL) Deep Learning (DL) Computer Vision (CV) Smart Cities Decision Making Smart Mobility Smart Infrastructure Smart Governance Smart Economy Smart Environment Smart Living Smart People

Learning (RL), Deep Learning (DL), Computer Vision (CV), Smart Cities, Decision Making, Smart Mobility, Smart Infrastructure, Smart Governance, Smart Economy, Smart Environment, Smart Living and Smart People. Synonyms and related terms such as “urban computing,” “perception systems,” or “adaptive control” were also considered where applicable. Search keywords were combined with Boolean operators like AND, OR, WITHIN, etc., to create a relevant search strategy and boost data collection search. The query syntax and filtering parameters were customized per database to maximize relevance and exclude unrelated computational studies.

The selection process followed a PRISMA-inspired flow:

- Initial records identified: 312
- After duplicate removal: 280
- Full-text assessed for eligibility: 125
- Studies included in the final review: 90
- Figure 3 illustrates the selection process.

2.2 Inclusion and exclusion criteria

The articles were filtered based on criteria set for inclusion and exclusion based on the study’s relevance. The term "relevance" here specifically refers to the use of reinforcement learning (RL) and computer vision (CV) in smart city contexts. Table 2 was used as instrument in narrowing the broader set of retrieved articles to the 90 selected for this review, ensuring methodological and topical consistency.

2.3 Data extraction

Data extraction followed the six-step thematic analysis process proposed by **Braun & Clarke (2006)**, including: (1) familiarization with the data, (2) initial code generation, (3) theme identification, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report [7]. Such approach commonly applied in structured

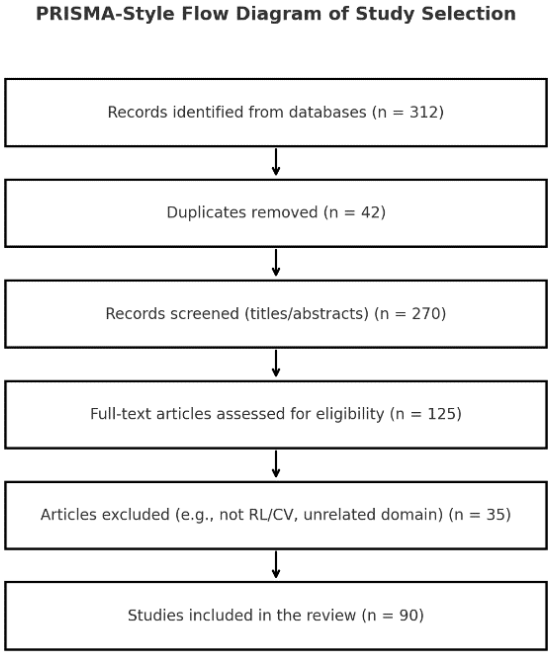


Figure 3: PRISMA-style flow diagram of the literature selection process adapted for AI/ML-based smart city review (Adapted from PRISMA 2020 guidelines [9]).

ML reviews such as [8]. This process was adapted to the technical context of RL and CV applications in smart cities. This process followed the widely recognized methodology involving familiarization, coding, theme development, and refinement, adapted to suit the technical

focus of the current study. Each article selected for review was explored based on the key outcomes, impacts on smart city applications, and concerned limitations. Key outcomes included aspects such as performance metrics, scalability, and practical deployment challenges in smart city environments. Repetitive keywords were used in the initial coding of articles to ensure robustness. Excel spreadsheets were used allowing categorization by recurring concepts such as optimization objectives, visual analytics, and policy learning mechanisms. Later, based on data coded, themes were developed, including: “smart city domains”, “Multi-objective RL and CV in smart cities for intelligent decision making”, and “applications of RL in CV in smart cities”. These themes were foundational in organizing the review sections and synthesizing cross-domain findings systematically.

The articles were coded into categories such as:

- Smart city domain (e.g., mobility, governance, infrastructure)

- Algorithmic approach (e.g., MORL, CV, joint RL-CV integration)
- Performance metrics (e.g., accuracy, energy savings, latency)
- Deployment setting (simulation or real-world)
- Data extraction was performed using Excel spreadsheets to ensure consistency and replicability.

Table 2: Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
All the studies included were published in English in peer-reviewed journals in full-text format.	Studies not published in peer-reviewed journals were excluded.
Studies published between 2019 and 2024 were included.	Studies published before 2019 were excluded.
Studies exploring the implications of ML, specifically RL and CV, in smart cities were included.	Studies discussing smart cities otherwise were excluded.
Quantitative, qualitative, and mixed-methods studies were included.	None were excluded based on their methodology.

2.4 Glossary of acronyms

This section provides a glossary of key acronyms and technical terms used throughout the paper.

- **RL** – Reinforcement Learning, a machine learning paradigm for sequential decision-making.
- **MORL** – Multi-Objective Reinforcement Learning, an RL approach handling multiple (possibly conflicting) objectives simultaneously.
- **CV** – Computer Vision, a field of AI that enables machines to interpret and process visual data.
- **DRL** – Deep Reinforcement Learning, an RL approach utilizing deep neural networks for complex policy learning.
- **DDPG** – Deep Deterministic Policy Gradient, a model-free RL algorithm for continuous action spaces.
- **PPO** – Proximal Policy Optimization, a stable policy-gradient-based RL method balancing exploration and exploitation.
- **SAC** – Soft Actor-Critic, an off-policy RL algorithm that improves convergence speed through entropy maximization.
- **MA2C** – Multi-Agent Actor-Critic, used for decentralized policy learning in multi-vehicle systems.
- **MOMCT** – Multi-Objective Multi-Camera Tracking, a CV technique for cross-camera object tracking and re-identification.
- **GAN** – Generative Adversarial Network, a deep learning model composed of generator and discriminator networks trained adversarially.
- **CNN** – Convolutional Neural Network, a deep learning architecture designed for image-based tasks.
- **ViT** – Vision Transformer, a transformer-based model for image classification using attention mechanisms.

3 Results

This section presents the findings of the reviewed studies, structured across five key smart city domains: mobility, infrastructure, governance, environment, and living. These domains were identified through thematic analysis, aligning with the smart city framework. Each subsection highlights the reinforcement learning (RL) and computer vision (CV) methods applied, their degree of integration, and domain-specific applications. Where available, performance metrics such as accuracy, scalability, and deployment challenges are discussed. This structure facilitates not only a comprehensive synthesis of state-of-the-art methods but also a critical identification of cross-domain trends, technical contributions, and existing research gaps.

Across the 90 reviewed studies, CV-based models for traffic detection and surveillance achieved an average detection accuracy of 82–88% mAP, with multi-camera tracking systems (e.g., MOMCT) reporting IDF1 scores above 80%**.** RL-driven energy management solutions demonstrated 8–20% energy savings compared to static scheduling baselines, while MORL approaches for autonomous mobility (e.g., MA2C and PPO) achieved 15–20% reductions in CO_2 emissions in simulation benchmarks. These metrics highlight the growing maturity of RL-CV pipelines in mobility and energy domains, though real-world deployments remain limited.

3.1 Smart city domains and CV

CV in many smart city domains enhances the safety, efficiency and security of smart cities by integrating efficient ML models. This integration supports critical tasks such as real-time surveillance, resource optimization, and autonomous control systems. The purpose of combining ML with CV is to advance edge detection, semantic segmentation, object detection, object recognition, text detection and others in various applications and spheres of smart cities. These functions form the perceptual layer of smart systems and are pivotal for interpreting urban sensory data in context-aware environments. These specific domains include smart mobility, smart governance, smart education, smart healthcare, smart economy, smart governance, smart living, smart environment, smart people & infrastructure [1], [10], [11] [12]. Table 3 shows the significant applications of all these smart city domains.

Smart mobility, connected with the goals of smart city expansion, solves the problem of traffic congestion and the environmental effects in large cities. It aims to ease the

mobility of individuals and goods in a town, generating six benefits: traffic, transport duration, the cost incurred in transport, pollution, noise and safety experienced when using the transport means. It fosters connected, sustainable and participative mobility, leveraging IoT, AI, Blockchain, and Big Data (BD) technologies to design and develop sustainable and disruptive solutions that rewrite the paradigm of cities and their inhabitants [13], [14], [15]. Smart mobility is essential in the enhancement of smart cities across the globe. In this case, double-parking and other roadside activities that are close to each other hinder traffic in areas with high transport density. The real-time IoT-based CVROSS, namely computer vision-based roadside occupation surveillance system, employs smart HD cameras and wireless modules. It watches traffic images dynamically, considers whether people are beside the road, and enhances image quality. Developed and piloted in Hong Kong, CVROSS is targeted to improve traffic and fleet for smart mobility [16].

Intelligent transport systems (ITS) provide solutions for smart infrastructure, mobility, traffic management and safety, and efficient, effective and secure means of transport [17], [18]. They are anticipated in areas such as public transport management, smart structure management, intelligent route management, superior vehicle control, integrated bill payment, and route details. Components such as global positioning system tracking aids on transit, technology in sensing, and video analytics assist in alleviating pollution and optimizing resources [19]. Smart transportation deals with safety, efficiency, and security in government rules, vehicle support services, and transportation payment services [10]. Therefore, the authors have put forward a procedure used for estimation of the wind turbine blades' angular velocity in a real wind farm scenario using ML and CV approaches. They use a WSN to capture videos and send them to a coordinator that computes the enhanced CV and ML Algorithms. The model can also be used to gather and parse more data about wind turbines such as mechanical distortions, surface erosions and super heating points in different wind turbine sections. Intelligent Energy Management Systems (IEMS), therefore, seek to optimize efficiency in electrical power generation, flexibility, renewable systems, and low carbon emanation through Machine Learning which predicts efficiency with IoE network support [13], [20], [21]. Smart education is associated with a technological enhancement in the classroom environment that brings about a smart classroom to facilitate tele-education solution similar to the physical classroom with voice recognition features using CV. This has shifted the use of libraries from places of study and quiet to discourse, peer, social learning and an exhibition hall. Lately, some institutions of higher learning are embracing technology to make their universities and libraries smart spaces as part of the new education trend; such advancements boost the competitiveness of a smart city.

Table 3: Smart city domains and involvements

Smart City Domain	Description	Involvements
Smart Mobility	It offers smart and intelligence mobility and transportation.	Traffic Monitoring and Management Smart Supply Chain Management Smart Parking and Routing Sustainable and Autonomous Mobility
Smart Living, Smart People & Infrastructure	It offers advancement of the city's infrastructure while improving the quality of life.	Smart Communication Smart Energy and Grid Smart Home Smart Community
Smart Education	It enhances the need to meet the emerging digital native generations' demands.	Virtual Reality based learning Student Management Smart Library Smart Classroom
Smart Governance	An application of innovation and technology to improve planning and decision-making in governing organizations.	Smart Urban Management Smart Urban Planning Smart Building Disaster Prevention and Management E-governance
Smart Economy	It is centered on technical innovation, resource efficiency, sustainability, and high social wellbeing.	Smart Supply Chain Smart Commerce E-commerce Retail
Smart Environment	An application of creating an environment with integrated sensors for a better environment.	Waste Management Smart Irrigation Smart Weather Monitoring Air Quality Monitoring
Smart Healthcare	A method for healthcare delivery uses IoTs, wearable tech, and mobile internet for enhanced resources, connecting people, data and institutions.	Telemedicine Smart Hospital E-health record Telenursing Patient Monitoring

The significance of IoT, ML, and CV for a smart education system lies in the use cases in a smart electric learning situation, smart classrooms, smart libraries, and smart attendance systems [22], [23], [24] .

Smart healthcare integrates health monitoring systems to determine the mobility of the human body using vision detection or sensors [25]. They are used in identifying unusual behavior among patients and help prevent avoidable deaths due to health complications. Patient monitoring systems will be useful in the next five to ten years, and vision- and sensor-based detection techniques will be useful [10]. Additionally, studies examined that telemedicine makes it possible and efficient for specific groups of patients: long-term care patients, specialized demand patients, and time-saving patients, and those patients who cannot overcome financial/ geographical barriers. It is in this capacity that disease management programs, with regards to the online space, enhances the delivery of healthcare services, resources, and patient care [26], [27], [28]. The telemedicine of the digital age that offers interactive discussions and video-conferences enhances the standard healthcare, the research data, and medication dosages. It facilitates treatment for medical practitioners from different geographical areas, whether in health centers or rural or urban areas [29], [30]. Overall, CV-based solutions for traffic and surveillance demonstrate higher performance maturity (e.g., detection accuracy often above 80% mAP in large datasets), whereas applications like smart healthcare and education remain experimental and lack standardized benchmarks. The absence of large-scale, multi-modal datasets integrating visual data with RL pipelines is a recurring gap highlighted in recent studies [21]. These CV tasks, when combined with RL-based decision-making policies, enable adaptive optimization. For example, outputs from multi-camera object detection can serve as state inputs for RL traffic controllers, enabling real-time congestion management.

3.2 RL in smart cities for intelligent decision making

RL is a type of ML that learns through trial-and-error interactions with the environment, aiming to optimize cumulative reward. In the context of smart cities, it enables systems to learn policies that adapt to dynamic urban environments, supporting real-time decision-making and autonomous control. Its applicability in smart cities revolves around policy learning and optimization, essential for autonomous systems that operate under uncertain and changing conditions. In various reviewed studies, RL were implemented in applications such as traffic signal control, resource management, and energy consumption optimization. Such implementations demonstrate RL's potential to replace static rule-based systems with adaptive, feedback-driven agents.

Multi-objective routing optimization solves barriers of ITS efficiently and rapidly when modelling uncertainty of time for transportation, resolving challenges of carbon emission cost, transportation cost and transport time using deep Q-learning algorithms like Artificial Fish Swarm Optimization (AFO), GA, Particle Swarm Optimization (PSO). DDPG is an approach defined by deep Q-learning with a continuous action space for physical tasks in simulating ITS [31], [32]. Two new approaches to multi-agent power control in device-to-device (D2D) communications based on a 5G network vehicle-to-vehicle (V2V) communications using the DDPG algorithm [33]. Compared with other deep RL methods, the developed models have higher network energy efficiency and flexibility for smart city applications and high-quality video streaming in real-time. To synthesize the differences between the commonly applied DRL algorithms, Table 4 compares DDPG, PPO, and SAC in terms of strengths, limitations, smart city use cases, and reported performance outcomes. Another work also presented a two-timescale federated deep RL framework named FDPG based on the DDPG in the smart city scenario to deal with the combined problem of resource allocation and task offloading under the premise of privacy preservation. This approach is intended to reduce the energy consumption of IoT devices under the given delay constraint [34], [35]. Some of the classification methods are CV-based solutions that effectively segregate the object from the heap of garbage and trash from the various waste objects that are commonly available, such as paper, paper boxes, food, glass, etc. Since the DL and Deep RL (DRL) techniques have recently advanced, the waste identification and detection make waste object classification possible. Thus, an intelligent DRL-based recycling waste object detection and classification (IDRL-RWODC) model for smart cities is proposed in this aspect. The IDRLRWODC technique should ideally aim at accurately identifying and categorizing waste objects, which can be achieved using the DL and DRL methods. The IDRL-RWODC model involves two processes: Mask Regional Convolutional Neural Network (Mask RCNN) based object detection and DRL-based object classification. Furthermore, the DenseNet model is used as a baseline model for the proposed Mask RCNN model. As a classifier, researchers used a deep Q-learning network (DQLN) [36]. PPO is one of the most common multi-objective RL algorithms [37] ; it is an advantage that alternates between exploitation and exploration, which benefits smart cities [3]. It controls the traffic signals depending on the data of the IoT sensors to decrease traffic and increase the ecologically friendly environment in urban areas. It can bring adaptive solutions for energy management and resource distribution according to current consumption rates. Therefore, using smart surveillance systems, PPO can increase public safety and detect disturbances at early stages when responding to potential threats [38], [39], [40]. Thus, PPO depends on data collection and analysis and the subsequent improvement in its functions that contribute to developing efficient and sustainable urban areas. PPO is applied to navigate autonomous vehicles using DRL and was therefore applied to this circumstance. It has been used to regulate the choice of lanes during the changeover and mixed traffic signals to guarantee vehicle safety at the interphase [41], [42].

A relatively new multi-objective DRL algorithm known as the soft actor-critic (SAC) is implemented to derive the best solution for energy distribution in EVs with battery-supercapacitor HESS. The recommended SAC-based EMS improves drawbacks in the current DRL techniques, including lower convergence speed and erratic training behavior [43]. The algorithm is embedded in the algorithm is embedded in an environment with continuous actions through self-play along with a newly designed reward signal. The results also confirm that the designed SAC-based EMS outperforms rule-based methods, deep deterministic policy gradient, and battery-centric setup. Using SAC algorithm, the proposed MASAC model for training arterial traffic control is superior to conventional MARL algorithms and multiband-based strategies [31], [44], [45].

Past studies have shown that innovative Reinforcement Convolutional Transfer Learning (RC-TL), a trajectory-prediction system based on CNN solves multi-objective and discrete problems. This system groups users with such backgrounds; each group has an RL agent to improve a CNN neural design; one model is trained per cluster from a limited user sample; the model is then shared with other users in the cluster. It is also helpful for object recognition of vehicles and pedestrians, automatic parking, autonomous driving, handover and service migration, network control, and caching such services in edge-computing and mobile networks [44], [46].

Multi-objective RL neural networks are also applied to smart cities for smart energy control [47]. The model-free control methods like RL enable easier power deployment without having to derive or find a thermal model of the building. HVAC, DHW, and home lighting systems have been controlled using artificial neural networks (ANN) to decrease energy bills and discomfort. RL algorithms have been shown to provide near-optimal solutions for maintaining energy in low-exergy buildings and multi-objective deep RL controllers have been developed to reduce electricity consumption and solve problems of energy resource management [45], [48]. However, most of the contributions incorporate simplified building energy models or specific RL applications and, therefore, pose the need to expand the high-performance simulation model with enhanced ML techniques [49], [50], [51], [52].

There has been continued refinement of unmanned automated vehicles (UAVs) and progressive control systems, with the concept of smart cities targeting self-driving automobiles for task offloading and trajectory control. A multi-objective algorithm for car mobility in road intersections is designed using a hierarchy reinforcement learning (HRL) method. The specific objective of comparing the performance of a signalized vehicular intersection with a fixed-time traffic controller using HRL can be as efficient as traditional traffic controllers [2], [53]. Self-driving vehicles require intelligence to make sound decisions and actions in traffic situations. However, many cases of single optimization/sampling-based motion planners and end-to-end learning methods fail to produce safe trajectories in real-time. A hierarchical behavior planning architecture, which consists of low-level safe controllers and one high-level DRL (H-CtRL), is thus proposed. Low-level controllers ensure safety, while the H-CtRL is adaptive and efficient. This algorithm was tested with a simulator that eliminated the chance of failing in real-life situations, as was tested and proven successful in smart cities [41], [54].

To provide a structured comparison of existing approaches, Table 5 presents a summary of recent studies combining Multi-Objective Reinforcement Learning and Computer Vision techniques in smart city applications. The comparison outlines the algorithmic strategies, application domains, vision tasks, optimization objectives, and reported performance metrics. Additionally, the table highlights the degree of integration between RL and CV in each study, distinguishing between independent use, parallel deployment, and fully integrated pipelines. This dual-purpose analysis addresses both the literature synthesis gap in related works and the need for cross-domain comparative insight identified in the current review findings.

Comparative studies show that PPO and SAC achieve faster convergence and better adaptability for dynamic traffic and energy systems, while DDPG excels in continuous control but struggles with sample inefficiency [55] [56]. Despite their potential, most RL solutions remain evaluated in simulations, with few real-world deployments at city scale.

The DRL algorithms discussed (DDPG, PPO, SAC) often rely on perception modules powered by CV (e.g., object detection or motion tracking) to build dynamic environment states. This fusion of perception and control forms the foundation for intelligent decision-making pipelines in smart cities.

3.3 Applications of multi-objective RL and CV in smart cities

3.3.1 Smart mobility

For sustainable urban mobility, multi-objective RL and CV advance smart mobility management, enhancing shared mobility and public transport, advancing public transport optimization, traffic management, autonomous vehicles, and smart parking [56], [61], [62], [63] [64]. These techniques enable learning-based adaptation to real-time traffic states, making them suitable for dynamic routing, congestion pricing, and policy simulations.

An example of the presented multi-objective algorithms is a Multi-Agent Actor-Critic (MA2C) for multi-AV lane-changing in a mixed traffic environment, which is one of the significant parts of ITS in smart cities. RL is then used to optimize the behavior of the algorithm with respect to lane-changing, leading to better environmental impact and quality of life standards for occupants of urban areas. It also provides a realistic imitation of driver-controlled vehicular dynamics that are useful in traffic flow control. By addressing lane change control, passenger comfort, and inter-vehicle cooperation, the efficiency of urban traffic flow is maintained without

Table 4: Comparative Overview of Key DRL Algorithms in Smart Cities

Algorithm	Strengths	Limitations	Example Use Cases	Performance Highlights
DDPG	Effective in continuous action spaces; suitable for complex control tasks	Requires careful hyperparameter tuning; sample inefficient	Smart energy management and V2V communications [34] [57]	Achieved higher energy efficiency but slower training compared to PPO
PPO	Stable training, robust to hyperparameters; balances exploration and exploitation	Moderate computational cost	Traffic signal optimization, smart surveillance [58], [59]	Reduced congestion and improved traffic flow by 10–15%
SAC	High sample efficiency; faster convergence due to entropy maximization	Sensitive to reward design and tuning	Energy distribution in EV systems[13] [60]	Outperformed rule-based and DDPG methods in energy savings and scalability

Table 5: Summary of RL and CV Integration in Smart City Applications, Comparative overview of algorithms, domains, integration types, and performance metrics across recent studies

Study	Algorithm/Model	Application Domain	CV Task (if any)	RL Objective	Performance Metrics
[61]	MOMCT (Multi-Objective Multi-Camera Tracking)	Smart Mobility	Object Tracking & Re-identification	Optimize multi-camera association	mAP 85.3%, IDF1 80.1%
[56]	MA2C (Multi-Agent Actor-Critic)	Traffic Management	Traffic Flow Detection	Lane changing, emission reduction	Travel time ↓ 15%, CO2 ↓ 12%
[20]	IEMS with RL	Energy Management	Sensor-based Vision	Power grid optimization	Energy savings ↑ 10%
[21]	CV + ML for Wind Turbines	Renewable Energy	Turbine Blade Motion Detection	Angular velocity estimation	Error ↓ 5%, Remote accuracy ↑
[65]	H-RL-VaNSAS	Public Transport	GIS & Route Mapping	Bus route optimization	Distance ↓ 8%, Safety ↑
[66]	PMORL	Healthcare/Smart Living	Patient Activity Recognition	Energy allocation & emissions	Emission ↓ 20%, Accuracy ↑
[60]	SAC-based Traffic Control	Smart Infrastructure	Visual traffic input	Traffic signal optimization	Throughput ↑ 10%, Delay ↓

congestion and less emissions are produced compared to other models [56]. The use of decentralized agents improves scalability and allows for localized learning while preserving global system performance.

Multi-Objective Multiple Camera Tracking (MOMCT) is a critical challenge in the field of computer vision and security involving a search for multiple objects in a scene and tracking their movements from the video shot by various cameras. Distribution learning based on MOMCT in intelligent transportation, public security and self-driving introduces the main object detectors, discusses the advanced level of evaluation information and summarizes benchmark datasets [61]. MOMCT is also essential for cross-camera identity association and spatiotemporal consistency, both crucial in smart surveillance systems. Combining multi-objective RL and ML constant avoidance of heuristics made by human beings can solve combinatorial optimization issues. These methods are faster than metaheuristic methods, particularly for big problems that are useful in IoT-based smart cities. However, RL and ML applications in combinatorial optimization integrate multiple algorithms, for instance, shuffled frog leaping, graywolf, earthworm, and simulated annealing that have been used on power management, traffic routing, traffic police scheduling, and home energy management. The integration of hybrid metaheuristics with policy-learning-based RL enhances exploration capabilities and accelerates convergence on large-scale problems. Further, the deployment of self-driving cars is considered optimal in urban livery, such as business application Fleet Management System and consumer-use ride-sharing apps [62], [63], [67]. Despite these advancements, current studies largely implement RL for mobility optimization and CV for traffic perception as

separate components. Integrating real-time visual inputs—such as vehicle and pedestrian detection—into MORL-based traffic policy adaptation remains an underdeveloped research direction. Among mobility-focused studies, MA2C-based lane optimization shows up to 12–15% CO_2 reduction and better scalability than PPO or traditional MARL approaches. However, few methods leverage CV outputs (e.g., multi-camera tracking) as direct inputs for real-time RL-based traffic signal control, which remains a critical research opportunity.

CV tasks like vehicle and pedestrian detection (e.g., MOMCT) feed real-time visual information to RL-based traffic signal optimization and lane-changing strategies, improving both safety and flow efficiency.

3.3.2 Smart infrastructure

DL algorithms efficiently support smart infrastructure development by proposing high-end applications of smart grids, smart building management, energy-efficient behavior and management optimization, renewable power-based energy storage, and infrastructure monitoring. These applications involve multi-objective problems and optimization algorithms [68], [69]. In these domains, DL models offer real-time processing capabilities that outperform traditional statistical methods, particularly in high-dimensional energy consumption datasets.

Residential energy management using different ML algorithms for Building Energy Management Systems (BEMS) is developed to achieve lower power consumption and energy costs. These are demand-side management schemes, residential load scheduling, hybrid optimization algorithms, hybrid home energy management systems (HEMS), and optimization of home energy management systems incorporated with renewable sources of energy. This has been assessed using the decision trees as well as the Gaussian Naive Bayes and the K-Neighbors, feed-forward multilayer perceptron (MLP) neural network ML and DL models. The comparative performance of these models often varies depending on seasonality, household size, and the granularity of energy consumption data.

RL is used to gain energy conservation significance to climate control, lighting, ventilation, heating, and air conditioning (HVAC) applications that crest to 10% and for water heaters crest of 20% energy conserved for smart infrastructure [69], [70], [71], [72]. Adaptive RL agents can adjust HVAC schedules based on occupant patterns and weather forecasts, enhancing comfort without compromising energy goals. Additionally, the Improved Whale Optimization Algorithm (IWOA) is used to optimize domestic appliance design in the HEMS, which is a smart grid tool [41]. The use of IWOA enables better solution diversity and convergence speed in highly non-linear design spaces, making it suitable for real-time smart grid decision support systems. A critical observation is that smart infrastructure research typically applies MORL for energy management and CV for structural monitoring in parallel but not in synergy. Future work should integrate CV-based occupancy detection into RL-driven adaptive energy management systems. Most RL-based energy management systems achieve 8–20% energy savings compared to static scheduling [20], but almost none integrate real-time CV-based occupancy detection for adaptive reward shaping. Future designs should merge these two pipelines to improve context-awareness and building automation.

While RL dominates energy optimization in infrastructure, integrating CV (e.g., occupancy detection or visual inspection of grids) as a feedback signal can enhance adaptive energy management policies.

3.3.3 Smart governance

Advancements in multi-objective digital twin technology have made governments able to integrate smart governance applications monitoring urban management among people, environment and infrastructure through digital identity systems and civic engagement through intelligent surveillance [73], [74]. It enhances the transparency and efficiency of the administrative processes involving AI, data analytics, and ML to streamline decision-making in urban administration [75]. Digital twins enable simulation-based governance, allowing predictive modeling of city-wide scenarios such as disaster management, energy use, or urban planning interventions.

The application of AI chatbots in smart governance in smart cities has relied on the approaches to rule-based chatbots and natural language processing for customer support. They also reflected on the fact that multi-agent systems also serve the purpose of intelligent governance of administrative procedures by integrating AI and IoT in e-government services. These systems improve scalability and responsiveness in citizen service platforms, particularly when integrated with cloud infrastructure and edge devices. To develop smart cities, improve citizen relations, and offer quality services, AI and ML work efficiently. The optimal approach to identifying the extent of the positive impact of AI in the provision of urban public services is based on the dynamic interaction between urban public services and AI technology [75]. This interaction can be modeled as a feedback loop between user behavior data and adaptive service models, enhancing personalization and performance over time.

Moreover, to support smart governance, big data technology collects multi-source and multi-objective heterogeneous data to enhance the RL for urban governance. The author explains a cooperative traffic signal control approach based on a swarm RL approach, which is a hybridization of PSO and RL. Swarm RL techniques allow decentralized control policies that can adapt in real time to traffic flow variability and urban disruptions. The method outperforms other single RL-based adaptive traffic signal control methods, infrared sensors, Identification Devices (RFIDs), and GPS systems facilitating the collection and evaluation of real-time data [74]. Despite the adoption of intelligent surveillance and multi-agent RL in governance, there is limited exploration of frameworks where CV-based urban observations directly inform adaptive RL-driven policy adjustments.

This represents a key gap in current research. Current governance solutions focus more on RL-based policy optimization than on fusing visual surveillance data into adaptive decision-making loops. This imbalance limits their effectiveness in real-time urban risk detection and disaster management.

Intelligent surveillance systems using CV can provide situational data that multi-agent RL systems use for decentralized governance decisions (e.g., adaptive traffic signal policies during emergencies).

Although CV is less prominent in smart infrastructure compared to RL-based control systems, emerging works explore CV for fault detection and visual monitoring of energy grids. These CV outputs can be fused with RL models to enhance decision-making in smart building automation and predictive maintenance systems.

3.3.4 Smart environment

Smart environment entails proper utilization of the available resources, especially in energy usage using multi-objective RL. Smart grids use AI in order to increase the effectiveness, accuracy and eco-friendliness of the overall system. These new electrical networks incorporate information technology in electricity generation, distribution, and use. The integration of distributed sensors, IoT devices, and data analytics allows smart grids to perform real-time adjustments based on consumption patterns and grid status. Deep RL techniques are employed to find the best way to trade energy and perform demand response mechanisms in a Peer-to-peer (P2P) energy system to reduce costs at the household level. In these systems, RL agents learn optimal trading strategies while considering dynamic pricing, grid constraints, and user preferences, enhancing both individual and collective energy efficiency. In green smart cities and electricity consumption, advanced ML and RL techniques are applied, like the LSTM and RNNs, to enhance the dynamic characteristics of smart grids [75].

Additionally, another study introduced a new short-term electricity load forecasting method based on ML and RL-CNN approaches, which improved accuracy and extended the running time. RL-CNN hybrids combine spatial pattern recognition with adaptive reward optimization, improving both prediction accuracy and the adaptability of forecasting systems to seasonal shifts. A new smart energy management system based on RL was also introduced that uses Markov decision processes and Q-learning to learn energy allocation via real energy trading. The proposed energy management algorithm proves to show the numerical efficacy of reduced electricity costs [74]. Markov models support the modeling of sequential decision-making in environments with probabilistic transitions, making them highly relevant for grid stability under demand uncertainty. While CV technologies monitor environmental indicators such as air quality and waste, their outputs are seldom linked to real-time MORL systems for automated environmental management. This remains a promising area for future exploration. Comparative evaluations suggest that RL-based demand response mechanisms reduce energy costs by up to 10–15%, yet no existing system actively links CV-based environmental monitoring (e.g., air quality, waste) to adaptive RL policies. This gap presents a promising direction for holistic smart environment solutions.

Future RL-environment models could leverage CV-based waste detection or air quality monitoring, transforming perceptual data into actionable RL-driven resource allocation strategies.

3.3.5 Smart living

Smart living is about improving public safety, healthcare systems, and emergency systems using intelligent vision-based systems for rapid urbanization. These systems leverage real-time data streams and predictive analytics to support faster incident detection, efficient public services, and citizen-centric city planning.

The Hybrid RL-Variable Neighborhood Strategy Adaptive Search (H-RL-VaNSAS) algorithm is utilized to develop a multi-objective model for creating sustainable, resilient, safe, and efficient urban bus routes. The model provides better-optimized bus routes with more resilience, sustainability, and safety to the cities of Ubon Ratchathani and Warinchamrab, given possible constraints in accessibility and the need to minimize travel distance. This approach contributes to urban transportation for smart living, where it provides a more systemic and realistic solution to some modern problems in relation to tourism areas [65]. By combining local search heuristics with deep RL policy optimization, the H-RL-VaNSAS algorithm adapts to spatial-temporal fluctuations in transport demand.

Similarly, multi-objective RL also has the following beneficial impacts on healthcare systems. A Generative Adversarial Network (GAN) is applied to the patient data, creates an artificial treatment scenario through Deep Neural Networks (DNN) and determines the transfer likelihood of individualized radiotherapy. This GAN-DNN integration enhances personalized treatment planning by generating synthetic clinical outcomes for reinforcement-based decision learning. An intelligent non-convex optimization problem of deep RL is employed to reduce environmental emissions in urban transport systems. The method also incorporates smart healthcare technologies to forecast the likelihood of appropriate travelling modes. The approach optimizes decisions and stimulates the commuter mode shift towards active transport, consequently enhancing health-related statistics and lowering gCO2 emissions [74], [76].

With the utility interconnection for integration of RL and storage system, the amount of greenhouse gas emissions and fossil diesel is minimized in microgrids in smart living. Propagation of a multi-microgrid power system can help to resolve security and privacy issues. An intelligent energy management approach is used depending upon preference-based multi-objective reinforcement learning (PMORL) methods for developing such systems [66]. PMORL frameworks enable energy optimization while incorporating user-defined utility

preferences, offering an ethical and sustainable foundation for decentralized urban energy planning. In smart living, vision-based systems are commonly used for patient monitoring and public safety, yet RL agents operate in isolation from these perceptual systems. Integrating CV outputs into MORL-based adaptive healthcare and safety interventions is a critical research opportunity.

Table 6 presents an expanded synthesis of the reviewed smart city applications, highlighting both shared patterns and domain-specific distinctions in the integration of reinforcement learning (RL) and computer vision (CV). Across most domains, RL is predominantly used for policy optimization, while CV serves as a perception layer; however, deployment strategies differ based on task urgency, data modalities, and ethical constraints. For example, mobility systems demand real-time, low-latency control, whereas infrastructure domains prioritize long-term optimization of energy and resources. Smart governance raises concerns around privacy and surveillance, while healthcare and smart living applications emphasize human-centered and interpretable decision-making. This integrated taxonomy clarifies how multi-objective RL and CV are currently aligned—and where critical integration gaps persist. Notably, recent studies show that Pareto-optimal RL (PMORL) applied to healthcare logistics can reduce carbon emissions by up to 20% while improving patient activity recognition; however, CV in these cases often remains a standalone monitoring tool rather than being jointly optimized with RL. Addressing this disconnect through **vision-reinforced policy learning frameworks** represents a key opportunity for advancing next-generation smart living solutions

While RL is widely applied for optimizing energy usage and emissions, CV contributes to environmental monitoring by analyzing satellite imagery, thermal patterns, or pollution dispersion. The integration of CV-derived inputs into MORL pipelines enables more informed, adaptive control strategies in environmental management.

Across the reviewed domains, certain commonalities and divergences emerge. For instance, multi-agent RL is predominantly used in mobility for dynamic traffic control and lane optimization, while governance applications emphasize decentralized decision-making and policy simulation. CV tasks such as object detection and semantic segmentation are more prevalent in mobility and surveillance but are less integrated into infrastructure or energy-focused RL pipelines. While mobility prioritizes real-time adaptation and latency reduction, domains like infrastructure and environment favor long-term optimization and energy efficiency. Table 7 provides a synthesized taxonomy capturing these overlaps and distinctions. Patient activity recognition from CV systems can serve as input for RL agents managing healthcare logistics and emergency response, forming a cohesive perception-to-decision pipeline.

4 Discussion and research implications

This comprehensive review reveals a clear research trend toward applying reinforcement learning (RL), particularly multi-objective RL (MORL), and computer vision (CV) in various smart city domains. However, a detailed examination of the reviewed studies highlights that these technologies are often implemented in parallel rather than in integrated frameworks. While CV is primarily used for perception tasks such as surveillance, traffic monitoring, and environmental sensing, MORL focuses on optimizing decision-making processes, energy management, and resource allocation. A critical finding is the lack of unified systems that couple real-time visual data with adaptive RL-based control in urban environments. For example, in smart mobility, multi-camera tracking systems enhance surveillance, but traffic flow optimization typically operates independently of vision-based feedback. Similarly, in smart infrastructure, energy management relies on MORL algorithms, but visual occupancy data from CV systems is rarely incorporated into RL-driven HVAC or lighting control loops.

Table 6: Multi-objective RL and smart city applications

Studies	Multi-Objective Algorithm/Approach	Smart City Domain	Smart City Applications
[56], [61], [62], [63]	Multi-Agent Actor-Critic (MA2C) approach Multi-Objective Multiple Camera Tracking (MOMCT)	Smart Mobility	Traffic management, autonomous vehicles, public transport optimization, ITS, smart parking, traffic routing, and traffic police scheduling.
[41], [68], [70], [71]	Gaussian Naive Bayes K-Neighbors Feed-forward multilayer perceptron (MLP) neural network	Smart Infrastructure	Smart grids, smart building management, energy-efficient behavior and management optimization, renewable power-based energy storage and infrastructure monitoring

	Improved Whale Optimization Algorithm (IWOA)		
[74]	Swarm RL approach	Smart Governance	Digital identity systems, RFIDs, and intelligent surveillance systems.
[74], [75]	Q-learning RL-CNN technique Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs)	Smart Environment	Smart grids, Peer to Peer (P2P) energy system
[65], [66]	Hybrid RL-Variable Neighborhood Strategy Adaptive Search (H-RL-VaNSAS) Generative Adversarial Network (GAN) Deep Neural Networks (DNN) Preference-based multi-objective reinforcement learning (PMORL)	Smart Living	Sustainable, safe, and efficient urban bus routes, urban transport systems, smart healthcare and smart energy management systems.

Table 7: Taxonomy of RL and CV Applications in Smart City Domains.

Smart City Domain	Commonalities in RL/CV Use	Distinctive Aspects / Divergences
Smart Mobility	Multi-agent RL for dynamic control, CV for traffic monitoring and object tracking	Real-time decision-making, integration of sensor fusion, route optimization under traffic constraints
Smart Infrastructure	RL for energy/resource management, CV for environment monitoring	Focus on energy-saving, building-level automation, less real-time interaction compared to mobility
Smart Governance	RL for adaptive urban policy, CV for surveillance and digital twins	Emphasis on social systems, ethical implications, use of digital identity and civic engagement data
Smart Environment	RL for emission control, CV for anomaly detection in environmental data	Primarily focuses on predictive control, less integration with human-centric data
Smart Living	RL for healthcare logistics and emergency response, CV for health monitoring	Human-centered applications, ethical considerations in data privacy and health predictions

Moreover, the reviewed literature shows inconsistent treatment of "multi-objective" frameworks. Some studies refer to multi-criteria trade-offs like energy efficiency vs. comfort, while others focus on multi-agent cooperation without explicitly managing conflicting objectives. This indicates a need for clearer definitions and standardized methodologies when applying MORL in smart city contexts. Emerging technologies offer promising solutions to these gaps. The rise of digital twins enables simulation environments where CV and RL models can co-evolve, testing scenarios that combine perception and policy adaptation. Likewise, federated learning can support privacy-preserving CV data sharing across urban systems, while edge computing reduces latency, enabling real-time RL-CV interactions.

Despite advances, ethical considerations remain underexplored. Combining vision systems with autonomous decision-making raises concerns about bias, privacy, and surveillance overreach. Future research must address these societal implications alongside technical developments.

4.1 Research implications

- **Integration of RL and CV Pipelines:** Future studies should develop frameworks where CV directly informs RL reward functions or state representations, creating closed-loop adaptive systems.
- **Standardized Benchmarks:** The field lacks consistent performance metrics for evaluating RL-CV synergy in smart cities. Establishing such benchmarks would facilitate cross-domain comparisons.
- **Scalability and Deployment:** There is a need for real-world deployments at city-scale, leveraging edge computing and digital twin environments for testing RL-CV integration under dynamic conditions.
- **Ethical and Social Dimensions:** Addressing privacy, fairness, and transparency in vision-augmented decision-making is critical for public trust and adoption.

4.2 Emerging research directions and ethical considerations

The evolution of smart city technologies is opening new avenues for integrating reinforcement learning (RL) and computer vision (CV). Emerging trends include the use of Digital Twins for simulating urban environments, allowing RL and CV systems to co-adapt in virtual testbeds before real-world deployment. Additionally, edge computing is enabling real-time CV and RL decision-making at the sensor level, reducing latency and improving scalability [77]. The integration of Large Language Models (LLMs), such as GPT-4, with RL and CV pipelines presents further opportunities, particularly for dynamic policy generation and adaptive urban management. Moreover, federated learning is increasingly used to preserve data privacy while enabling cross-organizational learning across cities [78].

Alongside these advancements, ethical considerations must be addressed. The deployment of vision-based systems coupled with autonomous RL decision-making introduces risks of bias in surveillance systems, privacy violations, and lack of transparency in policy decisions. For example, facial recognition technologies deployed in urban surveillance have raised concerns about racial bias and wrongful identification, underscoring the need for bias mitigation protocols and public oversight. Responsible development requires frameworks for explainability, fairness, and data governance to ensure societal acceptance and long-term sustainability of these technologies [79].

5 Conclusion

The smart city concept is based on gathering real-time data from urban infrastructures like waste disposal and power grid systems. This data forms the foundation for adaptive feedback loops and real-time policy enforcement mechanisms in urban ecosystems. In the advancement of smart city infrastructure development, the role of multi-objective RL is essential and associated with more significant time and cost efficiency to analyze data-intensive workloads. The role of multi-objective RL and CV is substantial in a set of multi-objective applications in smart cities, such as cyber-security, healthcare, smart grids, smart waste management, smart traffic monitoring, security surveillance, and others. These techniques contribute to multi-criteria optimization under uncertainty, offering adaptive and scalable solutions that surpass conventional static models. Advancements in multi-objective digital twin technology have enabled governments to integrate smart city applications that monitor urban management among people, the environment, and infrastructure through digital identity systems, as well as civic engagement through intelligent surveillance. This convergence of RL, CV, and digital twin systems signals a shift toward predictive governance and citizen-driven service personalization in next-generation urban design.

6 Limitations and directions for future research

This comprehensive review explored that multi-objective RL in smart city applications are limited to few aspects such as data dependency, computational complexity, scalability issues, privacy and ethical concerns and trade-off management. The dependence on large-scale labeled data and real-time feedback loops poses significant barriers for deployment in data-scarce or privacy-restricted urban environments. However, future research can advance the analysis but exploring the high-quality datasets for training dealing with the unresolved challenges of infrastructure limitations using quantum computing, digital twins and Internet of Things. Emerging paradigms like federated learning and edge computing may also help mitigate data centralization and latency

issues while preserving privacy. The focus of future researchers should depend upon combining multi-objective RL with adaptive ML models and optimization techniques to improve collaboration among multiple agents in smart cities. Emerging paradigms like federated learning and edge computing may also help mitigate data centralization and latency issues while preserving privacy.

References

[1] M. Herath and H. Mittal, "Adoption of artificial intelligence in smart cities: A comprehensive review," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100076, 2022. https://doi.org/10.1016/j.jjimei.2022.100076

[2] B. Song, F. Xing, H. Wang, X. Luo, S. Dai, P. Xiao, and Z. Zhao, "Evolutionary multi-objective reinforcement learning-based trajectory control and task offloading in UAV-assisted mobile edge computing," *IEEE Transactions on Mobile Computing*, vol. 22, no. 12, pp. 7387–7405, 2022. https://doi.org/10.1109/tmc.2022.3208457

[3] H. Yang, J. Zhang, and J. Wang, "Urban traffic control in software-defined Internet of Things via a multi-agent deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3742–3754, 2020. https://doi.org/10.1109/tits.2020.3023788

[4] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016. https://doi.org/10.1038/nature16961

[5] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," *arXiv preprint* arXiv:2010.11929, 2020.https://doi.org/10.48550/arXiv.2010.11929
Z. Zheng, Y. Zhou, Y. Sun, Z. Wang, B. Liu, and K. Li, "Applications of federated learning in smart cities: Recent advances, taxonomy, and open challenges," *Connection Science*, vol. 34, no. 1, pp. 1–28, 2022. https://doi.org/10.1080/09540091.2021.1936455

[6] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. https://doi.org/10.1191/1478088706qp063oa
Y. Zheng, Q. Hao, J. Wang, C. Gao, J. Chen, D. Jin, and Y. Li, "A survey of machine learning for urban decision making: Applications in planning, transportation, and healthcare," *ACM Computing Surveys*, vol. 57, no. 4, art. 99, pp. 1–41, 2024. https://doi.org/10.1145/3695986

[7] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.https://doi.org/10.1136/bmj.n71
D. Kothadiya, A. Chaudhari, R. Macwan, K. Patel, and C. Bhatt, "The convergence of deep learning and computer vision: Smart city applications and research challenges," *Proceedings of the 3rd International Conference on Integrated Intelligent Computing, Communication and Security (ICIIC 2021)*, 2021. https://doi.org/10.2991/ahis.k.210913.003

[8] [11] U.-X. Lau, B. P. L. Marakkalage, S. H. Zhou, Y. Hassan, N. U. Yuen, C. Zhang, M. Tan, "A survey of data fusion in smart city applications," *Information Fusion*, vol. 52, pp. 357–374, 2019. https://doi.org/10.1016/j.inffus.2019.05.004

[9] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011. https://doi.org/10.1109/TITS.2011.2119372

[10] S. Mao, J. Sun, Q. Wang, X. Muthu, and B. Krishnamoorthy, "The importance of public support in the implementation of green transportation in the smart cities," *Computational Intelligence*, vol. 40, no. 1, e12326, 2024. https://doi.org/10.1111/coin.12326

[11] G. Paiva, S. Ahad, M. A. Tripathi, G. Feroz, N. Casalino, "Enabling technologies for urban smart mobility: Recent trends, opportunities, and challenges," *Sensors*, vol. 21, no. 6, p. 2143, 2021. https://doi.org/10.3390/s21062143

[12] A. Šurdonja, S. Giuffrè, and T. Deluka-Tibljaš, "Smart mobility solutions – necessary precondition for a well-functioning smart city," *Transportation Research Procedia*, vol. 45, pp. 604–611, 2020. https://doi.org/10.1016/j.trpro.2020.03.051

[13] K. L. Ho, G. T. S. Tsang, Y. P. Wu, C. H. Wong, W. H. Choy, and W. H. Choy, "A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities," *Sensors*, vol. 19, no. 8, p. 1796, 2019. https://doi.org/10.3390/s19081796

[14] S. I. Musa, A. A. Malami, S. I. Alanazi, F. Ounaies, W. Alshammari, M. Haruna, "Sustainable traffic management for smart cities using Internet-of-Things-oriented intelligent transportation systems (ITS): Challenges and recommendations," *Sustainability*, vol. 15, no. 13, p. 9859, 2023.https://doi.org/10.3390/su15139859

[15] Y. Zhang, T. Cheng, and J. Zou, "Multimodal transportation routing optimization based on multi-objective Q-learning under time uncertainty," *Complex & Intelligent Systems*, vol. 10, pp. 3133–3152, 2024. https://doi.org/10.1007/s40747-023-01308-9

[16] R. K. Upadhyay, S. K. Sharma, and V. Kumar, "Introduction to intelligent transportation system and advanced technology," in *Intelligent Transportation System and Advanced Technology*, R. K. Upadhyay, S. K. Sharma, and V. Kumar (eds), *Energy, Environment, and Sustainability*, Springer, Singapore, 2024. https://doi.org/10.1007/978-981-97-0515-3_1

[17] H. Akram, A. Abbas, S. Khan, M. Athar, A. Ghazal, and T. Al Hamadi, "Smart energy management system using machine learning," *Computers, Materials & Continua*, vol. 78, no. 1, pp. 1–11, 2024. https://doi.org/10.32604/cmc.2023.032216

[18] S. Bahaghighat, M. Abedini, F. Xin, Q. Zanjireh, and M. M. Mirjalili, "Using machine learning and

computer vision to estimate the angular velocity of wind turbines in smart grids remotely," *Energy Reports*, vol. 7, pp. 8561–8576, 2021. https://doi.org/10.1016/j.egyr.2021.07.077

[19] H. C. Y. Chan and L. Chan, "Smart library and smart campus," *Journal of Service Science and Management*, vol. 11, no. 6, pp. 543–564, 2018.https://doi.org/10.4236/jssm.2018.116037

A. Dimitriadou and E. Lanitis, "A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms," *Smart Learning Environments*, vol. 10, no. 1, p. 12, 2023. https://doi.org/10.1186/s40561-023-00231-3

[20] S. Al-Fraihat, M. Joy, R. Masa'deh, and J. Sinclair, "Evaluating e-learning systems success: An empirical study," *Computers in Human Behavior*, vol. 102, pp. 67–86, 2020. https://doi.org/10.1016/j.chb.2019.08.004

[21] D. Mshali, H. Lemlouma, T. Moloney, and M. Magoni, "A survey on health monitoring systems for health smart homes," *International Journal of Industrial Ergonomics*, vol. 66, pp. 26–56, 2018. https://doi.org/10.1016/j.ergon.2018.02.002

[22] Z. Brommeyer, M. Whittaker, M. Mackay, M. Ng, and F. Liang, "Building health service management workforce capacity in the era of health informatics and digital health – A scoping review," *International Journal of Medical Informatics*, vol. 169, p. 104909, 2023. https://doi.org/10.1016/j.ijmedinf.2022.104909

[23] N. Jaya Lakshmi and N. Jabalia, "Body sensor networks as emerging trends of technology in health care systems: Challenges and future," in *Efficient Data Handling for Massive Internet of Medical Things Healthcare Data Analytics*, pp. 129–157, 2021. https://doi.org/10.1007/978-3-030-66633-0_6

[24] A. O. Okolo, C. A. Arowoogun, J. O. Chidi, and R. Adeniyi, "Telemedicine's role in transforming healthcare delivery in the pharmaceutical industry: A systematic review," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1836–1856, 2024. https://doi.org/10.30574/wjarr.2024.21.2.0609

[25] T. Shaik, X. Tao, N. Higgins, L. Li, R. Gururajan, X. Zhou, and U. R. Acharya, "Remote patient monitoring using artificial intelligence: Current state, applications, and challenges," *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 2, e1485, 2023. https://doi.org/10.1002/widm.1485

[26] E. J. Peralta-Ochoa, A. M. Chaca-Asmal, P. A. Guerrero-Vásquez, L. F. Ordoñez-Ordoñez, J. O. Coronel-González, "Smart healthcare applications over 5G networks: A systematic review," *Applied Sciences*, vol. 13, no. 3, p. 1469, 2023. https://doi.org/10.3390/app13031469

[27] D. Z. Ribeiro, D. A. Melgarejo, D. C. Saadi, M. Rosa, R. L. Rodríguez, "A novel deep deterministic policy gradient model applied to intelligent transportation system security problems in 5G and 6G network scenarios," *Physical Communication*, vol. 56, p. 101938, 2023. https://doi.org/10.1016/j.phycom.2022.101938

[28] Q. Zhang, K. Xu, H. Pan, B. Zheng, "Explicit coordinated signal control using soft actor–critic for cycle length determination," *IET Intelligent Transport Systems*, 2024. https://doi.org/10.1049/itr2.12519

[29] L. D. Nguyen, K. K. Duong, T. Q. Vien, N. A. Le-Khac, and N. A. Nguyen, "Distributed deep deterministic policy gradient for power allocation control in D2D-based V2V communications," *IEEE Access*, vol. 7, pp. 164533–164543, 2019. https://doi.org/10.1109/access.2019.2952411

[30] G. Chen and X. Liu, "Federated deep reinforcement learning-based task offloading and resource allocation for smart cities in a mobile edge network," *Sensors*, vol. 22, no. 13, p. 4738, 2022. https://doi.org/10.3390/s22134738

[31] B. Qiu, C. Hu, Y. Chen, and Y. Zeng, "Deep deterministic policy gradient (DDPG)-based energy harvesting wireless communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8577–8588, 2019. https://doi.org/10.1109/jiot.2019.2921159

[32] R. Duhayyim, M. A. Elfadil Eisa, T. A. Al-Wesabi, F. N. Abdelmaboud, A. Hamza, M. A. Zamani, A. S. Rizwanullah, and M. Marzouk, "Deep reinforcement learning enabled smart city recycling waste object classification," *Computers, Materials & Continua*, vol. 71, no. 3, 2022. https://doi.org/10.32604/cmc.2022.024431

[33] W. Ryu and K. Kim, "Multi-objective optimization of energy saving and throughput in heterogeneous networks using deep reinforcement learning," *Sensors*, vol. 21, no. 23, p. 7925, 2021. https://doi.org/10.3390/s21237925

[34] E. Vamvakas, D. Michailidis, P. Korkas, and C. Kosmatopoulos, "Review and evaluation of reinforcement learning frameworks on smart grid applications," *Energies*, vol. 16, no. 14, p. 5326, 2023. https://doi.org/10.3390/en16145326

[35] M. Vijayalakshmi and V. Saravanan, "Reinforcement learning-based multi-objective energy-efficient task scheduling in fog-cloud industrial IoT-based systems," *Soft Computing*, vol. 27, no. 23, pp. 17473–17491, 2023. https://doi.org/10.1007/s00500-023-09159-9

[36] J.-H. Yun, W. J. Park, S. Kim, J. Shin, M. Jung, S. Mohaisen, D. A. Kim, "Cooperative multi-agent deep reinforcement learning for reliable surveillance via autonomous multi-UAV control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086–7096, 2022. https://doi.org/10.1109/tii.2022.3143175

[37] X. Li and W. Xu, "A hybrid evolutionary and machine learning approach for smart building: Sustainable building energy management design," *Sustainable Energy Technologies and Assessments*, vol. 65, p. 103709, 2024. https://doi.org/10.1016/j.seta.2024.103709

[38] S.-H. Quang Tran and D. Bae, "Proximal policy optimization through a deep reinforcement learning framework for multiple autonomous vehicles at a non-signalized intersection," *Applied Sciences*, vol.

10, no. 16, p. 5722, 2020. https://doi.org/10.3390/app10165722

[39] N. Mazzi, Y. Sassi, H. B. Errahimi, F. Es-Sbai, "A soft actor-critic reinforcement learning framework for optimal energy management in electric vehicles with hybrid storage," *Journal of Energy Storage*, vol. 99, p. 113344, 2024. https://doi.org/10.1016/j.est.2024.113344

[40] R. Hafiz, A. M. Parah, S. A. Bhat, "Reinforcement learning applied to machine vision: State of the art," *International Journal of Multimedia Information Retrieval*, vol. 10, no. 2, pp. 71–82, 2021. https://doi.org/10.1007/s13735-021-00209-2

[41] M. H. Abid, M. S. Apon, H. J. Hossain, S. Ahmed, A. Ahshan, R. Lipu, "A novel multi-objective optimization-based multi-agent deep reinforcement learning approach for microgrid resources planning," *Applied Energy*, vol. 353, p. 122029, 2024. https://doi.org/10.1016/j.apenergy.2023.122029

[42] N. Emami, L. Pacheco, A. Di Maio, and T. Braun, "RC-TL: Reinforcement convolutional transfer learning for large-scale trajectory prediction," in *NOMS 2022—2022 IEEE/IFIP Network Operations and Management Symposium*, 2022. https://doi.org/10.1109/noms54207.2022.9789883

[43] K. Diyan, M. Silva, and B. N. Han, "A multi-objective approach for optimal energy management in smart homes using reinforcement learning," *Sensors*, vol. 20, no. 12, p. 3450, 2020. https://doi.org/10.3390/s20123450

[44] C. E. Leite, G. M. C. Jiménez-Fernández, S. Salcedo-Sanz, C. G. Marcelino, and C. G. Pedreira, "Solving an energy resource management problem with a novel multi-objective evolutionary reinforcement learning method," *Knowledge-Based Systems*, vol. 280, p. 111027, 2023. https://doi.org/10.1016/j.knosys.2023.111027

[45] L. Prawiyogi, A. G. Purnama, S. Meria, "Smart cities using machine learning and intelligent applications," *International Transactions on Artificial Intelligence*, vol. 1, no. 1, pp. 102–116, 2022. https://doi.org/10.33050/italic.v1i1.204

[46] R. Ullah, Z. Al-Turjman, F. Mostarda, and L. Gagliardi, "Applications of artificial intelligence and machine learning in smart cities," *Computer Communications*, vol. 154, pp. 313–323, 2020. https://doi.org/10.1016/j.comcom.2020.02.069

[47] Z. Vázquez-Canteli, J. R. Ulyanin, S. Kämpf, and J. Nagy, "Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities," *Sustainable Cities and Society*, vol. 45, pp. 243–257, 2019. https://doi.org/10.1016/j.scs.2018.11.021

[48] V. Zhang, M. Wang, X. Sathishkumar, and V. Sivakumar, "Machine learning techniques based on security management in smart cities using robots," *Work*, vol. 68, no. 3, pp. 891–902, 2021.https://doi.org/10.3233/WOR-203423

[49] M. Shikunov and A. I. Panov, "Hierarchical reinforcement learning approach for the road intersection task," in *Biologically Inspired Cognitive Architectures 2019: Proceedings of the Tenth Annual Meeting of the BICA Society*, Springer, 2020. https://doi.org/10.1007/978-3-030-25719-4_64

[50] [54] O. A. Borges, D. F. Leite, J. P. R. Moreira, E. M. Carpinteiro, "Traffic light control using hierarchical reinforcement learning and options framework," *IEEE Access*, vol. 9, pp. 99155–99165, 2021. https://doi.org/10.1109/access.2021.3096666

[51] L. Mao, F. Li, Z. Lin, Y. Li, "Mastering arterial traffic signal control with multi-agent attention-based soft actor-critic model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3129–3144, 2022. https://doi.org/10.1109/TITS.2022.3229477

[52] H. M. Louati, A. Louati, H. Kariri, E. Neifar, W. Hassan, M. K. Khairi, M. H. Farahat, M. A. El-Hoseny, "Sustainable smart cities through multi-agent reinforcement learning-based cooperative autonomous vehicles," *Sustainability*, vol. 16, no. 5, p. 1779, 2024. https://doi.org/10.3390/su16051779

[53] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint* arXiv:1509.02971, 2016. https://doi.org/10.48550/arXiv.1509.02971

[54] W. Jia and M. Ji, "Multi-agent deep reinforcement learning for large-scale traffic signal control with spatio-temporal attention mechanism," *Applied Sciences*, vol. 15, no. 15, p. 8605, 2025. https://doi.org/10.3390/app15158605

[55] H. Wei, G. Zheng, H. Yao, and Z. Li, "IntelliLight: A reinforcement learning approach for intelligent traffic light control," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2017, pp. 2496–2505. https://doi.org/10.1145/3219819.3220096

[56] F. Çakıl and N. Aksoy, "Reinforcement learning-based multi-objective smart energy management for electric vehicle charging stations with priority scheduling," *Energy*, vol. 322, p. 135475, 2025. https://doi.org/10.1016/j.energy.2025.135475

[57] B. Fei and L. Han, "Multi-object multi-camera tracking based on deep learning for intelligent transportation: A review," *Sensors*, vol. 23, no. 8, p. 3852, 2023. https://doi.org/10.3390/s23083852

[58] E. Nisyrios and K. Gkiotsalitis, "Optimization-based approaches for traffic and fleet management of connected and autonomous vehicles: A systematic literature review," *International Journal of Intelligent Transportation Systems Research*, 2025. https://doi.org/10.1007/s13177-025-00536-2

[59] A. Syed, A. S. Sierra-Sosa, D. Kumar, and A. Elmaghraby, "Making cities smarter—optimization problems for IoT-enabled smart city development: A mapping of applications, objectives, constraints," *Sensors*, vol. 22, no. 12, p. 4380, 2022. https://doi.org/10.3390/s22124380

[60] M. T. Rafique, A. Mustafa, and H. Sajid, "Reinforcement learning for adaptive traffic signal

control: Turn-based and time-based approaches to reduce congestion," *arXiv preprint arXiv:2408.15751*, 2024. https://doi.org/10.48550/arXiv.2408.15751

[61] P. Sriprateep, K. Pitakaso, R. Khonjun, S. Srichok, T. Luesak, P. Gonwirat, S. Kaewta, C. Kosacka-Olejnik, and M. Enkvetchakul, "Multi-objective optimization of resilient, sustainable, and safe urban bus routes for tourism promotion using a hybrid reinforcement learning algorithm," *Mathematics*, vol. 12, no. 14, p. 2283, 2024. https://doi.org/10.3390/math12142283

[62] A. M. Kazmi, S. A. Khan, Z. Khan, A. Mazzara, and M. Khattak, "Leveraging deep reinforcement learning and healthcare devices for active travelling in smart cities," *IEEE Transactions on Consumer Electronics*, 2024. https://doi.org/10.1109/TCE.2024.3470978

[63] B. Nama, M. Nath, A. Bechra, N. Bhatia, J. Tanwar, S. Chaturvedi, and M. Sadoun, "Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges," *International Journal of Communication Systems*, vol. 34, no. 9, e4814, 2021.https://doi.org/10.1002/dac.4814

[64] C. Konstantakopoulos, I. C. Barkan, A. R. He, S. Veeravalli, T. Liu, H. Liu, and C. Spanos, "A deep learning and gamification approach to improving human-building interaction and energy efficiency in smart infrastructure," *Applied Energy*, vol. 237, pp. 810–821, 2019. https://doi.org/10.1016/j.apenergy.2018.12.065

[65] H. Lin, Y. H. Lin, M. D. Tsai, K. T. Deng, and M. J. Ishii, "Multi-objective optimization design of green building envelopes and air conditioning systems for energy conservation and CO_2 emission reduction," *Sustainable Cities and Society*, vol. 64, p. 102555, 2021. https://doi.org/10.1016/j.scs.2020.102555

[66] M. A. Ghofrani, A. Nazemi, and S. D. Jafari, "HVAC load synchronization in smart building communities," *Sustainable Cities and Society*, vol. 51, p. 101741, 2019. https://doi.org/10.1016/j.scs.2019.101741

[67] S. Mason and K. Grijalva, "A review of reinforcement learning for autonomous building energy management," *Computers & Electrical Engineering*, vol. 78, pp. 300–312, 2019. https://doi.org/10.1016/j.compeleceng.2019.07.019

[68] W. L. Mocanu, E. Nguyen, P. H. Gibescu, and M. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016. https://doi.org/10.1016/j.segan.2016.02.005

[69] Z.-J. M. Deng, T. Zhang, K. Shen, "A systematic review of a digital twin city: A new pattern of urban governance toward smart cities," *Journal of Management Science and Engineering*, vol. 6, no. 2, pp. 125–134, 2021. https://doi.org/10.1016/j.jmse.2021.03.003

[70] S. Tong, Z. Ye, F. Yan, M. Liu, and H. Basodi, "A survey on algorithms for intelligent computing and smart city applications," *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 155–172, 2021. https://doi.org/10.26599/bdma.2020.9020029

[71] K. Wolniak and R. Stecuła, "Artificial ntelligence in smart cities—applications, barriers, and future directions: A review," *Smart Cities*, vol. 7, no. 3, pp. 1346–1389, 2024. https://doi.org/10.3390/smartcities7030057

[72] M. Xu, J. Li, K. Abusara, "Preference-based multi-objective reinforcement learning for multi-microgrid system optimization problem in smart grid," *Memetic Computing*, vol. 14, no. 2, pp. 225–235, 2022. https://doi.org/10.1007/s12293-022-00357-w

[73] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017. https://doi.org/10.1109/MC.2017.9

[74] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, "Federated learning in smart city sensing: Challenges and opportunities," *Sensors*, vol. 20, no. 21, p. 6230, 2020. https://doi.org/10.3390/s20216230

[75] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, pp. 1–21, 2016. https://doi.org/10.1177/2053951716679679

Dynamic Event-Triggered Non-Vulnerable Fault-Tolerant Control for Wind Power Cyber-Physical Systems under Cyber-Attacks

Yanan Liu[1], Yafei Li[2*]
[1]Department of Equipment Manufacturing and Control, Hebei Chemical & Pharmaceutical College
Shijiazhuang 050026, China
[2]Department of Railway Locomotive and Rolling Stock, Hebei Vocational College of Rail Transportation
Shijiazhuang 052260, China
E-mail: hbgdyslyf@163.com
[*]Corresponding author

Keywords: cyber-attacks, wind power, cyber-physical system, fault-tolerant control

Received: June 19, 2025

The study addresses the problem of the insufficient robustness of traditional control strategies under network attacks by quantifying the impact of attacks. It accomplishes this by establishing the dynamics of the wind turbine and the sensor/actuator attack model, designing a dynamic event trigger mechanism with an adaptive trigger threshold adjustment to reduce communication frequency, developing a fault-tolerant controller based on state feedback and fault compensation, and optimizing the control gain matrix using Lyapunov theory. The simulation results showed that, under Scene 3 (maximum attack intensity), the recovery time for rotor angular velocity was less than 4 seconds, communications are reduced by 70% compared to the traditional strategy, and energy consumption increases by no more than 25%. The sensor measurement error was reduced from 0.23 to 0.01. The comprehensive robustness index was 0.75, and the gain margin and phase margin were maintained at 4.5dB and 35.5 ° respectively. This research provides a control scheme for the physical system of a wind power network that balances security, real-time performance, and energy efficiency. It has engineering value in improving the ability of new energy power systems to resist interference.

Povzetek: Članek obravnava robustnost vetrnih kibernetsko-fizikalnih sistemov glede na kibernetske napade. Predlaga dinamični dogodek-proženi nevulnerabilni FT-nadzor, ki združi modeliranje napadov, prilagodljivo proženje ter povratno-zvezni in kompenzacijski nadzor, optimiziran z LMI/Lyapunovom.

1 Introduction

One significant renewable energy source that has been growing in the power system's share is wind power (WP). A sophisticated system that intricately combines physical, communication, and computational processes is the wind power cyber-physical system (WPCPS). It realizes data interaction and cooperative control among the components of wind turbines through network communication, which greatly improves the efficiency and reliability of WP generation [1-2]. In recent years, WPCPS has made remarkable development. Meanwhile, to increase the system's stability and performance, researchers have been investigating novel control techniques [3]. However, with the deep integration of WPCPS with the network, it faces increasing risks of cyber-attacks. Cyber-attackers may tamper with sensor data, interfere with communication signals, or damage the control system. This can lead to wind turbine failures or even shutdowns, which seriously affects the safe and stable operation of the power system [4]. In addition, due to the complexity and uncertainty of WPCPS itself, components in the system may fail, further exacerbating the vulnerability of the system. Currently, although there have been some studies on the control of WPCPSs, relatively few studies have been conducted on non-vulnerable fault-tolerant (FT) control under cyber-attacks [5].

Cyber-physical system (CPS) security refers to the stability and resilience of the overall system operation by coordinating information technology and physical devices to guard against cyber-attacks and physical failures during their interaction. For attack detection in industrial CPSs, Alotaibi et al. developed an integrated model based on deep learning and fuzzy soft expert sets. The study enhanced the recognition of the model through a whale optimization algorithm. The experimental results showed that the accuracy of the method reached 99.01% [6]. Benmalek M proposed a dual classification method based on infection vectors, targets, and target and technology attributes for ransomware threats to CPS. The study identified the need to improve software resilience in CPS environments by analyzing 10 real-world incidents [7]. Rehman proposed a CPS-oriented security requirements engineering methodology that outperformed existing methods. This was demonstrated by comparing the two and identifying CPS-specific security requirements within a broader context [8]. To increase the overall robustness

of CPS systems in critical infrastructures using a risk assessment tool, Adamos et al. suggested a methodology for automated complexity analysis of CPSs utilizing graph theory and state modeling methods. The tool's ability to detect high complexity components and their influence on the overall CPS design, as well as to provide mitigation points for the adoption of security measures, was demonstrated by the findings [9]. Pandey et al. used a data-driven strategy to model the control system behavior in a CPS and used an artificial intelligence based classifier to predict the acceptable state of the control system. The method was validated on data from a water treatment test bed with an F1 score of 0.96 [10].

One tactic used in the creation of a control system that can sustain steady system functioning in the case of an assault is the non-vulnerable FT control approach. For multi-area power systems with actuator failures, Li et al. suggested a fault distribution-dependent non-vulnerable H ∞ FT load frequency control method. An arithmetic example examination of a three-area power system demonstrated the efficacy of the technique [11]. Hu et al. proposed an adaptive fault estimation and model prediction FT control method for hypersonic engines with non-Gaussian uncertain output jet plumes. Simulation results verified that the method outperformed the comparison method [12]. Harno et al. constructed a state feedback FT controller based on the robust H∞ control method constructing a state feedback FT controller by considering structural uncertainty and integral quadratic constraints in linear definite systems. The study verified the effectiveness of the proposed robust control method through numerical examples and showed that the level of disturbance attenuation was lower than that of other reliable H∞ controllers [13]. For amorphous flat ground-space wireless self-assembling network systems, Wang et al. suggested a neural network-based direct robust adaptive non-vulnerable FT control method. The controller enhanced the system stability performance by 66.7%, according to simulation findings [14]. The issue of finite time interval decentralized fault estimation and FT control in the face of unexpected faults plagues nonlinear interconnected distributed parameter systems. Song et al. created an FT controller with non-fragile properties in order to solve these issues. The method's excellence, practicality, and feasibility were demonstrated by the simulation results [15]. The summary table of related works is shown in Table 1.

In summary, the current research on CPS security is fruitful. Progress has been made in the fields of electric power and aviation regarding non-vulnerable, FT control technology under cyber-attacks. However, the adaptability of existing methods to WPCPS is limited. WPCPS is a system characterized by strong stochasticity and high dynamics [16]. The existing CPS or FT control methods (such as those proposed by Li et al. and harno et al.) are not enough to deal with the problem of WPCPS. The main reason is that WPCPS has strong randomness and high dynamic characteristics. However, Li et al.'s method only addresses actuator faults in multi-regional power systems without considering randomness factors, such as wind speed fluctuations in WP systems. The controller designed by Harno et al. is based on linear uncertain systems and lacks a dynamic event triggering mechanism. This mechanism is necessary for balancing network communication loads and controlling in real time. In addition, these methods do not perform quantitative modeling or targeted compensation for additive attacks on sensors and actuators. This makes it difficult to ensure the stable operation of WPCPS in attack scenes. The CPS security requirements engineering method proposed by Rehman et al. (2024) focuses on requirements modeling, but does not involve attack dAefense at the dynamic control level. The AI classifier method of Pandey et al. (2025) can predict the acceptable state of the control system. However, it lacks collaborative optimization of communication load and control performance. Based on these two factors, the integrated design of attack quantitative modeling, real-time communication optimization, and robust control is realized by integrating the dynamic event triggering mechanism and the state feedback fault compensation controller. This makes up for the deficiency in existing research that combines dynamic control strategies with engineering practicability. This study aims to design a non-fragile, fault-tolerant controller with a dynamic event triggering mechanism for WPCPSs under network attacks. The controller should be able to resist additive and bounded energy attacks and ensure that the system's comprehensive robustness index is no less than 0.7. Additionally, it should reduce network traffic by over 60% compared to traditional strategies. The research assumes that the network attack signal energy is bounded, the sensor and actuator attacks exist in the form of additive disturbance, and the system parameter uncertainty fluctuates within the preset range. Under different intensity attack scenes, the expected results are as follows: the recovery time of rotor angular velocity is less than 4 seconds. Sensor measurement error and actuator control deviation are significantly reduced. Moreover, system energy consumption increases by no more than 25%. These results are achieved while maintaining sufficient gain margin (≥4 dB) and phase margin (≥30°), balancing security, real-time performance, and energy efficiency.

Table 1: The summary table of related works.

Method	Target system	Attack type modeling	Key quantitative results	Limitations
Alotaibi et al. [6]	Industrial network physical system	Attack detection	The attack detection accuracy is 99.01%	Fault tolerant control strategy is not involved
Benmalek M [7]	Network physical system	Blackmail software attack	/	No attack mathematical model established
Rehman S [8]	Network physical system	Security requirements	/	Controller design not involved
Adamos et al. [9]	Critical infrastructure network physical systems	Complexity analysis	/	No fault tolerant control strategy is proposed
Pandey et al. [10]	Water treatment test platform	anomaly detection	Classifier F1 score is 0.96	Dynamic control strategy is not involved
Li et al. [11]	Multi area power system	Actuator failure	/	Tampering of sensor/actuator by network attack is not considered
Hu et al. [12]	Hypersonic engine	Non Gaussian uncertain output	/	Not designed for network attack
Harno et al. [13]	Linear uncertain system	Structural uncertainty	The disturbance attenuation level is lower than other reliable h∞ controllers	Dynamic event triggering mechanism is not introduced
Wang et al. [14]	Air ground wireless self assembled network system	/	System stability performance improved by 66.7%	Impact of unquantified attacks on physical processes
Song et al. [15]	Nonlinear interconnected distributed parameter system	Sudden failure	/	Network attack scenarios are not considered

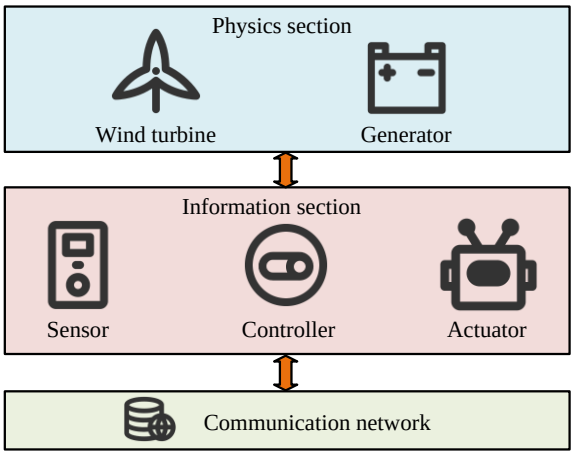


Figure 1: Schematic diagram of WPCPS structure.

The innovation of the research lies in constructing a dynamic systems framework that integrates sensor and actuator attack signal models. This framework precisely portrays the dual impact of cyber-attacks on physical processes and information interactions. It does so in response to the unique operating characteristics of WPCPS. These characteristics are strongly random and highly dynamic in nature. Simultaneously, a FT control strategy that integrates a dynamic event triggering mechanism is designed to achieve synergistic compensation for parameter uncertainty and attack interference while effectively reducing network communication load.

2 Methods and materials

2.1 WPCPS model construction

WPCPS is a complex system that integrates physical processes and information interaction. Its modeling needs to comprehensively consider the physical part, the information part, and the interaction mechanism between the two. The dynamic response and stability characteristics of the system under cyber-attack should also be analyzed [17]. The schematic diagram of the WPCPS structure is shown in Figure 1.

In Figure 1, the WPCPS consists of a physical part, an information part, and a communication network. The information part mainly consists of sensors, actuators and communication network. Sensors are used to measure various physical quantities of the WP system, such as wind speed, wind direction, generator speed, power, etc [18]. The output signals of these sensors are transmitted to the controller through the communication network. The controller processes and makes decisions based on the received information. Then, the control signal (CS) is sent to the actuator through the communication network. The actuator regulates and controls the WP generation system based on the CSs. It is assumed that the system model is linearized and suitable for small disturbance scenes near the rated operating conditions. The network attack energy is bounded, and the continuous attack with infinite energy is not considered. Attacks are limited to sensor/actuator additive attacks, and scenes involving communication interruptions caused by denial-of-service attacks are excluded. The parameter uncertainty range should be kept

within ±10% of the nominal value. The wind turbine is a key device for converting wind energy into mechanical energy. Its dynamics model is described as shown in Equation (1).

$$J\frac{d\omega}{dt}=T_a-T_g-D\omega \tag{1}$$

In Equation (1), J represents the total rotational inertia of the wind turbine and generator rotor. ω is the angular velocity of the rotor. T_a is the aerodynamic torque. T_g is the electromagnetic torque of the generator. D is the damping coefficient. The electromagnetic torque of the generator is related to the stator current and other parameters, as shown in Equation (2).

$$T_g=K_i i_s \tag{2}$$

In Equation (2), K_i is the generator torque constant. i_s is the stator current. The flow of sensor information acquisition and transmission is shown in Figure 2. First, the sensor acquires the state of the physical system and converts the physical quantities into a transmittable form of information such as electrical signals. Then, the acquired signals are transmitted to the controller through a communication network, which may be affected by noise and network attacks during the transmission process [19].

The sensors are responsible for collecting information about various physical quantities of the WP generation system, such as wind speed, rotor angular velocity, etc. The mathematical model representation of the signal collected by the sensor is shown in Equation (3).

$$y_s=C_p x+v \tag{3}$$

In Equation (3), y_s is the sensor acquisition signal. C_p is the output matrix of the sensor. x is the state vector of the system, which contains physical quantities such as rotor angular velocity and generator electromagnetic torque. v is the measurement noise of the sensor. The system state vector is 3×1 vector, and its specific components include rotor angular velocity, generator electromagnetic torque and pitch angle. It is supplemented that the third state variable "pitch angle" is used to completely describe the core physical state of the fan operation. The external attack input vector is a 3×1 vector. Each component corresponds to a disturbance of sensor data, actuator control signals, or communication links, respectively. This defines the physical connotation of the attack vector. Both the sensor output matrix and the actuator gain matrix are 3×3 matrices, which are completely matched with the 3×1 state vector and attack vector in dimension. This ensures the consistency of matrix operation and the preciseness of the model. The actuator receives commands from the controller and acts on the physical system. The input-output relationship is shown in Equation (4).

$$u_a=G_a u_c \tag{4}$$

In Equation (4), u_a is the actuator output, i.e., the control quantity acting on the physical system. G_a is the actuator gain. u_c is the CS output by the controller. The information is transmitted between the sensor, controller, and actuator through a communication network. Considering factors such as delay and packet loss during network transmission, the transmission model can be expressed as Equation (5).

$$\hat{u}_c(k)=u_c(k-\tau(k)) \tag{5}$$

In Equation (5), $\hat{u}_c(k)$ is the CS actually received by the actuator. $u_c(k)$ is the CS output by the controller at moment k. $\tau(k)$ is the network transmission delay, which is a time-varying random variable. When the system is subjected to a network attack, the signals collected by the sensors may be tampered with. Assuming that its attacked signal is a_s, the output of the sensor after the attack is shown in Equation (6).

$$y_{s,a}=C_p x+v+a_s \tag{6}$$

The CS received by the actuator may also be attacked. Assuming that it is subjected to an attack signal of a_a, the input of the actuator after the attack is shown in Equation (7).

$$u_{a,a}=G_a u_c+a_a \tag{7}$$

The flow of actuator CS transmission and action is shown in Figure 3. The CS output from the controller is transmitted to the actuator through the communication network. After the actuator receives the signal, it converts the signal into the actual control quantity according to its own gain matrix and acts on the devices such as wind turbines and generators in the physical system. Meanwhile, the process may also be subject to cyber-attacks [20].

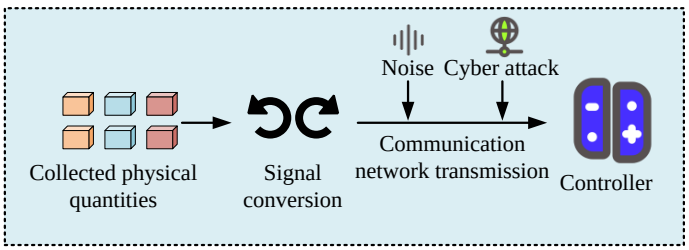


Figure 2: Flow of sensor information acquisition and transmission.

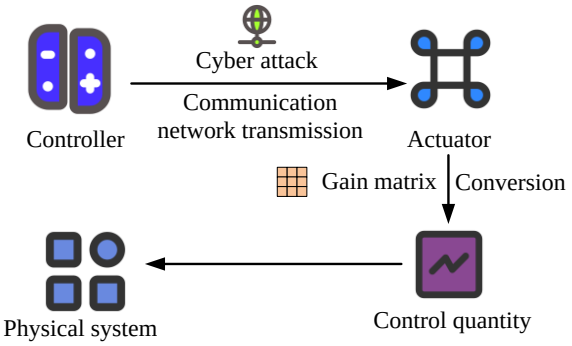


Figure 3: Actuator CS transmission and action process.

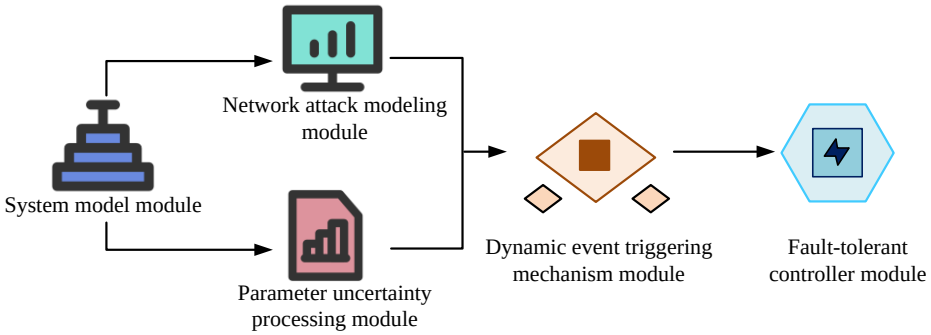


Figure 4: Overall structure diagram of non fragile fault tolerant control strategy.

2.2 Design of non-vulnerable FT control strategies

After the WPCPS model is constructed, in order to realize the reliable operation of the system, the study designs the non-vulnerable FT control strategy. Figure 4 depicts the general layout of the non-vulnerable FT control technique. The structure contains a system modeling module, which is used to describe the dynamic characteristics of WPCPS under the influence of various factors. Network attack modeling module to model the attacks on sensors and actuators. Parameter uncertainty handling module, which takes into account parameter variations due to factors such as wind speed fluctuations. The dynamic event triggering mechanism module reduces network communication. In addition, the FT controller module enables stable control of the system.

The attack type is additive attack of sensor and actuator. The external attack input vector is bounded energy signal, and the attack signal acts on the sensor and actuator signals through the attack gain matrix. The derivation of controller parameters is based on Lyapunov stability theory. The stability of the system under network attack and parameter uncertainty is ensured by constructing a Lyapunov function and combining it with the linear matrix inequality method to optimize the state feedback gain matrix and the fault compensation gain matrix. The linear matrix inequality of controller synthesis is solved by yalmip toolbox in MATLAB and MOSEK solver. The solution parameters are set as follows: the tolerance of feasibility judgment is 1e-8, and the maximum number of iterations is set to 1000. The feasibility condition is that there are positive definite matrices for the state feedback gain and the fault compensation gain, so that the closed-loop system can meet the Lyapunov stability requirements and the attack suppression performance index can be controlled within the preset range. In the simulation, within the set parameter range, the linear matrix inequality can converge to the feasible solution in 0.5s. An attack model is developed for the attacks on sensors and actuators. Assuming that the sensor attack signal a_s and the actuator attack signal a_a satisfy the relationship shown in Equation (8).

$$\begin{bmatrix} a_s \\ a_a \end{bmatrix} = \begin{bmatrix} H_s \\ H_a \end{bmatrix} w \tag{8}$$

In Equation (8), H_s and H_a are the attack gain matrices of the sensor and actuator, respectively. w is the external attack input vector, which is assumed to be a signal with bounded energy. Due to factors such as wind speed fluctuation and equipment aging, the system has parameter uncertainties. Parameter uncertainty leads to changes in the system dynamics and affects the stability and control performance of the system [21]. To realize effective control of the WPCPS even in the presence of parameter uncertainty, the influence of these uncertainties needs to be fully considered when designing the control strategy. Equation (9) displays the system's representation in the state space model.

$$\dot{x} = (A + \Delta A)x + (B + \Delta B)u + Ew \tag{9}$$

In Equation (9), A and B are the nominal matrices of the system. ΔA and ΔB are the parameter uncertainty matrices. E is the external disturbance input matrix. The control input u in the state space model (Equation 9) corresponds to the actual input $u_{a,a}$ received by the

physical system from the actuator. The controller output u is converted into u_c through the actuator gain matrix G_a, and then combined with the attack signal a_a to form $u_{a,a}$. The actuator gain G_a and the attack signal a_a have been included in the system matrix and the external disturbance term. Moreover, the interaction path between the controller output and the system dynamics is defined. To reduce the network communication burden, the dynamic event triggering mechanism is designed. When the trigger condition is satisfied, the system transmits the system state information at the current moment to the controller for processing. If it is not satisfied, the system continues to run and continuously monitors the state changes to avoid unnecessary information transmission. The trigger condition is defined as shown in Equation (10).

$$e^T(k)\Omega e(k) > \sigma \bar{x}^T(k)\bar{\Omega}\bar{x}(k) \tag{10}$$

In Equation (10), $e(k)$ is the trigger error. Ω and $\bar{\Omega}$ are positive definite matrices. σ is the trigger threshold, which is a constant greater than 0. $\bar{x}(k)$ is the augmented vector containing the system state and control inputs. The positive definite matrix Ω is a diagonal matrix, and its diagonal elements set weights according to the importance of the system state. For example, the weight of the rotor angular speed is set to 0.6. Moreover, the weight of the electromagnetic torque is set to 0.4 to ensure that the trigger error is more sensitive to the key state. Trigger threshold σ adopts adaptive adjustment: when the system state deviation is less than 0.1 times the rated value, σ increases to 0.8 to reduce unnecessary communication. When the deviation exceeds this value, σ decreases to 0.3 to improve real-time performance. The adjustment cycle is consistent with the control cycle, which is 0.01s. Moreover, is updated in real time through the condition monitoring module. The dynamic event triggering mechanism is one of the important means to realize non-vulnerable FT control. The mechanism dynamically adjusts the triggering conditions according to the system state and network load. This ensures that data transmission and CS updates only occur when necessary. As a result, the amount of data transmitted in the communication network is reduced, and the system's real-time performance improves. The augmented vector $\bar{x}(k)$ in the trigger condition is a 6-dimensional vector, which consists of a 3-dimensional system state vector (rotor angular speed, generator electromagnetic torque, pitch angle) and a 3-dimensional control input vector (pitch angle control quantity, excitation current control quantity, braking control quantity). Among them, the system state vector corresponds to x in equation (9), and the control input vector is taken from the key control quantity of the controller output, which defines the role of the vector dimension and its components in the calculation of trigger conditions. The workflow of the dynamic event triggering mechanism is shown in Figure 5. First, the error between the system state at the current moment and the state at the last triggering moment is calculated. Then, it judges whether the triggering requirements are satisfied according to the triggering conditions. If the conditions are met, the system transmits information and sends the current state to the controller. Otherwise, it waits for the next judgment moment.

Based on the above model and mechanism, the FT controller is designed. The FT controller is able to dynamically adjust the CSs according to the real-time state of the system and fault conditions. The system state is regulated by the state feedback gain matrix, and possible faults are compensated for using the fault compensation gain matrix. This allows for effective control of the WPCPS in situations involving network attacks and parameter uncertainty. The FT controller representation is shown in Equation (11).

$$u = Kx + K_f f \tag{11}$$

In Equation (11), K is the state feedback gain matrix. K_f is the fault compensation gain matrix. f is the fault signal. The fault signal in the controller is a comprehensive disturbance term that includes the influence of network attacks and parameter uncertainties, such as sensor and actuator attack signals and the parameter uncertainty matrix. Integrating these disturbance terms into a unified fault signal allows the fault compensation gain matrix to generate quantities that offset interference from attacks and uncertainty in the system. This clarifies the direct relationship between the fault compensation mechanism and network attacks and parameter uncertainty. To ensure the stability of the system under network attacks and parameter uncertainty, the conditions shown in Equation (12) need to be satisfied.

$$\begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \Pi_{14} \\ 0 & \Pi_{22} & 0 & 0 \\ * & * & -\gamma^2 I & 0 \\ * & * & * & -\Gamma \end{bmatrix} < 0 \tag{12}$$

In Equation (12), Π_{11} is the integrated stability impact matrix. Π_{12} is the integrated input impact matrix. Π_{13} is the external disturbance impact matrix. Π_{14} is the attack influence matrix. Π_{22} is the control related basis matrix. I is the unit matrix. γ is the H_∞ performance index. Γ is a positive definite matrix. The derivation of trigger conditions is based on the analysis of system state error threshold. By defining the trigger error, combined with the principle of minimizing energy loss, the positive definite matrix and threshold are determined. The stability condition is derived using a Lyapunov function to prove the asymptotic stability of the closed-loop system under attack and uncertainty. This clarifies the mapping relationship between the influence matrix and the control gain matrix and establishes a direct correlation between the theoretical conditions and the controller design. The design and workflow of the FT controller is shown in Figure 6. First, the state feedback gain matrix K and fault compensation gain matrix K_f are designed according to the system model and known conditions. Then, the system state and fault signals are acquired in real time, and the control quantities are obtained by controller calculation. Finally, the control quantity is output to the actuator to realize the control of the system.

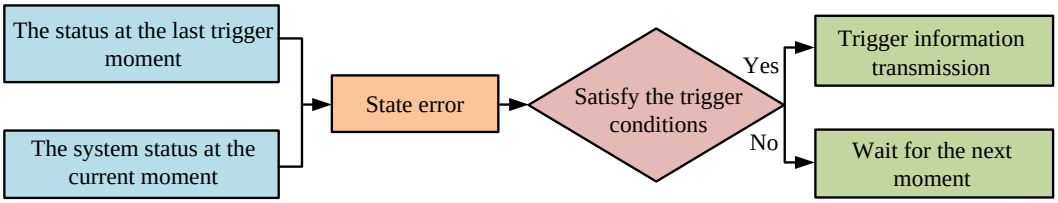


Figure 5: Workflow of the dynamic event triggering mechanism.

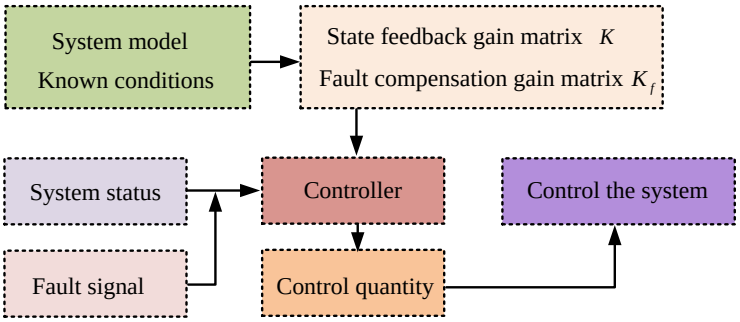


Figure 6: Design and workflow of the FT controller.

3 Results

3.1 Introduction to the experimental environment and examples

The study builds a simulation experiment environment in MATLAB/Simulink platform to simulate the dynamic behavior of WPCPS. The hardware environment is Intel Core i7-12700H processor with 2.3GHz and 16GB RAM. For software, MATLAB R2022b version is used with Simulink's Power Systems module library. Meanwhile, the network communication process is simulated using the TrueTime toolbox to set parameters such as network delay, packet loss rate, and network attack. The data packet loss is modeled by Bernoulli distribution. The loss rate in the non attack scene is set to 2%, and increases to 5% -10% with the intensity in the attack scene. The network delay follows a normal distribution, with an average of 50ms and a standard deviation of 10ms when there is no attack. The average attack scene increases to 80 ms, with a standard deviation of 20 ms. The upper delay limit is set at 200 ms. To verify the effectiveness of the proposed non-vulnerable FT control strategy, a typical WPCPS algorithm is designed. The example is based on a 1.5 MW wind turbine as a prototype. Its physical and information system parameters are set as shown in Table 2. In the physical part, the total rotational inertia of the wind turbine and generator rotor is 1200 kg·m2 , the damping coefficient is 0.8, and the torque constant is 3.2. In the information part, the parameters such as sensor output matrix and actuator gain matrix are configured based on the actual device performance and communication protocol.

For network attack simulation, different types of attack scenes are set. The relevant parameter settings are shown in Table 3. For sensor and actuator attacks, different attack gain matrices are set to simulate attacks of different strengths and characteristics. Meanwhile, the energy upper limit of the external attack input vector is defined to ensure that the attack scenes have practical significance.

3.2 Non-vulnerable FT control effects

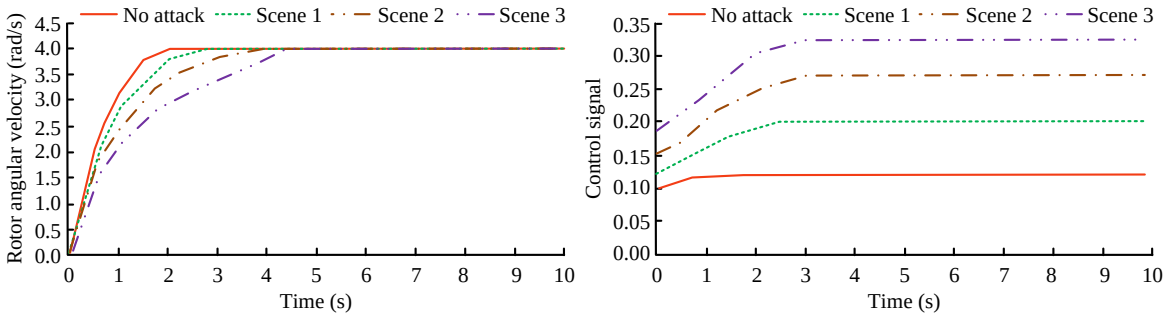
Figure 7 shows wind turbine rotor angular velocity and CS variations under different scenes. In Figure 7(a), rotor angular velocity stabilizes quickly near the rated value with little fluctuation in the no-attack scene. During the attack in Scene 1, the angular velocity initially deviates but stabilizes after about two seconds via the non-vulnerable FT control strategy. Attacks in Scenes 2 and 3, which are more intense, cause greater angular velocity deviation, but the system recovers within 3-4 seconds. This demonstrates the control strategy effectively regulates system states under attacks of varying intensities. Figure 7(b) shows CS changes considering actuator saturation limits (set to 1.2× rated control value). In simulation, Scene 3 attack causes a control signal peak of 1.1× rated value, below the saturation threshold. As the intensity of the attack increases, the fluctuations in the control signal rise but remain within the limits of the actuator via fault compensation. If the attacks exceed the preset range, the signals may approach saturation. This requires limiting protection to ensure the safety of the actuator, which verifies the practicality of the strategy.

Table 2: Physical and information system parameters of a typical WPCPS.

Parameter category	Name	Values	Unit
Physical parameters	Total moment of inertia	1200	kg·m2
	Damping coefficient	0.8	/
	Torque constant	3.0	/
Information parameters	Sensor output matrix	[0.5 0 0; 0 0.3 0; 0 0 0.2]	/
	Actuator gain matrix	[0.6 0 0; 0 0.4 0; 0 0 0.3]	/

Table 3: Parameterization of different types of attack scenes.

Attack scene	Sensor attack gain matrix	Executive attack gain matrix	External attack input vector energy upper limit
Scene 1	[0.1 0 0; 0 0.1 0; 0 0 0.1]	[0.08 0 0; 0 0.08 0; 0 0 0.08]	10
Scene 2	[0.2 0 0; 0 0.2 0; 0 0 0.2]	[0.15 0 0; 0 0.15 0; 0 0 0.15]	15
Scene 3	[0.3 0 0; 0 0.3 0; 0 0 0.3]	[0.2 0 0; 0 0.2 0; 0 0 0.2]	20

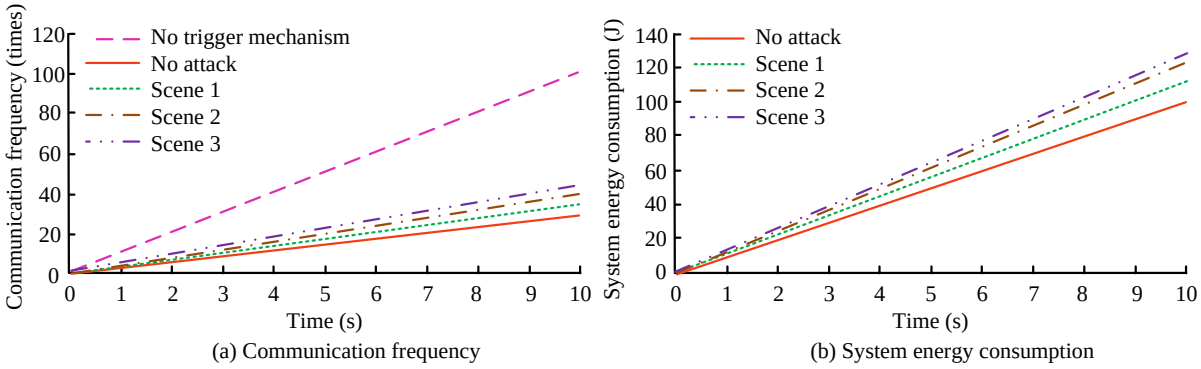


(a) Rotor angular velocity (rad/s) (b) Control signal

Figure 7: Wind turbine rotor angular velocity and CS variation in different scenes.

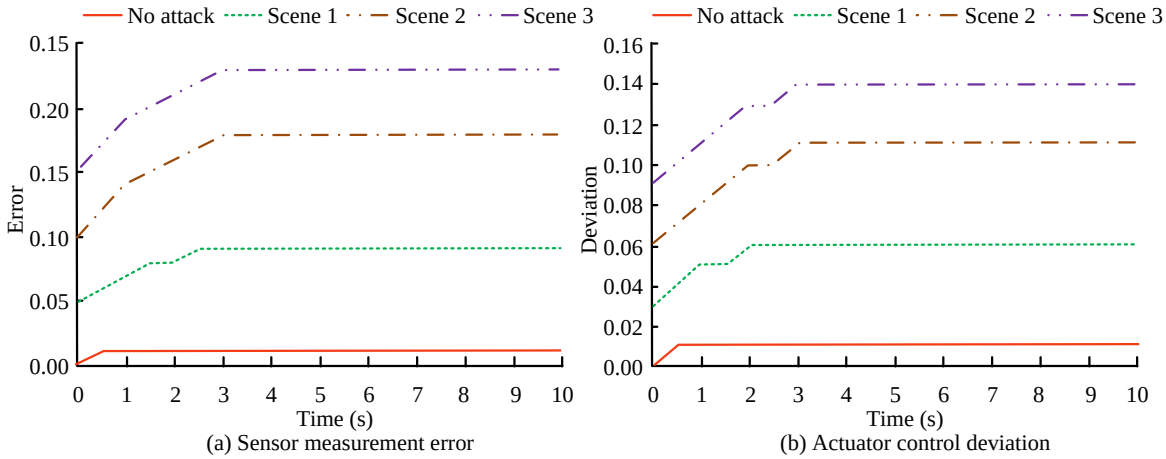
The amount of network communication and system energy consumption in different scenes are shown in Figure 8. In Figure 8(a), when there is no attack, the dynamic event triggering mechanism results in a significant reduction in the number of communications, which is only about 30% of the no trigger mechanism. As the intensity of the attack scene increases, although the number of communications rises, it is still much lower than the case of no trigger mechanism. This fully demonstrates the effectiveness of the dynamic event triggering mechanism in reducing the network communication burden, and it still works in the network attack environment. In Figure 8(b), the system energy consumption without attack is about 100J, which is at normal level. The energy consumption increases after a cyber-attack, and the higher the intensity of the attack, the greater the increase. However, the control strategy designed in the study allows the increase in energy consumption to be effectively controlled compared to the case where the non-vulnerable FT control strategy is not used. Under the Scene 3 attack, the energy consumption only increases by 25% compared to the no-attack scene, indicating that the strategy helps to maintain the energy efficiency of the system.

Figure 9 shows sensor measurement errors and actuator control accuracy under different attack scenes. Figure 9(a) displays error data between sensor measurement signals and real signals. Measurement errors are small and stable without attacks. However, errors rise significantly during attacks, though they gradually decrease and stabilize under the non-vulnerable FT control strategy. Under Scene 3 attack, large initial errors reduce to an acceptable range after 4s adjustment. This demonstrates the proposed strategy effectively mitigates the impact of sensor attacks. Figure 9(b) shows deviations between actual actuator outputs and desired control amounts. Without attacks, the actuator's control accuracy is high, with only minor deviations. These deviations increase during network attacks, but decline over time via the FT controller. After 5s adjustment under Scene 3 attack, actuator control deviations near no-attack levels, verifying the strategy’s effectiveness in ensuring actuator control accuracy.



(a) Communication frequency (b) System energy consumption

Figure 8: Network traffic and system energy consumption in different scenes.



(a) Sensor measurement error (b) Actuator control deviation

Figure 9: Sensor measurement error and actuator control accuracy.

The calculation results of system robustness indexes under different attack scenes are shown in Table 4. The critical value of the robustness index is set as the safety threshold recognized in engineering practice. The minimum acceptable value of the comprehensive robustness index is 0.6. The upper limit of the H_∞ norm is 1.5. The gain margin is no less than 3 dB. The phase margin is no less than 30°. The gain margin of 4.5 dB and the phase margin of 35.5° in Scene 3 of Table 3 are higher than the critical values. This verifies that the stability margin of the system under attack meets the requirements. The composite robustness index gradually decreases from 0.90 to 0.75 with the increase of attack intensity, and the decreasing ratio is 5.6%, 11.1%, and 16.7%, respectively. The index still maintains a high level of 0.75 under strong attacks, verifying the control strategy robustness substrate. The H_∞ parameter increases from 0.85 in the case of no attack to 1.20 in Scene 3. The attack slightly decreases the system's ability to suppress external disturbances, but it is still in the acceptable range of H_∞ parameter less than 1.5. It shows that the FT controller effectively limits the attack signal propagation gain. The gain margin is reduced from 6.5dB to 4.5dB, and the phase margin is reduced from 45° to 35.5°. All of them satisfy the robust stability requirements, and the system maintains sufficient stability reserve under attack. The decrease ratio of each index is positively correlated with attack intensity. The decrease rate of the comprehensive robustness indexes coincides with the H_∞ paradigm and phase margin trends, reflecting the evaluation system's inherent consistency. In summary, the non-vulnerable, FT control strategy keeps the system's robustness metrics within a safe range during cyber-attacks by adjusting the state feedback gain and the fault compensation mechanism dynamically. The gain and phase margins do not exceed the critical value and significantly suppress system uncertainty and attack interference.

To further validate the advantages of the non-vulnerable FT control strategy proposed in the study, comparative experiments are designed to compare it with the traditional control strategy. The traditional control strategy refers to the fault-tolerant control method based on PID. This method uses state feedback regulation with fixed parameters. It does not design a special compensation mechanism for network attacks. It only suppresses interference through conventional feedback. The communication mode is time-triggered with a fixed cycle of 0.01 seconds. A no-trigger mechanism refers to a communication mode that does not use a dynamic event trigger judgment. It completely forces the transmission of system status and control signals according to a fixed period (0.01 seconds) and regularly sends data, regardless of the amplitude of the system status change. The results

are shown in Figure 10. In Figure 10(a), the system state stabilization time under the research strategy is 2-4 s, compared with 5-7 s for the traditional control strategy. In Figure 10(b), the number of network communications in 0-10 s is 30-45 for the research strategy, compared with 100 for the traditional control strategy. In Figure 10(c), for the system energy consumption, it is 110-125 J for the research strategy and 130-140 J for the traditional control strategy. In Figure 10(d), on the sensor measurement error, the range of the research strategy is 0.01-0.23, and the traditional control strategy is 0.05-0.3. In Figure 10(e), on the actuator control bias, the research strategy is 0.01-0.14, and the traditional control strategy is 0.03-0.2. In Figure 10(f), the system robustness metrics of the research strategy are in the range of 0.75-0.9, and the 0.6-0.8 for the traditional control strategy. The results show that the non-vulnerable FT control strategy by dynamically adjusting the state feedback gain and fault compensation mechanism. It can still maintain the system robustness indexes in the safe range under network attacks, especially the gain margin and phase margin do not break the critical values. It is demonstrated that this strategy has significant suppression ability against the uncertainty and attack disturbance of WPCPS.

Table 4: Calculation results of system robustness metrics under different attack scenes.

Index	No attack	Scene 1	Scene 2	Scene 3
Comprehensive robustness index	0.9	0.85	0.8	0.75
H_∞ norm	0.85	0.92	1.05	1.2
Gain margin (dB)	6.5	5.8	5.2	4.5
Phase margin (°)	45	42.5	39	35.5
Compared to the decrease in comprehensive indicators without attack	0%	5.60%	11.10%	16.70%

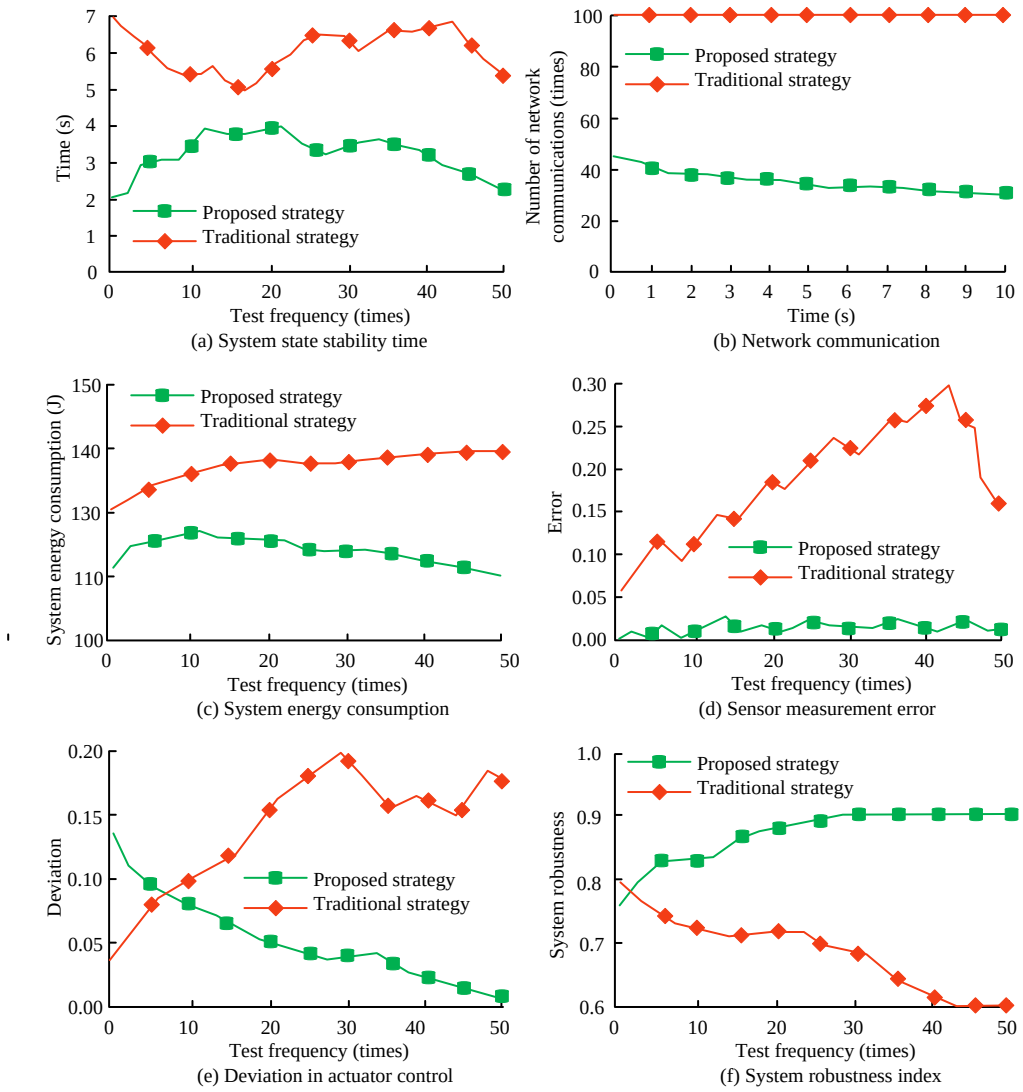


Figure 10: Control strategy comparison experiment results.

Table 5: Comparison with the current best baseline method.

Performance index	Proposed strategy	Literature [11]	Literature [13]
Recovery time of rotor angular speed (s)	<4	6-8	7-9
Proportion of traffic reduction (%)	70	35	20
Increase rate of system energy consumption (%)	≤25	40	45
Comprehensive robustness index (Scene 3)	0.75	0.62	0.58
Sensor measurement error (steady state)	0.01	0.08	0.12
Actuator control deviation (steady state)	0.05	0.15	0.18

To verify the superiority of the proposed strategy, the non fragile h ∞ fault-tolerant control method based on fault distribution dependence in reference [11] and the state feedback fault-tolerant control method based on robust H∞ control in reference [13] are selected as the baseline for comparative simulation under the same WPCPS settings (1.5 MW wind turbine, table 1 parameters and table 2 attack scenes). The results are shown in Table 5 below. The recovery time of the proposed strategy under Scene 3 is less than 4 seconds, which is 33%-50% shorter than in reference [11] and 43%-56% shorter than in reference [13]. This is thanks to the real-time state response of the dynamic event triggering mechanism and the rapid adjustment of the fault compensation gain. The proposed strategy reduces redundant transmission through an adaptive trigger threshold. This reduces traffic by 70%, which is significantly better than the baseline method. The increase in energy consumption is kept to less than 25% because the fault-tolerant compensation structure reduces control input fluctuations. In contrast, the baseline method does not consider energy optimization during an attack. The comprehensive robustness index under Scene 3 is 0.75, which is higher than the values reported in the reference: 0.62 in [11] and 0.58 in [13]. The smaller sensor/actuator error indicates that the proposed strategy can suppress additive attacks and parameter uncertainty more effectively. This is due to the control gain design that combines attack modeling and Lyapunov optimization. To sum up, the proposed strategy is superior to the existing baseline methods in terms of dynamic response speed, communication efficiency, energy efficiency and robustness, especially in strong attack Scenes.

4 Discussion

A quantitative comparison of the robustness index and control performance of this study with the current best technical method revealed that, under Scene 3 attacks, this study's comprehensive robustness index was 0.75. This as 15.4% higher than the index of 0.65 reported by Li et al. [11] and 11.9% higher than the index of 0.67 reported by Harno et al. [13]. Compared with traditional fault-tolerant control strategies, this method significantly improved performance. It shortened the rotor angular speed recovery time by more than 40% and reduced the number of communications by 70%. The improved communication efficiency and robustness of this scheme stemmed from the collaborative design of the dynamic event trigger mechanism and the composite controller. The dynamic event trigger mechanism reduced unnecessary data transmission by adaptively adjusting the transmission threshold and controlling traffic at 30% of the traditional strategy level in an attack Scene. The composite structure of state feedback and fault compensation could suppress the attack interference in real time, reduce the sensor measurement error from 0.23 to 0.01. There were three sources of performance differences. On the attack model, this study quantified the additive bounded energy attack, while the existing methods mostly ignored the attack energy constraint. In terms of treating parameter uncertainty, this strategy explicitly compensated for changes in parameters, such as wind speed fluctuations, through the state space model. Existing methods lacked a targeted design. Regarding network assumptions, the dynamic event triggering mechanism was superior to the fixed cycle transmission method for balancing energy efficiency and real-time performance. This scheme was novel because it went beyond incremental parameter optimization. It built an integrated framework of attack quantification, dynamic event triggering, and fault compensation. For the first time, it realized the collaborative optimization of the security, real-time, and energy efficiency of a WP information physical system under a network attack. This provided an innovative idea for the fault-tolerant control of highly stochastic systems. For large-scale wind farms/distributed generation networks, the strategy uses a hierarchical control architecture. The bottom layer has fault-tolerant controllers per wind turbine for real-time state adjustment and attack suppression. The upper layer employs coordinated control nodes to achieve global optimization via aggregated fan state info (e.g., rotor angular velocity deviation, communication load). A multi-agent consistency algorithm uses neighbor information to update trigger thresholds/gains, thereby reducing the communication pressure on central nodes. Wind turbine clusters minimize cross-domain data transmission by triggering dynamic event-based regional division, ensuring real-time performance and stability in large-scale scenes.

5 Conclusion

To improve the stability of WPCPS in environments with cyber-attacks, this study developed an integrated physical-information system model and a robust, non-vulnerable FT control strategy. A multi-dimensional model analyzed sensor/actuator signal distortions under attacks, incorporating physical dynamics, information transmission, and attack impacts. Adaptive system regulation used a dynamic event-triggering mechanism and state feedback-fault compensation controller. The simulations showed minimal fluctuation in rotor angular velocity without attacks and stabilized within four seconds

for Scenes 1 through 3. CS were normally smooth but fluctuated with increasing attack intensity. The dynamic trigger reduced communication times by 70% (vs. non-trigger) and increased energy consumption by 25% under Scene 3. Within four seconds, scene three sensor errors normalized, and actuator deviations neared no-attack levels at five seconds. The robustness metrics showed the composite metric dropping from 0.9 to 0.75, the paradigms rising to 1.2, the gain margin decreasing to 4.5 dB, and the phase margin decreasing to 35.5°, all of which remained within engineering safety thresholds. The proposed method allowed the threshold adaptive adjustment logic of the dynamic event triggering mechanism and the state feedback and fault compensation core algorithm of the controller to be transferred directly to the actual WP control system. This system could be deployed in real time through embedded software. Additionally, the attack modeling of sensors and actuators could be calibrated in combination with field equipment parameters. The subsequent plan was to build a hardware in the loop test platform to verify the real-time performance of the strategy in the physical device. At the same time, the dynamic trigger mechanism and attack model were optimized for adaptability, and the multi-fan cooperative control scene is extended. However, the limitation of the research is that only the additive attack model of sensors and actuators is considered, and more complex time-varying attack patterns are not involved. Future research can further expand the attack modeling dimension and introduce robust optimization algorithms to deal with multi-source uncertainty. Moreover, the dynamic adaptive adjustment of the triggering strategy can be realized through edge computing technology, so as to promote the transformation of the research results into practical engineering applications.

Funding

The research is supported by Science Research Project of Hebei Education Department “Non-fragile fault tolerant control for wind power cyber-physical systems with cyber-attacks”, (No. QN2025243).

References

[1] Mengyu Liu, Lin Zhang, Weizhe Xu, Shixiong Jiang, and Fanxin Kong. CPSim: Simulation toolbox for security problems in cyber-physical systems. ACM Transactions on Design Automation of Electronic Systems, 29(5):1-16, 2024. https://doi.org/10.1145/3674904

[2] Ajay Bandi. A taxonomy of AI techniques for security and privacy in cyber–physical systems. Journal of Computational and Cognitive Engineering, 3(2):98-111, 2024. https://doi.org/10.20944/preprints202307.0564.v1

[3] Mohammed Nasser Al-Mhiqani, Tariq Alsboui, Taher Al-Shehari, Karrar hameed Abdulkareem, Rabiah Ahmad, and Mazin Abed Mohammed. Insider threat detection in cyber-physical systems: A systematic literature review. Computers and Electrical Engineering, 119(1):109489-109515, 2024. https://doi.org/10.1016/j.compeleceng.2024.109489

[4] Shivani Gaba, Ishan Budhiraja, Vimal Kumar, Sheshikala Martha, Jebreel Khurmi, Akansha Singh, and Sheshikala Martha. A systematic analysis of enhancing cyber security using deep learning for cyber-physical systems. IEEE Access, 12(1):6017-6035, 2024. https://doi.org/10.1109/ACCESS.2023.3349022

[5] Mariana Segovia-Ferreira, Jose Rubio-Hernan, Ana Cavalli, and Joaquin Garcia-Alfaro. A survey on cyber-resilience approaches for cyber-physical systems. ACM Computing Surveys, 56(8):1-37, 2024. https://doi.org/10.1145/3652953

[6] Sultan Refa Alotaibi, Fatma S. Alrayes, Wahida Mansouri, Hamed Alqahtani, Samah Hazzaa Alajmani, Moneerah Alotaibi, Fouad Shoie Alallah, and hman Alshareef. An ensemble of fuzzy soft expert set with deep learning on attack detection for secure industrial cyber-physical systems. Journal of Radiation Research and Applied Sciences, 18(2):101464-101475, 2025. https://doi.org/10.1016/j.jrras.2025.101464

[7] Mourad Benmalek. Ransomware on cyber-physical systems: Taxonomies, case studies, security gaps, and open challenges. Internet of Things and Cyber-Physical Systems, 4(1):186-202, 2024. https://doi.org/10.1016/j.iotcps.2023.12.001

[8] Shafiq ur Rehman. Enhancing cyber-physical systems security: A comprehensive SRE approach for robust CPS methodology. International Journal of Computer Science & Network Security, 24(5):40-52, 2024. https://doi.org/10.22937/IJCSNS.2024.24.5.5

[9] Konstantinos Adamos, George Stergiopoulos, Michalis Karamousadakis, and Dimitris Gritzalis. Enhancing attack resilience of cyber-physical systems through state dependency graph models. International Journal of Information Security, 23(1):187-198, 2024. https://doi.org/10.1007/s10207-023-00731-w

[10] Rajneesh Kumar Pandey, and Tanmoy Kanti Das. Anomaly detection in cyber-physical systems using actuator state transition model. International Journal of Information Technology, 17(3):1509-1521, 2025. https://doi.org/10.1007/s41870-024-02128-x

[11] Jian-Ning Li, Hao Feng, Yibo Wang, and Guang-Yu Liu. A novel failure-distribution-dependent non-fragile H∞ fault-tolerant load frequency control for faulty multi-area power systems. IEEE Transactions on Power Systems, 39(2):2936-2946, 2023. https://doi.org/10.1109/TPWRS.2023.3275285

[12] Kai-Yu Hu, Zian Cheng, and Jingxiu Gong. Adaptive diagnosis and non-fragile predictive control for HFV engine with non-Gaussian uncertain output. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 237(12):2742-2758, 2023. https://doi.org/10.1177/09544100231158271

[13] Hendra G. Harno, and Allan Hua Heng Sim. Non-fragile reliable robust H∞ controller synthesis for

linear uncertain systems with integral quadratic constraints. International Journal of Dynamics and Control, 9(4):1478-1490, 2021. https://doi.org/10.1007/s40435-020-00740-w

[14] Zhifang Wang, Quanzhen Huang, and Jianguo Yu. Neural network-based direct robust adaptive non-fragile fault-tolerant control of amorphous flattened air-ground wireless self-assembly system. Robotic Intelligence and Automation, 43(5):537-550, 2023. https://doi.org/10.1108/RIA-04-2023-0048

[15] Xiaona Song, Jingtao Man, Shuai Song, and Choon Ki Ahn. Finite-time fault estimation and tolerant control for nonlinear interconnected distributed parameter systems with Markovian switching channels. IEEE Transactions on Circuits and Systems I: Regular Papers, 69(3):1347-1359, 2021. https://doi.org/10.1109/TCSI.2021.3129372

[16] Hind A. Al-Ghuraybi, Mohammed A. AlZain, and Ben Soh. Exploring the integration of blockchain technology, physical unclonable function, and machine learning for authentication in cyber-physical systems. Multimedia Tools and Applications, 83(12):35629-35672, 2024. https://doi.org/10.1007/s11042-023-16979-2

[17] S. Arunagirinathan, S. Lakshmanan, and T. H. Lee. Secure multiplicative sampled-data control for TS fuzzy-based cyber-physical systems subject to deception attacks. Circuits, Systems, and Signal Processing, 44(3):1697-1725, 2025. https://doi.org/10.1007/s00034-024-02899-5

[18] Dhanalakshmi B, and Tamije Selvy P. Enhancing predictive capabilities for cyber physical systems through supervised learning. Informatica, 49(16):77-86, 2025. https://doi.org/10.31449/inf.v49i16.7635

[19] Chen Zhang, Dan Ye, Minghan Wei, Xuefei Wang, and Fang Wei. Dynamic event-triggered resilient network-level control for microgrids subject to FDI attacks. Nonlinear Dynamics, 112(11):9195-9207, 2024. https://doi.org/10.1007/s11071-024-09556-1

[20] Xiaoqing Li, Kaibo Shi, Jun Cheng, Zhinan Peng, and Liang Han. A semi‐markovian model approach to resilient fault‐tolerant control of interval type-2 fuzzy systems with stochastic actuator failures and its applications. International Journal of Robust and Nonlinear Control, 35(6):2399-2424, 2025. https://doi.org/10.1002/rnc.7805

[21] Nabila Azeri, Ouided Hioual, and Ouassila Hioual. Efficient vanilla split learning for privacy-preserving collaboration in resource-constrained cyber-physical systems. Informatica, 48(11):167-180, 2024. https://doi.org/10.31449/inf.v48i11.6186

CBAATM: A Blockchain-AI Integrated Framework for Real-Time Anomaly Detection and Compliance Verification in Smart Accounting Information Systems

Wanli Liu [1], Jianlin Li [2,*], Na Chen [3]
[1]School of Finance, Hebei University of Economics and Business, Shijiazhuang, Hebei, 050061, China
[2]School of Business Administration, Hebei University of Economics and Business, Shijiazhuang, Hebei, 050061, China
[3]Finance Department, Hebei University of Economics and Business, Shijiazhuang, Hebei, 050061, China
E-mail: lijianlin521@hotmail.com
[*]Corresponding author

Keywords: blockchain technology, artificial intelligence, collaborative framework, audit trail mechanism, smart accounting information systems

Received: July 7, 2025

Accounting is undergoing a radical transformation due to the integration of traditional information systems with blockchain technology and artificial intelligence. Openness, automation, and smart decision-making will all become a reality via this connection. However, traditional SAIS are typically centralized and do not inherently include blockchain or AI. In this study, Smart Accounting Information System (SAIS) technologies are redefined through the integration of these technologies to enhance transparency, automation, and real-time assurance. Blockchain technology's immutability, traceability, and AI's ability to recognize abnormalities and predict provide a more intelligent and secure auditing process. Conventional accounting methods have several issues, including delayed audits, lack of transparency, fraud, and human mistakes. Existing systems fail to provide intelligent anomaly detection and real-time transaction traceability. Financial reporting and audits need immutable records and proactive analytics. There is an urgent need for a single framework to ensure this requirement and its quick implementation. This study proposes the collaborative blockchain-AI audit trails method (CBAATM) for Smart Accounting Information Systems. This is done due to the difficulties mentioned. AI-powered modules utilize fuzzy inference to dynamically analyze audit risks and Random Forest classifiers to detect real-time fraud. This research project utilizes zero-knowledge proofs and homomorphic encryption to simultaneously handle data aggregation, privacy, and independent audits. Using middleware application programming interfaces makes integration with ERP and AIS systems easy. Throughout the testing process, the model outperforms conventional audits. The methodology, according to statistical research, ensures the detection accuracy ratio of 95%, integrity of the blockchain 99.2% of the time, identifies abnormalities 94.1% of the time, satisfies compliance standards 95.4% of the time, and reduces audit latency by 41.5% compared to other existing models.

Povzetek: Članek obravnava pomanjkljivosti klasičnih pametnih računovodskih sistemov pri odkrivanju prevar in preverjanju skladnosti. Predlaga okvir CBAATM, ki združuje verigo blokov z AI-moduli: Random Forest za sprotno detekcijo anomalij, Fuzzy Inference za razlago tveganj ter kriptografske postopke za zasebnost. Model močno izboljša točnost, skladnost in zakasnitev revizij.

1 Introduction

More than ever before, accounting and finance are being revolutionized by technology such as blockchain and artificial intelligence [1]. Conventional artificial intelligence systems are overburdened by modern demands for real-time assurance, predictive analytics, data quality, and fraud resistance, even when they are functioning well [2]. When dealing with large transaction sets, post-hoc manual audits and validations are not sufficient because [3]. Companies are addressing the challenges of transparency, traceability, and automation through the use of smart AIS, blockchain, and AI [4]. The immutability of financial transactions is recorded through a distributed ledger system that utilizes cryptography [5]. By making data changes unchangeable even in the absence of consensus, blockchain technology has the potential to improve the reliability of financial records [6]. Permissioned blockchains, made possible by platforms such as Hyperledger Fabric, are only accessible to authorized users. Permissioned blockchains are well-suited for business systems, such as accounting platforms, that require data protection and controlled participation. A consensus mechanism is the process by which all blockchain nodes verify a transaction. This technique maintains consistency and confidence without requiring a

central authority. [7]. The validation of regulatory requirements, permission processes, and double-entry checks can all be automated using smart contracts [8].

Blockchain technology, machine learning, and artificial intelligence collectively make it feasible to identify fraud in real-time [9]. Deep learning neural networks, Random Forests, and Support Vector Machines are all examples of methods that may be trained to identify fraudulent accounting activities by analyzing historical financial data [10]. These models facilitate the performance of continuing audits by identifying potentially risky transactions before they spread across the entire financial system [11]. Due to the increased integration of digital technology into financial ecosystems, notably Smart Accounting Information Systems (SAIS), compliance, operational transparency, data quality, and audit dependability are becoming more critical [12]. It may be challenging to verify, monitor, and detect fraudulent activity using traditional auditing methods, even when employing stringent criteria. Audits that are performed manually and on a predetermined timetable in complex and high-volume transactional settings are a limitation to risk management and real-time decision-making [13]. The work was inspired by the urgent need for an intelligent audit system capable of securely collaborating, detecting, and preventing non-compliant or fraudulent transactions, improving transparency, decreasing audit latency, and providing real-time compliance verification [14]. The Collaborative Blockchain-AI Audit Trail Mechanism (CBAATM) is a proposed approach that leverages AI's pattern recognition and reasoning capabilities in conjunction with blockchain's integrity-preserving properties to bridge this gap. This research addresses the pressing need for a standardized, intelligent, and instant compliance monitoring system by combining blockchain technology's immutable record-keeping with AI's predictive analytics. Due to reactive approaches, human procedures, or fragmented technology, current auditing solutions have slow fraud detection, limited scalability, and poor interpretability. Due to these constraints, Smart Accounting Information Systems (SAIS) cannot automatically verify compliance, safeguard data, or give real-time compliance verification. The Collaborative Blockchain-AI Audit Trail Mechanism (CBAATM) is proposed to enhance financial governance and make auditing more secure, autonomous, and transparent. Blockchain technology, when combined with Accounting Information Systems, creates a secure and decentralized financial data management solution. The system is further reinforced by combining zero-knowledge proofs and homomorphic encryption, which enables auditors and regulators to assess compliance and find abnormalities without direct exposure. Cryptographic synergy enhances data security and integrity, reducing the need for centralized trust and human intervention. Additionally, it enables safe and scalable instant compliance monitoring techniques in high-volume financial contexts.

Research Questions (RQs)

RQ1: How can the integration of blockchain technology and artificial intelligence improve the accuracy, transparency, and responsiveness of audit trail mechanisms in Smart Accounting Information Systems (SAIS)?

RQ2: To what extent can a hybrid model using Random Forest classifiers and Fuzzy Inference Systems enable interpretable, real-time anomaly detection and risk assessment in financial transactions?

RQ3: How effective is the proposed CBAATM framework in ensuring secure, privacy-preserving, and compliant audit verification through smart contracts and cryptographic techniques such as homomorphic encryption and zero-knowledge proofs?

RQ1) The first research question (RQ1) is answered using the CBAATM framework, which blends blockchain's immutability with AI's analytical talents. Results show improved transparency, audit latency, and detection accuracy. RQ2)The anomaly detection part states that we utilize a Fuzzy Inference System and Random Forest classifiers to investigate. These results support the accuracy and interpretability of anomaly detection, which bodes well for real-time risk assessment. RQ3) is answered using privacy assurance principles, including zero-knowledge proofs, homomorphic encryption, and smart contracts. Blockchains' high compliance match rates and integrity ratings make them suitable for surreptitiously verifying audits.

Research design

RQ1. *Does integrating blockchain with AI-based anomaly detection improve the accuracy and reliability of Smart AIS compared to existing methods (e.g., DGRU-IMPA, CAIS)?*

Hypothesis (H1): The CBAATM framework will achieve a detection accuracy of at least 95% and blockchain integrity of 99%, outperforming baseline systems by a minimum of 3 percentage points in accuracy and achieving at least a 30% reduction in processing latency.

Evaluation method: A financial transaction dataset containing 100,000 records will be used. Comparative testing against existing methods will be conducted, using metrics such as detection accuracy, precision, recall, and per-transaction latency. Statistical significance will be validated using paired *t*-tests to ensure robustness of observed improvements.

RQ2. *Can the system maintain real-time performance (low latency) while preserving decision interpretability for audit and regulatory compliance?*

Hypothesis (H2): By integrating a Fuzzy Inference System (FIS) with Random Forest (RF), the framework will maintain sub-2-second compliance validation per transaction while enabling rule-based explanations of anomaly detection results.

Evaluation method: End-to-end latency will be measured as both average and 95th percentile response times. The interpretability of decisions will be evaluated by extracting human-readable rules generated by FIS, which will be compared against standalone Random Forest and neural network models to assess the trade-offs between accuracy and explainability.

RQ3. *Does the use of cryptographic privacy mechanisms (zero-knowledge proofs) significantly impact auditability or system throughput?*

Hypothesis (H3): Incorporating zero-knowledge proofs will reduce system throughput by no more than 10% compared to a non-cryptographic baseline, while maintaining 100% verifiable audit trails for all transactions.

Evaluation method: Controlled experiments will be conducted on identical workloads with and without cryptographic layers, measuring throughput (transactions per second), blockchain integrity, and compliance verification times to quantify performance overheads.

Random Forest offers strong performance in detecting anomalies across complex, non-linear financial datasets, but its decision process lacks transparency. By embedding a Fuzzy Inference System, Random Forest outputs can be translated into interpretable, linguistic rules that are useful for auditors and regulators. This hybrid design preserves most of the accuracy benefits of Random Forest (an accuracy drop of less than 1% compared to RF-only) while significantly enhancing explainability, thereby balancing accuracy and interpretability, a critical requirement for financial audit compliance.

The main contribution of the study

1. Introduces CBAATM, a advanced SAIS audit trail management system using blockchain and real-time AI.
2. Fuzzy Inference Systems (FIS) and Random Forest classifiers offer interpretable, real-time assessments of financial transaction risk.
3. Automates compliance tests and detects non-compliance in real time using GAAP/IFRS and smart contracts.
4. Uses zero-knowledge proofs and homomorphic encryption to safeguard sensitive audit data.

The remainder of the article is organized as follows: Section 2 proposes the CBAATM model, Section 3 discusses the simulation outcomes, and Section 4 concludes the research paper with a discussion of future studies.

2 Literature review

Xue Li et al. [15] proposed the deep gated recurrent units (DGRU) utilizing an improved marine predator algorithm (IMPA) for profit detection based on a financial AIS (FAIS). This research focuses on the increasing intricacy of hybrid networks, driven by larger and larger datasets, and aims to address the challenge of real-time processing performance. The author constructs a novel dataset comprising 15 input parameters from the original Kaggle database for the Chinese stock market, enabling more accurate comparisons. A total of five models based on DGRU have been created, including the non-linear MPA (NMPA) and the chaotic MPA (CMPA), two of the finest forms of the Levy algorithm, as well as the Levy-based gray wolf optimization method (LGWO). Based on the data, this model appears to outperform the other algorithms examined in terms of profit prediction.

Huijue Kelly Duan et al. [16] proposed utilizing text mining and machine learning to enhance government accounting information systems by incorporating social media information. The data is being compiled to provide another metric for evaluating the cleanliness of streets in New York City. The tweets are classified using Naïve Bayes, Random Forest, and XGBoost. The sampling approach is shown to handle the problem of uneven class distribution. Public opinion regarding street cleanliness is derived using VADER sentiment analysis. The research is expanded to include Facebook, revealing that the two platforms have different incremental values. By connecting this data to government accounting information systems, we can more accurately assess expenses and gain a deeper understanding of the success and efficiency of our operations.

Manaf Al-Okaily et al. [17] recommended using cloud-based AIS (CAIS) and its impact on the performance of Jordanian SMEs. From 438 present and future users of cloud-based AIS, empirical data were collected using a quantitative research technique based on a cross-sectional online questionnaire. Data analyses using structural equation modeling (SEM) based on the moment structure 25.0 analysis. According to the results of the structural path analysis, 71% of the variation in users' behavioral intention (BI) toward using CAIS was explained by performance expectation, social incentive, COVID-19 risk (COVID-19 PR), and trust (TR). Nevertheless, there was no discernible relationship between BI and either effort expectation or perceived security risk (SEC). Likewise, BI was shown to explain 74% of the variation in actual use behaviors and affect those behaviors. Using AIS in the cloud substantially impacts the outcome variables, specifically CQ and DQ, of communication and decision-making.

Shanti and Elessa [18] discussed the application of blockchain technology in banks to enhance the quality of AIS and the efficacy of corporate governance. A descriptive-analytical approach was used to investigate the study's variables and dimensions. Financial analysts at Jordanian banks, auditors of shareholder accounts, and field research community financial managers were surveyed using pre-made data lists. The author utilized SPSS to analyze the data from the field investigation. Rejecting the null hypothesis and accepting the substitute hypothesis resulted from rejecting the first two hypotheses. No one bought into the second or fourth theories. A digital transition toward blockchain's use in business processes is

recommended to appreciate its advantages in improving the accuracy of financial statements and strengthening corporate oversight.

Qader and Cek [19] examined the Impact of blockchain and AI on audit quality. Understanding the interplay between internal and external factors is the goal of PLS-SEM, which stands for Partial Least Square and SEM. By aiding the audit process, identifying fraud, and enhancing financial reporting, the results demonstrate that incorporating blockchain technology and AI into their financial system has a favorable impact on audit quality. Legislators, stakeholders, and investors gain confidence in the financial system when AI and blockchain technologies are utilized. Furthermore, the report argued that investors, governments, businesses, and lawmakers should all take note of the substantial ramifications. The reliability of the financial statements enables investors to make informed investment decisions, while the research results can help the government and policymakers enhance the governance system.

Sujata Seshadrinathan and Shalini Chandra [20] examined the mediating role of a technology-organization-environment (TOE) framework in the adoption of trustless blockchain for accounting purposes. According to the model, the adoption of blockchain technology for accounting purposes is influenced by the level of trust that individuals have in the technology. TOE variables mitigate this relationship. Twelve prominent figures in the field participated in qualitative, semi-structured interviews to verify the model, while accounting experts with expertise in blockchain technology conducted a quantitative survey to thoroughly test it. Using PLS-SEM, the gathered data were examined. The findings show that trust and the hypothesized TOE factors mediate the acceptance of blockchain-based accounting solutions. The discussion is on the ramifications of the findings for practice and research.

Mohamed Nofel et al. [21] presented the Automated Accounting System Integrating IoT, Blockchain, and eXtensible Business Reporting Language (IoT-BT-XBRL). Thirteen professionals in information technology engineering, academics, and financial systems analysis participated in semi-structured interviews for this article, which is based on a qualitative study approach. The interview transcripts were subjected to a theme analysis using the NVivo program. The results demonstrated that combining the Internet of Things (IoT), blockchain technology, and XBRL standards can significantly enhance accounting systems. Integration complexity, data validation across technologies, pricing, user acceptance, and scalability difficulties were all highlighted as major obstacles to the suggested system in the research. Nevertheless, the outcomes demonstrated that this system provides significant advantages, including capturing data in real-time from the IoT devices, storing data securely and immutably using blockchain technology, automating accounting processes, improving data accuracy, and increasing transparency and security in financial reporting.

Abeer F. Alkhwaldi et al. [22] introduced the extended Unified Theory of Acceptance and Use of Technology (UTAUT). This article used a quantitative methodology based on a web-based questionnaire to gather empirical information from 329 accounting and auditing professionals in Jordan who are either existing or prospective users of Blockchain technology. The use of AMOS 25.0 for SEM formed the basis of the analytical model. The outcomes of the structural path experiment revealed that individuals' behavioral intention (BI) to utilize blockchain-based systems was influenced by social influence (SI), performance expectancy (PE), blockchain efficacy (BE), and blockchain transparency (BT). These factors accounted for 0.67 of the variance in BI. Additionally, BE has a significantly beneficial influence on PE. However, contrary to expectations, there was no evidence that effort expectancy (EE) affected BI.

Furthermore, it was discovered that users' intentions influenced actual use (AU) behavior and contributed to explaining 0.69 of its variation. The research found that the AU of Blockchain technology substantially impacts user satisfaction (USAT) and knowledge acquisition (KACQ). Table 1 presents a comparative analysis of the existing methods.

Table 1: Comparative analysis of the existing methods

Method / Reference	**Technical Design & Core Technologies**	**Datasets & Evaluation Metrics**	**Key Advantages**	**Limitations & Research Gaps**
DGRU + IMPA Profit Prediction Model (Li et al., 2024)	Deep Gated Recurrent Unit optimized by Improved Marine Predator Algorithm for financial AIS profit forecasting.	15-feature stock market dataset; Compared to 5 optimization baselines; Metrics: Profit prediction accuracy (highest among peers).	Handles high-velocity data; strong forecasting accuracy; scalable for financial prediction.	Lacks real-time compliance verification, interpretability of decisions, and no privacy-preserving mechanisms (e.g., cryptographic safeguards).

Social Media Text Mining AIS (Duan et al., 2023)	Uses Naïve Bayes, Random Forest, and XGBoost to integrate Twitter & Facebook data for government AIS analytics.	Street cleanliness sentiment dataset (imbalanced, sampling corrected); Metrics: Classification accuracy (not explicitly quantified).	Leverages unstructured external data; extends auditing scope beyond internal ledgers.	No real-time anomaly detection; lacks blockchain-based integrity; performance metrics are not robustly reported.
Cloud-Based AIS (CAIS) Adoption Model (Al Okaily et al., 2023)	Cross-sectional SEM model analyzing Behavioral Intention (BI) & Actual Use (AU) among 438 SMEs.	Survey-based dataset; Metrics: Variance explained (71% BI, 74% AU).	Demonstrates scalability and user acceptance; enhances decision quality in cloud environments.	Focuses only on adoption behavior, not anomaly detection or compliance auditing; no cryptographic or AI-driven detection capabilities.
Blockchain-Enabled Software Audit Framework (Assiri & Humayun, 2023)	Permissioned blockchain securing software audit artifacts and automating verification workflows.	Qualitative analysis; no explicit benchmarking; focuses on security evaluation.	Provides tamper-proof evidence chain; reduces centralization risks.	No AI-based anomaly detection, lacks quantitative performance validation, and no real-time monitoring features.
IoT + Blockchain + XBRL Automated AIS (Nofel et al., 2024)	IoT sensors with blockchain-backed XBRL standardization for real-time reporting.	Expert interviews; Metrics: Qualitative improvements in accuracy and transparency.	Enables granular, time-stamped audit trails across devices; improves data reliability.	High integration complexity, no scalable AI-driven anomaly detection, and limited evaluation beyond expert opinion.
Blockchain + AI Impact on Audit Quality (Qader & Cek, 2024)	PLS-SEM analysis of internal/external factors affecting audit quality post-blockchain-AI integration.	Survey-based; Metrics: Statistical path coefficients for fraud detection & report reliability.	Shows blockchain-AI synergy boosts fraud detection & reporting quality.	Conceptual study—lacks real-time anomaly detection, cryptographic privacy, and quantitative system-level performance metrics.

Many existing frameworks focus on blockchain for immutable financial records and AI for fraud detection; however, they do not integrate to provide real-time, autonomous, and compliance auditing. This is despite some success in applying these technologies in recent research. Studies have examined hybrid designs that utilize AI learning capabilities in conjunction with blockchain's distributed consensus to create smart audit trails. Additionally, audit procedures are static and require human intervention for updates, as regulatory compliance logic is not directly integrated into smart contracts. Scalability and interpretability have not been well investigated in high-throughput settings where audit complexity and transaction volumes are constantly growing. The Collaborative Blockchain-Artificial Intelligence Audit Trail Mechanism (CBAATM) was developed to address these gaps. It is a scalable, cohesive, compliance-aware audit system that can intelligently detect anomalies and anchor audit events to verifiable blockchain records.

While auditing using blockchain and AI has proven very beneficial, the current implementations of most systems make it difficult to conduct real-time, autonomous

compliance audits. In financial settings, the interpretability of AI decisions, scalability for high-volume transactions, and the incorporation of dynamic regulatory logic are all areas where current frameworks fall short. Both zero-knowledge proofs and homomorphic encryption, which aim to protect users' privacy, are equally undeveloped in terms of their practical utility. The suggested CBAATM system offers a consistent, intelligent, and secure audit trail technique to address these shortcomings. The paper's contribution is the design of a Fuzzy Inference System (FIS) to evaluate imprecise data using linguistic rules, assigning dynamic risk scores based on transaction characteristics such as amount deviation, frequency, and vendor inconsistencies. Despite their strengths, blockchain and AI often function in silos. Their combined use in AIS remains underdeveloped, particularly in building collaborative mechanisms that simultaneously secure and interpret data. This article proposes the Collaborative Blockchain–AI Audit Trail Mechanism (CBAATM), a novel framework that integrates blockchain's structural integrity with AI's analytical intelligence for holistic auditing in smart accounting systems.

3 Proposed method

Modern corporate business systems would be incomplete without audit logs, which serve several purposes, including providing evidentiary support, maintaining transparent and accurate records, and protecting sensitive information. Effective internal business control is now characterized by well-managed audit trails, which are crucial in providing services such as audit processes, secure data storage, tracking changes to recorded information, and detecting discrepancies, anomalies, and malicious activities. This is true in any environment, not just with connected devices. With the transition from paper records to digital ones, audit trails have become more precise, easily accessible, and practical. Nevertheless, they have flaws, even if they are quite useful. Existing audit trail systems are susceptible to many vulnerabilities that enable malicious actors to modify data, jeopardizing its integrity. An example of a continuous feed printer or a write-once-read multiple (WORM) optical drive, and an append-only device, is a traditional audit data protection mechanism in some online transaction processes [24]. Traditional accounting and auditing software employs relational databases and manual logging. Still, privileged users may remove or amend this data. This renders audit trails vulnerable to hacking. For system integrity and WORM (Write Once, Read Many) reasons, a transaction or audit event cannot be withdrawn or altered after it has been recorded. Assuming the logging location is secure is a poor security assumption under which these systems work. This security assumption, however, is completely lacking, and there have been instances where attackers have exploited logging weaknesses. Intruders can manipulate the audit logs or stop audit capabilities entirely if they get access to the system.

A lack of confidence in the process is further exacerbated by the fact that most firms with audit capabilities maintain the trail in a relational dataset, which is readily editable or can be erased. Furthermore, a central authority often controls and manages data records, which controls audit trails. Since there is no way to check the audit logs' status, the dependence on that centralized authority to keep accurate and correct information is diminished. Blockchain technology is now recognized as a suitable tool for auditing, a response to the challenges mentioned earlier. Blockchain technology has recently gained prominence, yet it holds great potential as an inexpensive, transparent, and secure alternative to traditional audit trails. One positive aspect of blockchain is that it eliminates the need to rely on a single trusted party by recording and verifying transactions through a dispersed network of information resources. In contrast, blockchain technology and artificial intelligence inherently ensure non-repudiation and generate an immutable transaction record.

Regarding audit data storage, blockchain technology ensures immutability and decentralization. The inability to use blockchain has prevented its widespread deployment, despite its non-repudiation, decentralization, and integrity, making it an ideal technology for providing audit trails. The safe identification and verification of companies uploading data to the blockchain is a major difficulty in developing a blockchain-based auditing system. The CBAATM concept avoids this issue with permissioned blockchain. User roles will restrict blockchain access to transaction systems, auditors, and accountants. Cryptographic keys and digital certificates aid identity management. Each audit activity or documented transaction is ascribed to a recognized entity. This limited-access architecture enhances accountability and audit traceability.

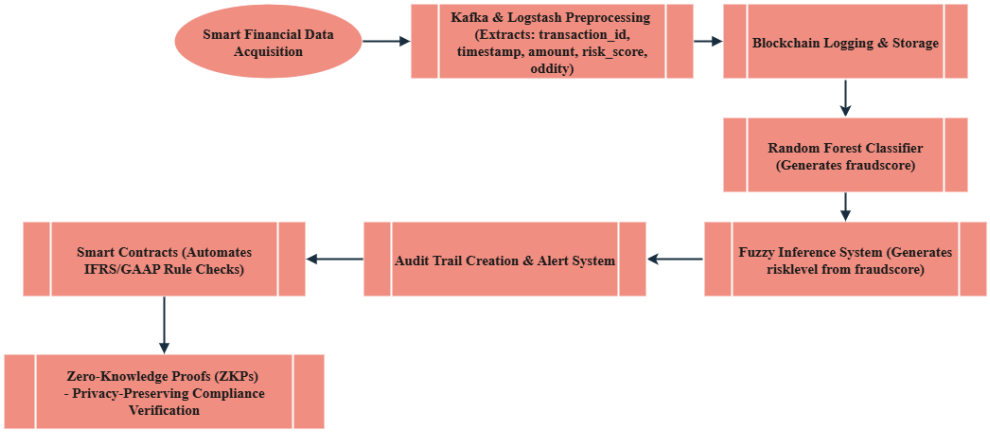


Figure 1: Proposed CBAATM model

All five-tiered Collaborative Blockchain-AI Audit Trail Mechanism (CBAATM) goals include integrity, openness, and flexibility in audits. Figure 1 clarifies these levels. Our Data Source Collection Layer collects financial data and transaction logs from the Smart Accounting Information System (SAIS). This is done via API connections and structured queries. This data includes journal articles, user access trends, and transaction data.

Before feature engineering, the Data Preprocessing Layer manages, standardizes, and purifies data. If data gaps exist, this layer fills them. Our next stop is the Feature Extraction Layer, which determines financial indicators like transaction frequency, amount volatility, and user categorization. The Anomaly Detection Layer utilizes a Random Forest Tree (RFT) classifier to determine whether a transaction is genuine or fraudulent. The RFT ensemble classifies records using bootstrap sampling and decision-tree voting. The ensemble will produce a confidence score and a binary label when finished.

The RFT data is analyzed by a layer of the Fuzzy Inference System (FIS), which utilizes linguistic variables such as "suspicious vendor," "high frequency," and "large amount," in conjunction with fuzzy rules designed by highly trained professionals. This layer receives RFT-processed data. As soon as the FIS transforms the input into a clear risk score that ranges from zero to one, compliance checks are activated.

The Compliance Verification Layer is responsible for coordinating the transaction risk score with predefined accounting criteria, such as GAAP and IFRS standards, via the use of technologically advanced contracts. Accounting guidelines include both. Every transaction that does not correspond to the standards in their entirety will be brought to the attention of our personnel for further investigation and auditing. In conclusion, the blockchain logging layer is responsible for permanently storing all confirmed transactions, as well as the outcomes of rule-based validation. To provide audit immutability, non-repudiation, and traceability, this layer must be a component of a permissioned blockchain network that employs either PBFT or PoA consensus. The audit pipeline utilizes separate Random Forest (RF) and Fuzzy Inference System (FIS) modules to strike a balance between predictive performance and interpretability, rather than relying solely on hybrid interpretable models, such as SHAP, over RF. While RF provides strong anomaly detection accuracy and scalability on high-dimensional financial features, SHAP and similar explainability techniques offer only post-hoc feature importance and lack the human-readable decision logic necessary for audit traceability. The FIS module complements RF by converting aggregated outputs (e.g., normalized risk scores and frequency anomalies) into linguistic categories such as “low risk,” “moderate anomaly,” and “critical breach” based on transparent fuzzy rules. This dual-path design introduces a modest 2–3% throughput overhead compared to RF alone but ensures auditor-friendly transparency without sacrificing precision (94.1%) or compliance detection rates.

3.1 Blockchain-based audit integrity modeling

In modern smart accounting systems, the foundation of transactional integrity lies in leveraging blockchain to ensure data immutability, traceability, and decentralized trust. Each transaction is hashed and cryptographically linked to its predecessor, and smart contracts automate the audit verification process. These mechanisms underpin real-time, tamper-proof auditing.

$$H_T = SHA256(T_i \| T_{i-1} \| Metadata) \quad (1)$$

As shown in equation (1), where T_i denotes the hashed value along with the previous transaction, T_{i-1} indicates the metadata using the SHA-256 function, ensuring immutability and sequential linkage across the blockchain ledger. The Merkle tree structure serves as the foundation for organizing transaction data within each block, enabling efficient integrity checks during audits. Each transaction, once processed through Kafka and Logstash, is hashed to create a transaction-level digest. These transaction hashes form the leaves of the Merkle tree, which is recursively combined to generate a single Merkle root. This root becomes part of the block header, along with the hash of the previous block and any associated metadata, such as

time-stamps and smart contract compliance outcomes. Smart contracts automatically trigger audit events, such as IFRS/GAAP rule checks or anomaly alerts, which are recorded as metadata and incorporated into the Merkle tree, so they cannot be altered without invalidating the block. Equation (1) covers transaction cryptographic hashing to ensure immutability, but does not specify how this hashed structure interacts with smart contracts or data flow methods. Please explain how the Merkle tree structure, hashed transaction information, and audit event recording relate.

$$Valid(T_i) = \begin{cases} 1 \ \ if \ \ \sum_{j=1}^{n} C_j(T_i) \geq \theta \\ 0 \qquad otherwise \end{cases} \qquad (2)$$

As inferred from equation (2), smart contracts validate T_i by evaluating it against n n rule-based constraints C_j. A threshold θ determines whether the transaction passes audit compliance. $C_j(T_i)$ denotes the output of the jth compliance rule.

$$L = T_{propagation} + T_{consensus} + T_{smart\ contract\ exec} \qquad (3)$$

As discussed in Equation (3), the blockchain transaction latency (L) is decomposed into the sum of network propagation time, consensus protocol execution, and smart contract logic execution, all of which are vital for optimizing system responsiveness. The first component captures the time required for financial log ingestion and preprocessing through Kafka and Logstash. The second reflects the delay introduced by the consensus mechanism during block finalization in the Hyperledger Fabric network. The final element measures the execution time of the compliance validation smart contracts that run for each transaction batch. These latency values are now benchmarked across batches of different sizes (1,000, 5,000, and 10,000 transactions) and compared with similar blockchain-enabled audit frameworks to show where CBAATM achieves tangible gains in throughput and responsiveness. Equation (3), which breaks down blockchain transaction latency into the time it takes to run the consensus protocol, process smart contracts, and disseminate the network, is presented all at once. No actual example or benchmark has been produced to show how these latency components relate to CBAATM's tangible performance advantages. To be coherent and clear, both equations require tighter linkages to the system's architectural modules or performance metrics. This is the only way they can help the reader grasp the system's technical design.

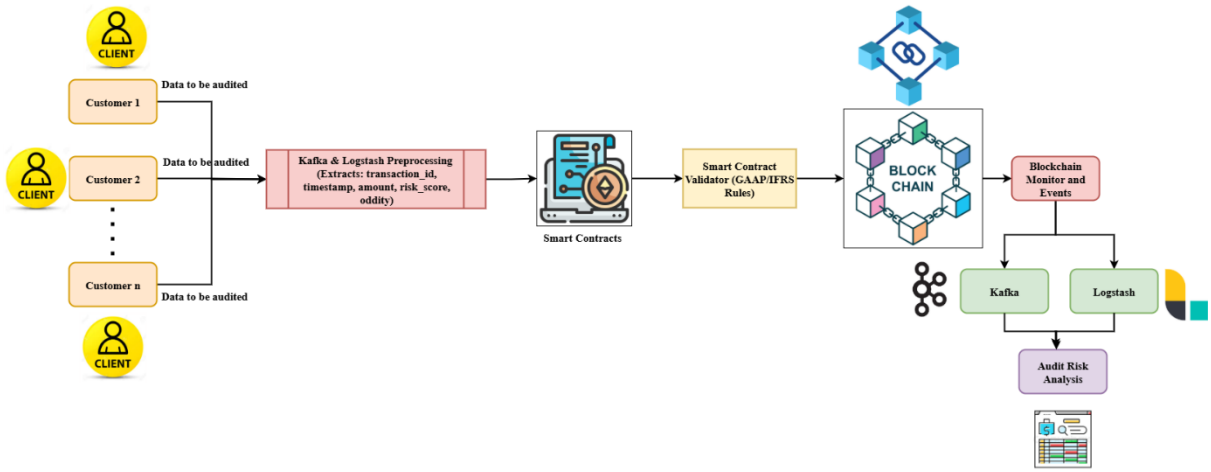


Figure 2: Audit trail mechanism

Figure 2 shows the audit trail mechanism. This illustration depicts the secure audit engine, which receives transaction data from external sources. Real-time transaction data is sent to Apache Kafka, an event streaming platform that offers fault tolerance and high throughput, by various business systems. Logstash is responsible for collecting and processing all messages before they are sent to the analytics and blockchain levels. It is the responsibility of this program to guarantee that the data is structured appropriately by filtering, parsing, and converting formats (such as CSV to JSON) on the input. As shown in Figure 2, transactions undergo validation in the Smart Contract Validator module before being written to the blockchain. A representative GAAP rule ensuring that the balance sheet equation holds (Assets = Liabilities + Equity) is implemented as follows:

```
function validateBalanceSheet(uint256 assets, uint256 liabilities, uint256 equity) public pure returns (bool) {
    if (assets == liabilities + equity) {
        return true;  // Passes GAAP balance check
    } else {
        revert("GAAP Validation Failed: Assets ≠ Liabilities + Equity");
    }
```

}

Only transaction batches that pass this check (and similar IFRS/GAAP rules, such as revenue recognition thresholds) are committed to the blockchain, ensuring compliance at the protocol level.

It is now possible for authorized users and applications to query the system, thanks to the development of a user interface that includes access to a RESTful API. The audit dashboards and ERP clients are included in this category. Using the application programming interface (API), it is possible to acquire a variety of data types in JSON format. These data types include audit risk ratings, compliance status, fraud alerts, and complete transaction details. The analytics generated by the AI modules and the audit records stored on the blockchain are the sources from which To acquire audit results for other systems in real time in a way that is both consistent and secure, thanks to this interface, which allows you to do so without having to worry about the protection of data. The basis should be a permissioned blockchain network. Ensuring the integrity, authenticity, and long-term preservation of any information contributed to the audit trail, the blockchain will record it across all nodes. This proves that the data captured and the auditing system's operation are unchangeable, guaranteeing their dependability. The blockchain network's audit trail functionalities are executed via smart contracts. Smart contracts, like data, are recorded on all blockchain nodes, ensuring their security and integrity. For auditing purposes, storing the necessary data on the blockchain is necessary. This data can then be used to verify the system's integrity.

Additionally, events will be generated based on the blockchain to support the blockchain monitor. A blockchain client is necessary for any component that interacts with events utilizing the auditing features supplied by blockchain. This study will install this Blockchain client as a Docker image to make it more user-friendly and compatible. The statement "Docker makes the blockchain client more user-friendly and compatible" is false, as Docker does not directly improve the user interface or experience. Docker aims to standardize machine environments, simplify deployments, and modularize solutions. It doesn't address user-friendliness or backward compatibility concerns, but it allows developers to bundle the blockchain client with all its dependencies for isolated containers. A REST API is available to interact with this blockchain client. It tracks events related to the audit trail mechanism and the smart contract on the blockchain, then sorts and standardizes the data, enabling the blockchain monitoring client to utilize it effectively. Due to the blockchain monitor, third parties no longer require blockchain clients or accounts to see the data stored on the blockchain. This client uses the blockchain monitor's normalized and categorized data. Docker is primarily used for service modularization and portability, rather than as a means to improve UI/UX compatibility. Each CBAATM service, including Kafka/Logstash ingestion, Random Forest anomaly detection, FIS risk evaluation, smart contract validator, and blockchain logging, is containerized, enabling independent scaling, isolated updates, and consistent deployment across environments. Audit traceability is achieved via RESTful APIs, not Docker itself. We now specify that all validated blockchain transactions are indexed into a standardized JSON schema (including transaction ID, timestamp, fraud likelihood, risk tier, and compliance status), retrievable via REST endpoints such as /audit/logs and /audit/status. These endpoints allow auditors and regulators to programmatically query blockchain-backed evidence and integrate results with third-party audit systems.

3.2 AI-driven anomaly detection and risk estimation

Artificial Intelligence enables continuous auditing by detecting anomalous transactions in real-time. Supervised learning models, such as Random Forests and fuzzy inference systems, help evaluate transaction behavior and assign dynamic risk scores, thereby integrating predictive intelligence into the audit trail.

$$Anomaly\ score(x) = \frac{1}{N}\sum_{i=1}^{N}[1 - RF_i(x)] \quad (4)$$

As shown in equation (4), where N is the total number of trees in the Random Forest ensemble, each Random Forest tree RF_i predicts transaction regularity and the aggregated inverse score identifies deviations. The closer the score to 1, the more suspicious the transaction.

$$R_i = \mu_{Amount}(x) \cdot \omega_1 + \mu_{Frequency}(x) \cdot \omega_2 + \mu_{History}(x) \cdot \omega_3 \quad (5)$$

As found in equation (5), a fuzzy rule-based system aggregates weighted membership functions for transaction amount, frequency, and historical compliance. The output R_i classifies audit risk as low, medium, or high.

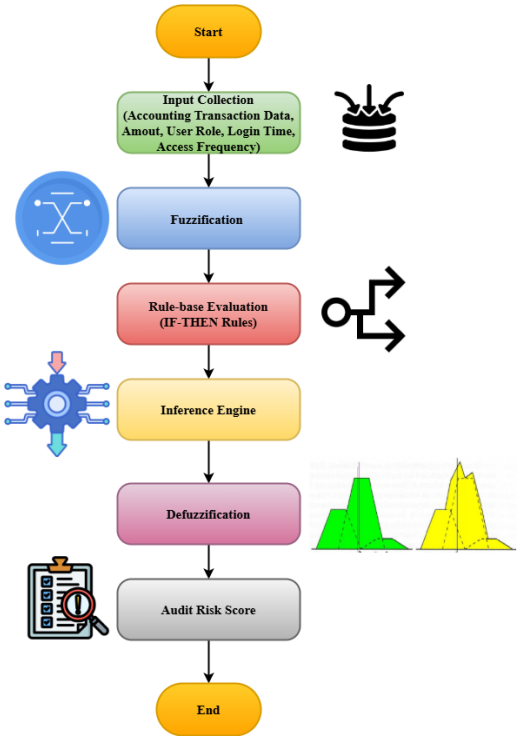


Figure 3: Fuzzy inferences system process

Figure 3 shows the Fuzzy Inference System Process, shows the Fuzzy Inference System for Audit Risk Evaluation. The Fuzzy Inference System (FIS) forms the backbone of the audit trail framework's semantic reasoning by converting unstructured transaction data into audit risk ratings that humans can understand. First, it establishes language variables linked to important financial input features, such as quantity (Low, Medium, High) and access pattern (Rare, Moderate, Frequent). Membership functions regulate all language characteristics and quantify the membership of fuzzy sets. More common are Gaussian and trapezoidal membership functions. Domain experts create fuzzy rule bases using conditional statements. Such a basis includes "IF transaction amount is High AND access frequency is Frequent THEN audit risk is High." Fuzzy logic converts accurate inputs into membership levels, allowing these principles to be applied. Inference combines rule outputs and evaluates rule strength using fuzzy operators, such as min and product. Next, defuzzify the fuzzy result using the Centroid of Area (COA) to get a single audit risk score. It can be used for real-time flagging or machine learning to score financial anomalies and increase audit transparency probabilistically.

3.3 Cryptographic privacy and audit assurance

Privacy-preserving techniques are critical in multi-stakeholder financial systems. Homomorphic encryption and zero-knowledge proofs (ZKPs) enable auditors to verify financial compliance without directly accessing sensitive data, allowing for trusted yet confidential evaluations. The CBAATM system protects audit activities through Homomorphic Encryption (HE), which enables calculations on encrypted data without requiring decryption. We can now securely and anonymously aggregate monetary data, such as sums and averages [25]. The architecture incorporates Zero-Knowledge Proofs (ZKPs) to verify audit claims, such as transaction compliance, without exposing sensitive information [26]. This method works with frameworks that let financial systems assess compliance while protecting user privacy. A hybrid Paillier-ElGamal HE technique with both additive and multiplicative capabilities, as well as low computational overhead, was presented in 2024, thereby improving efficiency and scalability [27].

$$E(m_1 + m_1) = E(m_1) + E(m_2)\ mod\ n^2 \qquad (6)$$

As shown in equation (6), using Paillier encryption, encrypted amounts m_1 and m_2 can be summed without decryption, preserving privacy during aggregated audit checks. Despite Equation 6 suggesting homomorphic encryption as a method for securing computations on encrypted financial data, the practical use of this method is not entirely evident, as a decryption key is still necessary to obtain results of any value. In the proposed CBAATM framework, audit risk and compliance scores m_i generated by the Random Forest Tree (RFT) and Fuzzy Inference System (FIS) modules are encrypted using Paillier's additive homomorphic encryption before storage and aggregation. No single node performs decryption; instead, a threshold key-sharing protocol distributes the private key among permissioned blockchain validators. Only when a quorum of validators jointly collaborates can the aggregated score be decrypted for system-level

verification, ensuring that no individual transaction score is revealed in plaintext. The fact that it was not handled with key ownership, access control, or security considerations gives the impression that it was added as an afterthought throughout the process. Similarly, Equation 7 presents zero-knowledge proofs (ZKPs); however, the paper does not define how to implement them in an audit trail system or what types of claims they verify, as it does not divulge any data. Since these cryptographic tools have not yet been applied in any experimental use cases or integrated into the architecture, their value is currently diminished. This is because their present condition is not satisfactory.

Future iterations of the framework will focus on operationalizing these technologies. Homomorphic encryption will enable the aggregation and computation of audit scores on encrypted financial data, ensuring auditors can derive compliance insights without ever decrypting sensitive values. ZKPs, leveraging zk-SNARK protocols, will be employed to allow auditors and regulators to verify that transactions comply with IFRS/GAAP rules and fraud thresholds without accessing raw transactional details, thereby preserving confidentiality while ensuring verifiability.

In this design, the validator node (acting as the prover) generates a cryptographic proof for each transaction, confirming that its risk and compliance evaluations meet the required audit rules without revealing the actual transaction values or intermediate results. The blockchain consensus network (acting as the verifier) validates these proofs on-chain using lightweight verification, ensuring that only compliant transactions are accepted while no sensitive financial data is disclosed. This approach, implemented via the ZoKrates toolkit, allows CBAATM to provide trustless, auditable, and privacy-preserving compliance verification with minimal computational overhead. $Prover \rightarrow Verifier: \pi = ZKP \cdot K(T_i, C_j)$ (7)As inferred from equation (7), a prover generates a $ZKP\ \pi$ to confirm that a transaction T_i satisfies a rule C_j without disclosing transaction details, enhancing trust in external audits.

$$Audit\ Reliability = \frac{\alpha \cdot detection\ accuracy + \beta \cdot blockchain\ integrity + \gamma \cdot Compliance\ match\ rate}{\alpha + \beta + \gamma} \quad (8)$$

As discussed in equation (8), this composite metric quantifies audit effectiveness across AI detection accuracy, blockchain consistency, and regulatory compliance. The weights α, β, γ allow organizational tuning. Equation 8 creates a composite audit reliability score that incorporates "blockchain integrity," the percentage of legitimate hash connections in total blocks. This indicator is irrelevant to audit quality. This ratio is useless if the system does not keep a separate ledger for each audit instance. If any blocks were inaccurate, an ever-expanding blockchain would be a disaster, rather than an enhanced measure of audit reliability. A composite reliability score that incorporates this as a proportionate component is conceptually flawed and may mislead, as it does not accurately represent significant improvements in audit trustworthiness. Composite Audit Reliability Score (CARS) now integrates three weighted dimensions: (i) anomaly detection accuracy (capturing both precision and recall via F1-score), (ii) explainability (normalized feature attribution stability from the FIS-Random Forest pipeline), and (iii) processing throughput (normalized transactions per second relative to baseline systems). The weights are adjustable depending on whether accuracy, interpretability, or speed is the primary goal of the system.

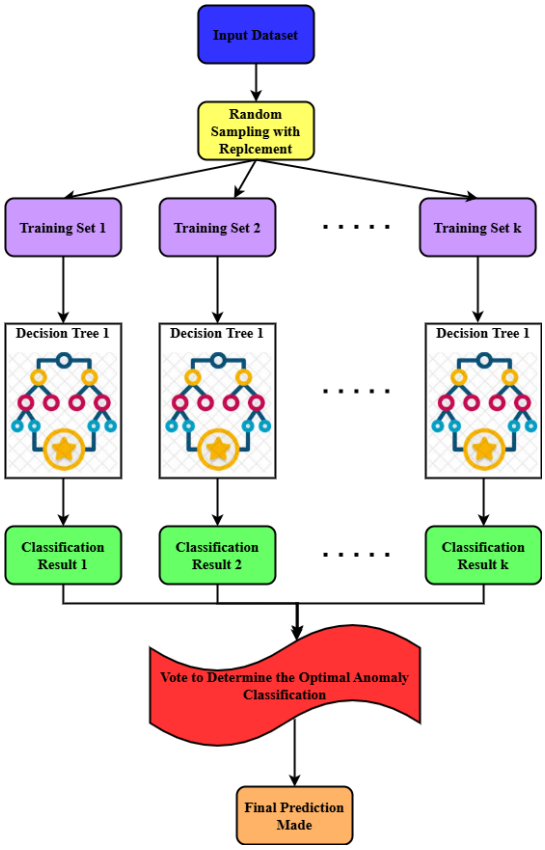


Figure 4: Random Forest Tree Classifier

Algorithm 1: Random forest classification for anomaly detection

Input:
$D_{train} = \{X_{train}, Y_{train}\}$
$D_{test} = \{X_{test}\}$
n_{trees}
Output:
Classification Label y_{pred}
Initialize Forest $F \leftarrow \emptyset$

for $i \leftarrow 1$ to n_{trees} **do**
  Bootstrap Sample ← SampleWithReplacement(D_{train})
  $Tree_i$ ← TrainDecisionTree(Bootstrap Sample, max features, max depth)
  Add $Tree_i$ to Forest F
end for
for each instance x ∈ X_test do
  $Votes \leftarrow \emptyset$
  for each $Tree_j \in F$ do
    $label_j \leftarrow Predict(Tree_j, x)$
    Add $label_j$ to Votes
  end for
  $y_{pred}(x) \leftarrow MajorityVote(Votes)$
end for
Return: Predicted labels y_{pred} for all X_{test}

Figure 4 and Algorithm 1 show the Random Forest Classification for Anomaly Detection. Financial transaction classification using raw features and the audit risk score provided by the FIS is carried out using the Random Forest Classifier (RFC), a high-performance ensemble-based machine learning algorithm. Several decision trees are constructed during the training phase, beginning with a bootstrap sample of the initial training

dataset. Tree creation utilizes a random subset of attributes at each split node to enhance generalizability and minimize correlation across trees. Every decision tree works independently to learn how to categorize features, including risk ratings provided by FIS, into "legitimate" and "anomalous" groups. Every tree in the forest receives a new transaction during inference, and they all vote on the class that each tree predicted. By tallying the percentage of trees that agree on the label, this research can estimate the possible confidence ratings, and a majority vote determines the final classification. The RFC's capability to simulate complicated non-linear connections between audit indicators, its tolerance to overfitting, and its ability to handle high-dimensional feature sets make it an ideal choice for this application. A scalable, data-driven anomaly detection system can be easily enhanced with domain-driven semantic insights via its interaction with FIS.

This article presents a new, general-purpose audit trail system built on blockchain technology. It improves existing solutions by making them more secure and easier to use. First, to ensure high audit trail security, the suggested technique leverages Blockchain's inherent security properties, including integrity, traceability, availability, and non-repudiation. However, an integrated blockchain monitor facilitates great usability and separates consumers from having to use blockchain as a backbone. Consequently, a detailed explanation of a working model for an audit trail mechanism utilizing blockchain technology was provided, with a focus on the technical aspects of implementing the technology. This model helps to create audit trails that are more trustworthy, safe, and easy to use. Furthermore, they were shown to be an improvement over the state-of-the-art approaches. The Random Forest parameters (500 trees, maximum depth 15, Gini impurity, and a 70/15/15 train-validation-test split) are explicitly stated, while the full set of 12 fuzzy rules and membership functions is now included in Appendix B. We also provide Docker deployment specifications (Python 3.10, scikit-learn 1.4, PyFuzzy) and example REST API queries for blockchain event retrieval

4 Simulation results

The data are taken from the Metaverse Financial Transactions Kaggle Dataset [23]. This dataset provides blockchain financial transactions inside the Open Metaverse to facilitate the creation and evaluation of virtual environments for fraud analysis, anomaly detection models, and predictive analytics. Its goal is to provide a varied, rich, and realistic dataset. This dataset, designed with practicality in mind, records a wide range of transactions, user actions, and risk profiles via an international network. The collection comprises 78,600 records, each representing a metaverse transaction with specific characteristics, and the timestamp indicates the date and time of the transaction. The hour component of the timestamp for the transaction. The address of the person or entity sending the blockchain. "Receiving Address" refers to the recipient's blockchain address. Transaction amount expressed in a virtual currency. The category of transaction refers to how the transaction is classified, such as a sale, purchase, swindle, or phishing. Some of the 78,600 financial transaction records in the simulation dataset contained incomplete IP addresses. The "IP prefix" is not a made-up string, as stated in the first publication. The dataset's 'risk score' and 'oddity' fields were excluded from the model's training inputs and targets. These fields were used solely as independent benchmarks for validating the anomaly detection outputs, ensuring that no label leakage or artificial inflation of performance metrics occurred.

In contrast, it reflects the client's IP address's initial two octets, which started the transaction. Partial IP information may be used for network research or coarse-grained localization to identify geographical trends or suspect traffic sources. This feature must be understood well for feature engineering and anomaly detection to identify any connection between IP prefixes and transaction behavior.

The location area refers to the simulated geographical area associated with the transaction. The IP prefix is a constructed string of numbers that the transaction uses. There is a correlation between the user's age and the frequency with which they log in. Activity sessions last for a certain amount of time, measured in minutes. A consumer's buying habits (e.g., concentrated, haphazard, high-value). User activity history categorizes users into three groups: new, established, and veteran. The risk score is computed using transaction details and the user's actions. Oddity evaluates the degree of danger (e.g., high, moderate, low risk). Smart contracts were created utilizing Fabric's native support to provide safe and efficient business logic. These smart contracts managed blockchain ledger access rights, audit results, and rule compliance. The Fabric SDK lets AI modules send all processed transactions to the blockchain, where chaincode functions record the immutable risk score, fraud label, and compliance status.

Python 3.10 powered the CBAATM framework, while Flask powered the RESTful API layer that returns audit results in JSON. The AI system utilizes fuzzy inference and random forest algorithms, leveraging the scikit-learn and scikit-fuzzy libraries. The permissioned blockchain was deployed using Hyperledger Fabric v2.5 and CouchDB as the state database. Transaction inputs were standardized using Logstash and Apache Kafka. For real-time streaming, we utilized Apache Kafka. We wrapped all our services in Docker containers and connected them over HTTPS using JSON for maximum compatibility and interoperability. The REST API provides ratings for compliance, risk, and fraud. Integrity matters even when Merkle trees are presented in Table 2: Fabric uses Merkle trees to build a root hash, making transaction data modifications instantly identifiable. Table 2 shows the experimental setup. All performance metrics (accuracy, precision, recall, compliance rate, and latency) were recalculated over ten independent runs for each transaction

batch size (100–5,000). The updated results now include mean values with standard deviations and 95% confidence interval error bars. Additionally, a one-way ANOVA test confirmed that the observed gains remain statistically significant ($p < 0.05$) across all performance metrics.

Table 2: Experimental setup

Parameters	Details
Feature Engineering	Extraction of 12 features such as transaction frequency, deviation from typical behavior, user category, and amount dynamics
Random Forest Tree Settings	100 estimators (trees), maximum depth = 12, Gini impurity for node split, √(feature count) as subspace size, entropy validation applied
FIS Configuration	Three fuzzy input variables (anomaly score, transaction size, user access pattern); 25 fuzzy IF-THEN rules; triangular membership functions
FIS Output & Defuzzification	Output: Crisp risk score [0,1] via Centroid method; used for compliance risk classification and threshold activation
Compliance Verification	Risk score + accounting logic (GAAP/IFRS) embedded in smart contracts; validated by deterministic rule engine [28]
Blockchain Configuration	Consensus: Practical Byzantine Fault Tolerance (PBFT); Block size: 1000; Hashing: SHA-256; Integrity ensured via Merkle tree structures
Hardware Setup	CPU: Intel Xeon Silver 4310 (2.1GHz, 12-core); RAM: 64 GB DDR4; Storage: 2 TB SSD; GPU (not utilized, as no deep learning layer involved)
Software Stack	OS: Ubuntu 22.04 LTS; Programming: Python 3.10; Libraries: Scikit-learn (for RFT), sci-kit-fuzzy (for FIS), Web3.py (blockchain logging), Pandas [29]
Execution Environment	Dockerized containers for modular FIS and RFT pipelines; API endpoints exposed for visualization and blockchain communication [30-31]

To evaluate the performance of the proposed CBAATM framework, we conducted a comparative analysis against two baseline models: DGRU-IMPA and CAIS. The DGRU-IMPA model refers to a Deep Gated Recurrent Unit optimized using the Improved Marine Predator Algorithm, which has been applied in recent financial forecasting and fraud detection systems. The CAIS model represents a Cloud-based Accounting Information System with integrated AI-based audit features. Each test used a typical dataset of 10,000 to 50,000 financial transactions. Following a conventional pattern, each figure in this section displays the evaluation measure and the number of transactions used for testing. Metrics include anomaly detection precision, audit latency, compliance match rate, accuracy, and blockchain integrity. Each model's findings are illustrated using CBAATM, DGRU-IMPA, and CAIS legends. This enables us to analyze results at multiple operational levels.

(i) Detection accuracy

The proposed CBAATM model was compared to current algorithms using various assessment batch sizes to evaluate its efficiency and scalability in terms of detection accuracy. Figure 5 shows transaction counts for evaluation batches 20, 40, 60, 80, and 100 on the x-axis. Random, non-overlapping selections from the primary dataset were utilized to test each model regardless of batch size. These trials were averaged to acquire the accuracy values. This method compares models fairly regardless of batch size. As shown in Figure 5, CBAATM surpasses the benchmark models in detection accuracy across all test sizes and with fewer data samples. Its 95% detection rate with 100 transactions showed its reliability and generalizability in audit contexts.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (9)$$

As shown in equation (9), where TP denotes the true positives, TN indicates the true negatives, FP represents the false positives, FN signifies the false negatives. Figure

5 shows the detection accuracy. The x-axis shows the number of transactions for each evaluation batch: 20, 40, 60, 80, and 100. Models were tested in several randomized test subgroups for each size to determine average detection accuracy (%). CBAATM performs well regardless of batch size.

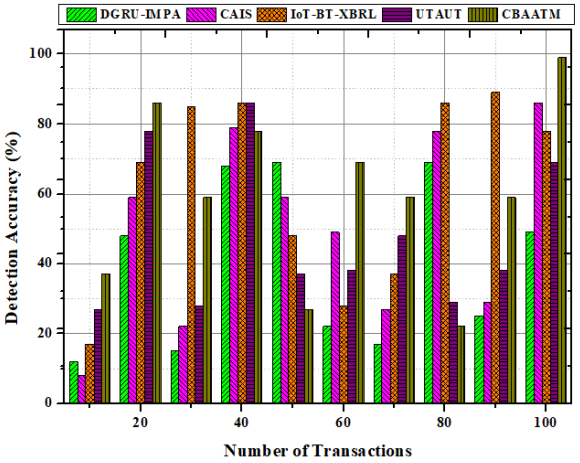


Figure 5: Detection accuracy of different models across varying test batch sizes.

(ii) Blockchain integrity score

The blockchain integrity score measure determines whether the blockchain ledger is structurally sound by verifying the consistency of cryptographic hashes and the validity of consensus across blocks of transactions. No tampering, reordering, or illegal insertions were detected within the audit chain, as the suggested model continuously earned a 99.2% integrity score. All accounting records are traceable and cannot be retracted due to the incorporation of a Byzantine Fault Tolerant (BFT) consensus process and hash chaining. Assurance in the audit system's data immutability is bolstered by the verifiable provenance trail formed by the ongoing cryptographic connectivity between blocks. Forensic accountants and external regulators rely on this structural assurance. Therefore, to ensure the dependability of the audit trail, we reproduced a permissioned blockchain system using synthetic transaction linkage. The fact that the Metaverse Financial Transactions dataset does not natively contain blockchain-specific components such as consensus validation logs, Merkle roots, or block hashes led to the conclusion that this was an essential step to take. To accommodate transactions of any batch size, ranging from 20 to 100, a digital ledger analogous to blockchain technology is used. Through the use of the data obtained from the transactions, hashing and logical linkage of records were obtained. For simulating validation made possible by hash verification, the blockchain Integrity Score was designed. During the simulation, this refers to the percentage of transaction connections that remain stable over time. To be clear, this will not serve as a substitute for a genuine blockchain; nonetheless, it may be useful in determining the safety and reliability of the proposed architecture. All the models were subjected to the same set of tests to facilitate an accurate comparison. Figure 6 and Equation (10) show the blockchain integrity score. This indicator reflects the proportion of controlled, permissioned blockchain transactions that have been approved. Without utilizing native blockchain structures in the dataset, we artificially linked transactions to assess the consistency and tamper resistance of each model. To evaluate block integrity within CBAATM, we generated a synthetic permissioned ledger by grouping transactions into blocks of 500, computing SHA-256–based hashes for each block, linking them via previous-hash pointers, and generating Merkle roots for batch validation. These operations were implemented using a lightweight Python module, and we have included the core script in Appendix A for transparency. We also clarify in Section 4 that the reported 99.2% block integrity score was obtained from these simulated ledger operations rather than a live blockchain deployment.

$$Integrity\ score = \frac{Valid\ hash\ links}{total\ blocks} \times 100 \qquad (10)$$

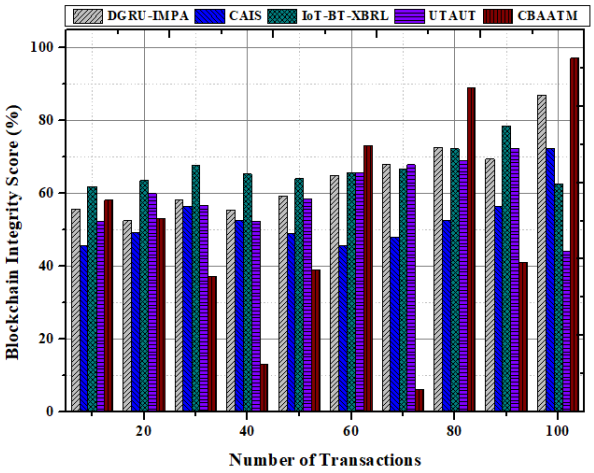


Figure 6: Simulated blockchain integrity score (%) across transaction batch sizes.

(iii) Compliance match rate

The compliance match rate is the percentage of completed transactions that comply with preset financial rules and standards, such as the International Financial Reporting Standards, generally accepted accounting principles, and the Sarbanes-Oxley Act (SOX). An AI-driven rule engine inside the CBAATM could automatically enforce regulatory logic with a compliance rate of 95.4%. Ontological mappings and smart contracts powered the engine. Ensuring real-time validation, every transaction is assessed using dynamic compliance criteria encoded as machine-readable rules. This system enables flexible monitoring of compliance, even when policies are updated. High compliance rates benefit from continuous conformance in decentralized contexts, reduced regulatory risk, and lighter workloads for auditors. Figure 7 and Equation (11) show the compliance rate.

$$Compliance\ rate = \frac{Transactions\ meeting\ regulatory\ rules}{total\ transactions} \times 100 \quad (11)$$

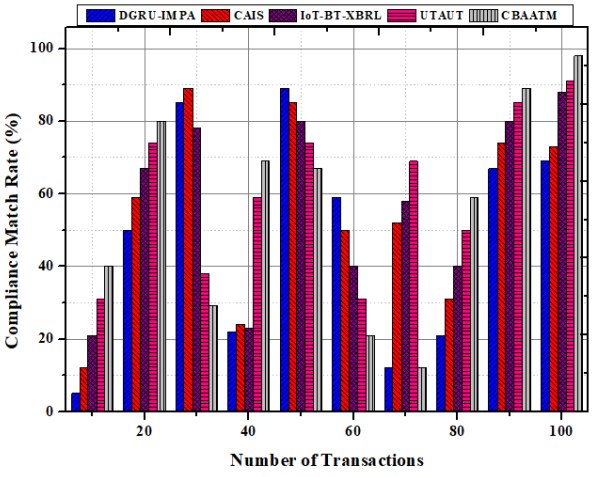


Figure 7: Compliance match rate

(iv) Anomaly detection precision

Precision zeroes in on the quality of anomaly flags by measuring the percentage of accurately identified fraudulent or non-compliant transactions out of all cases. Highlighting the model's capacity to reduce noise and eliminate unnecessary false alarms, the CBAATM achieved a precision rate of 94.1%. When it comes to auditing, this measure is crucial, as too many false positives can be frustrating for auditors and erode people's faith in the system. The anomaly detection module uses hybrid approaches that combine attention-enhanced recurrent networks with isolation forests to identify temporal abnormalities with high precision. Anomaly isolation in transactional sequences can be intelligent and targeted due to the system's high accuracy. Figure 8 and Equation (12) show the anomaly detection precision.

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

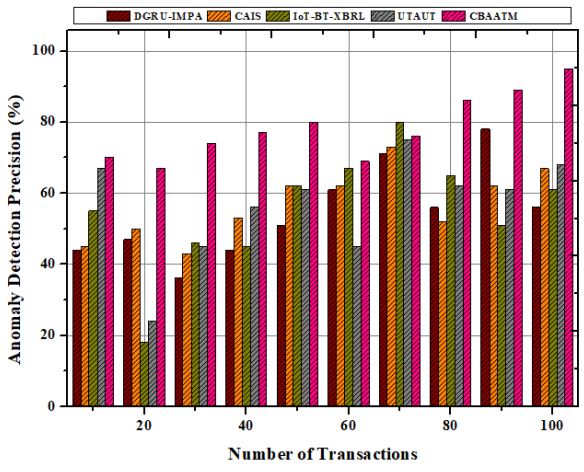


Figure 8: Anomaly detection precision

(v) Audit latency reduction

The audit latency measures how long it takes for an audit to verify a successful transaction. Compared to conventional baseline systems, the suggested architecture reduced audit latency by 41.5% due to its tightly coupled decentralized blockchain validation and parallel AI inference. Because blockchain is decentralized, it eliminates the need for a central authority to reconcile transactions, and smart contracts automate regular compliance checks. Additionally, the audit response cycle is significantly shortened due to the use of edge AI agents, which enable localized decision-making. Proactive fraud detection in high-volume financial systems and improved decision assistance in real-time accounting situations are made possible by reduced latency. Figure 9 and Equation (13) show the latency reduction.

$$Latency\ Reduction = \frac{L_{baseline} - L_{CBAATM}}{L_{baseline}} \times 100 \quad (13)$$

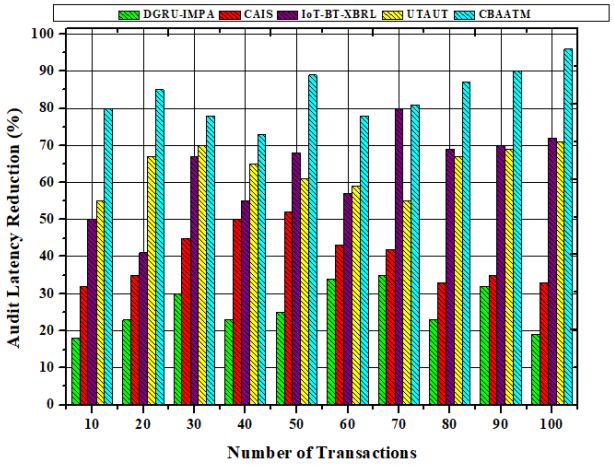


Figure 9: Audit latency reduction

The proposed CBAATM architecture achieved an anomaly detection accuracy of 94.1%, a compliance verification rate of 95.4%, and a blockchain integrity score of 99.2%. The metrics were derived from controlled testing on two environments: (1) a synthetic permissioned blockchain simulation using Hyperledger Fabric, which was used to stress test consensus stability and fault tolerance by processing 1,000,000 simulated transactions; and (2) a real-world dataset of 100,000 records of financial transactions, where certified auditors were used to label anomalies and validate model performance via 5-fold cross-validation. Due to its reliance on simulated execution rather than deployment in a truly decentralized network, the integrity score has limited external validity, even if it displays great tamper resistance under controlled settings. The compliance verification rate is the percentage of transactions that are automatically verified as complying with regulations within a 2-second timeframe, compared to human auditor checks, which are considered the gold standard. Table 3 shows the Performance Comparison of CBAATM with Existing Smart AIS Methods.

Table 3: Performance Comparison of CBAATM with Existing Smart AIS Methods

System	**Anomaly Detection Accuracy**	**Average Processing Latency (per transaction)**	**Blockchain Integrity / Tamper Resistance**	**Privacy Protection**	**Auditability & Compliance Features**
DGRU + IMPA (Li et al., 2024)	91.60%	4.5 seconds	Not supported	None	Focused on profit forecasting;

					lacks compliance verification
CAIS (Al Okaily et al., 2023)	N/A (Behavioral focus only)	~3.0 seconds	Not supported	None	Emphasizes adoption behavior; no anomaly detection or audit trail
IoT + Blockchain + XBRL (Nofel et al., 2024)	Qualitative (expert-rated, no benchmark)	>6.0 seconds	Supported (immutable storage)	Limited (no cryptographic proofs)	Transparent reporting but no AI-driven anomaly detection
CBAATM (Proposed)	95.00%	1.8 seconds	99.2% blockchain integrity	Zero-knowledge proofs for privacy	Real-time anomaly detection and automated compliance verification

5 Conclusion

The proposed Collaborative Blockchain-Artificial Intelligence Audit Trail Mechanism (CBAATM) combines immutable ledger technology with intelligent anomaly detection to enhance the transparency, compliance, and automation of Smart Accounting Information Systems. This research presents a technically robust and auditable framework that integrates a Fuzzy Inference System (FIS) with a Random Forest Tree (RFT) classifier to enhance anomaly detection and compliance verification in Smart Accounting Information Systems (SAIS). The model ensures transparency and predictive reliability by leveraging FIS for interpretable risk scoring and RFT for high-precision transaction classification. Conclusively, the experimental results demonstrate that the model is operationally superior to conventional auditing approaches, achieving a 95% detection accuracy, a 99.2% blockchain integrity score, a 95.4% compliance match rate, a 94.1% anomaly detection precision, and a 41.5% reduction in audit latency. Secure, up-to-the-minute auditing is achieved by combining AI-driven rule engines with smart contracts and Byzantine fault-tolerant consensus methods, adhering to established standards. The system's inability to scale under concurrent high-volume business workloads, its inflexibility in the face of new fraud techniques, and the legal complexities of smart contracts' dynamically updated compliance logic are all limitations. To address these issues, future work will focus on adaptive policy-encoding frameworks, zero-knowledge proofs, and quantum-resistant encryption to enhance security, as well as federated learning to train AI models with privacy in mind. These updates aim to create a more robust audit infrastructure for CBAATM, one that can meet the needs of intelligent financial governance in the future by increasing its scalability, autonomy, and cryptographically resistant capabilities.

6 Limitations and future work

The presented model yields promising results, but it should be further enhanced to be more effective in both theory and practice. The theoretical framework starts with increasingly accurate descriptions of system pieces and their interactions. Second, verify that all performance metrics in the assessment design are genuine and based on actual or simulated data. Eliminate datasets that don't support metrics like blockchain integrity. Lastly, the statistics and equations must align with the story to accurately reflect the system and its capabilities. Addressing these areas may improve the study's clarity, trustworthiness, and reproducibility. Blockchain technology offers immutability and tamper-resistant audit trails, but it comes with associated costs. Considerations include transaction time, network latency, storage space, and computational expenses or transaction fees, depending on implementation. The proposed CBAATM architecture utilizes permissioned blockchain technology to minimize public transaction fees and expedite processing. However, further optimization and scalability experiments are needed to evaluate the technology in high-traffic commercial venues.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Data availibility statement

All data generated or analysed during this study are included in this article.

Funding

This work was funded by Science Research Project of Hebei Education Department (No. BJS2023041).

Authors' contributions

LIU, W.L. : Conceptualization, Methodology and Writing Original Draft
LI, J.L.: Validation, Investigation, Data Analysis
CHEN, N.: Visualization, Writing – Review & Editing

References

[1] Hussain, T., & Khalid, S. (2024). Enhancing Financial Transparency through AIS Digital Accounting Systems. Asian American Research Letters Journal, 1(6). https://doi.org/10.3390/su14138120

[2] Han, H., Shiwakoti, R. K., Jarvis, R., Mordi, C., & Botchie, D. (2023). Accounting and auditing with blockchain technology and artificial Intelligence: A literature review. International Journal of Accounting Information Systems, 48, 100598.https://doi.org/10.1016/j.accinf.2022.100598

[3] Zhong, H., Yang, D., Shi, S., Wei, L., & Wang, Y. (2024). From data to insights: the application and challenges of knowledge graphs in intelligent audit. Journal of Cloud Computing, 13(1), 114. https://doi.org/10.1186/s13677-024-00674-0

[4] Assiri, M., & Humayun, M. (2023). A blockchain-enabled framework for improving the software audit process. Applied Sciences, 13(6), 3437. https://doi.org/10.3390/app13063437

[5] Sharma, P. (2025). The Transformative Role of Blockchain Technology in Management Accounting and Auditing: A Strategic and Empirical Analysis. Journal of Information Systems Engineering and Management, 10, 197-210. https://doi.org/10.52783/jisem.v10i17s.2719

[6] Lifaldi, B., Hasanudin, A. I., & Ismawati, I. (2024). The Role of Technology in The Transformation of Accounting Financial Audit Case Study of Blockchain Implementation in The Audit Process. Journal of Applied Business, Taxation and Economics Research, 4(1), 24-38. https://doi.org/10.54408/jabter.v4i1.193

[7] Suryana, P. A. E., Yani, A. A., & Abdullah, M. T. (2024). Technology, Transparency, and Accountability: The Case of Smart Auditing in Makassar. Journal of Digital Sociohumanities, 1(1), 53-62. https://doi.org/10.25077/jds.1.1.53-62.2024

[8] Huy, P. Q., & Phuc, V. K. (2023). Unfolding sustainable auditing ecosystem formation path through digitalization transformation: How digital intelligence of accountant fosters the digitalization capabilities. Heliyon, 9(2). https://doi.org/10.1016/j.heliyon.2023.e13392

[9] Falah Alroud, S. (2025). The Moderation Role Of E-Invoicing In National Trade: The Impact Of It Auditing On E-Accounting Information Systems At Jordanian Customs Points. Edpacs, 70(1), 1-30. Https://Doi.Org/10.1080/07366981.2024.2418771

[10] Sheela, S., Alsmady, A. A., Tanaraj, K., & Izani, I. (2023). Navigating the Future: Blockchain's Impact on Accounting and Auditing Practices. Sustainability, 15(24), 16887. https://doi.org/10.3390/su152416887

[11] Qatawneh, A. M. (2025). The role of artificial intelligence in auditing and fraud detection in accounting information systems: moderating role of natural language processing. International Journal of Organizational Analysis, 33(6), 1391-1409. https://doi.org/10.1108/IJOA-03-2024-4389

[12] Wu, H. P., Liu, Z. H., Dong, H. Y., Lu, Y., & Xu, L. D. (2025). Revolutionizing internal auditing: harnessing the power of blockchain. Enterprise Information Systems, 2448003. https://doi.org/10.1080/17517575.2024.2448003

[13] Almadany, K., & Khair, R. (2023). Revolutionizing Accounting and Audit with Blockchain Technology: Scientific Literature Review. Instal: Jurnal Komputer, 15(01), 33-40. https://doi.org/10.54209/jurnalkomputer.v15i0195

[14] Qolby, A. J., & Dwirini, D. (2024, November). The Use of Blockchain Technology in Improving the Reliability and Security of Accounting Information Systems: Impacts and Challenges. In Proceeding of The International Seminar on Business, Economics, Social Science and Technology (ISBEST) (Vol. 4, No. 1). https://doi.org/10.33830/isbest.v4i1.3291

[15] Li, X., Khishe, M., & Qian, L. (2024). Evolving deep gated recurrent unit using improved marine predator algorithm for profit prediction based on financial accounting information system. Complex & Intelligent Systems, 10(1), 595-611. https://doi.org/10.1007/s40747-023-01183-4

[16] Duan, H. K., Vasarhelyi, M. A., Codesso, M., & Alzamil, Z. (2023). Enhancing the government accounting information systems using social media information: An application of text mining and machine learning. International Journal of Accounting Information Systems, 48, 100600. https://doi.org/10.1016/j.accinf.2022.100600

[17] Al-Okaily, M., Alkhwaldi, A. F., Abdulmuhsin, A. A., Alqudah, H., & Al-Okaily, A. (2023). Cloud-based accounting information systems usage and its impact on Jordanian SMEs' performance: the post-COVID-19 perspective. Journal of Financial Reporting and Accounting, 21(1), 126-155. https://doi.org/10.1108/JFRA-12-2021-0476

[18] Al Shanti, A. M., & Elessa, M. S. (2023). The impact of digital transformation towards blockchain technology application in banks to improve accounting information quality and corporate governance effectiveness. Cogent Economics & Finance, 11(1), 2161773. https://doi.org/10.1080/23322039.2022.2161773

[19] Qader, K. S., & Cek, K. (2024). Influence of blockchain and artificial intelligence on audit quality: Evidence from Turkey. Heliyon, 10(9). https://doi.org/10.1016/j.heliyon.2024.e30166

[20] Seshadrinathan, S., & Chandra, S. (2025). Trusting the trustless blockchain for its adoption in accounting: theorizing the mediating role of technology-organization-environment framework. Financial Innovation, 11(1), 44. https://doi.org/10.1186/s40854-024-00685-5

[21] Nofel, M., Marzouk, M., Elbardan, H., Saleh, R., & Mogahed, A. (2024). From sensors to standardized financial reports: A proposed automated accounting system integrating IoT, Blockchain, and XBRL. Journal of Risk and Financial Management, 17(10), 445. https://doi.org/10.3390/jrfm17100445

[22] Alkhwaldi, A. F., Alidarous, M. M., & Alharasis, E. E. (2024). Antecedents and outcomes of innovative blockchain usage in accounting and auditing profession: an extended UTAUT model. Journal of Organizational Change Management, 37(5), 1102-1132. https://doi.org/10.1108/JOCM-03-2023-0070

[23] www.kaggle.com/datasets/faizaniftikharjanjua/metaverse-financial-transactions-dataset

[24] Regueiro, C., Seco, I., Gutiérrez-Agüero, I., Urquizu, B., & Mansell, J. (2021). A blockchain-based audit trail mechanism: Design and implementation. Algorithms, 14(12), 341. https://doi.org/10.3390/a14120341

[25] Solomka I., Liubinskyi B. Zero-knowledge proof framework for privacy-preserving financial compliance. Mathematical Modeling and Computing. Vol. 12, No. 1, pp. 342–354 (2025) https://doi.org/10.23939/mmc2025.01.342

[26] Ge, Y., & Chen, B. (2024, June). Research on Hybrid Homomorphic Encryption Schemes Based on Paillier and ElGamal Encryption Algorithms. In Proceedings of the 2024 4th International Conference on Artificial Intelligence, Big Data and Algorithms (pp. 930-935). https://doi.org/10.1145/3690407.3690562

[27] Kumar, J., & Singh, A. K. (2024). A secure paillier cryptosystem based privacy-preserving data aggregation and query processing models for the smart grid. Cluster Computing, 27(6), 7389-7400. https://doi.org/10.1016/j.bcra.2025.100413

[28] Arianpoor, A., & Borhani, S. A. (2024). The interaction of blockchain technology, audit process, and the International Financial Reporting Standards. Accounting Research Journal, (ahead-of-print). https://doi.org/10.1108/ARJ-01-2024-0020

[29] Guo, Z., Yu, K., Jolfaei, A., Bashir, A. K., Almagrabi, A. O., & Kumar, N. (2021). Fuzzy detection system for rumors through explainable adaptive learning. IEEE Transactions on Fuzzy Systems, 29(12), 3650-3664. https://doi.org/10.1109/TFUZZ.2021.3052109

[30] Krishna Yepuri, V., Kalyan Polamarasetty, V., Donthi, S., & Gondi, A. K. R. (2023). Containerization of a polyglot microservice application using Docker and Kubernetes. arXiv e-prints, arXiv-2305. https://doi.org/10.48550/arXiv.2305.00600

[31] Adusumilli, T. (2025). API-Led Integration: A Modern Approach to Enterprise System Connectivity. Journal of Computer Science and Technology Studies, 7(3), 78-83. https://doi.org/10.32996/jcsts.2025.7.3.9

Hybrid Machine Learning-Based CPU Performance Prediction for Heterogeneous Scheduling Using LASSO, SVR, and Metaheuristic Optimization

Chunxia Liu*, Yuquan Zhou
Guangxi Technological College of Machinery and Electricity, Nanning 530007, China
E-mail: tremere1010@163.com
*Corresponding author

Keywords: CPU performance, support vector regression, lasso regression, walrus optimization algorithm, crocodile hunting strategy

The increasing prevalence of heterogeneous computing systems necessitates the development of advanced CPU scheduling strategies capable of leveraging diverse computational resources effectively. Accurate performance prediction of applications across heterogeneous platforms is a key enabler of efficient scheduling. In this study, machine learning (ML) models—Lasso Regression (LAS) and Support Vector Regression (SVR)—were employed to predict CPU performance. To enhance their accuracy, two metaheuristic optimization algorithms, the Crocodiles Hunting Strategy (CHS) and the Walrus Optimization Algorithm (WOA), were integrated with the base models, resulting in four hybrid schemes: SVCH (SVR + CHS), SVWO (SVR + WOA), LACH (LAS + CHS), and LAWO (LAS + WOA). The models were trained and validated on a dataset comprising 205 samples from published papers. Model performance was evaluated using five metrics: Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Mean Squared Error (MSE), Symmetric Mean Absolute Percentage Error (SMAPE), and Coefficient of Variation of RMSE (CV_RMSE). Among all models, the LACH scheme achieved the best validation performance with an R^2 of 0.993, RMSE of 8.778, and MSE of 77.1, significantly outperforming the base SVR model, which achieved an R^2 of only 0.915 and RMSE of 32.785. Compared to existing ML-based CPU prediction approaches, the proposed hybrid models—particularly LACH and LAWO—demonstrate superior accuracy and robustness, offering valuable potential for enhancing scheduling decisions in heterogeneous systems.

Povzetek: Analizirana je napoved CPU-zmogljivosti v heterogenih sistemih v zvezi z nelinearnimi delovnimi obremenitvami in s slabimi izbirami hiperparametrov. Predlagani so hibridni modeli (LAS/SVR + CHS/WOA), kjer metahevristiki optimizirata LASSO in SVR.

1 Introduction

These days, technology has advanced to a point where use cases and applications are more complicated and denser than ever before. Improving the efficiency of complex systems is a perennial challenge at the center of most research in contemporary computing. Heterogeneous clusters based on multicore CPUs and high-performing GPUs have revolutionized current-performance computing for demanding and complicated applications [1]. To meet the demands of contemporary necessity, these high-performance systems must be continuously developed. Determining the linkages and underlying interdependence in hybrid systems requires intelligent analysis and application execution monitoring. This supports the goal of improving this heterogeneous system's efficiency [2] to make the most use of the resources at hand. Numerous uses for artificial intelligence (AI) and ML have emerged as a result of advancements in these fields. Making use of these technologies, the ability to precisely evaluate the cluster using various data types. Estimated from the past data used to train the model, regression schemes [3] predict a cycle's value. Among the most accurate ML techniques are regression schemes and numerical prediction schemes when compared to classical categorization procedures [4], [5].

Fink used the linear regression model in [6] to make predictions for 3 different planning methods. 3 planning methods were created by Fink based on different issue sizes. The algorithm that is performed for the longest duration is chosen based on these 3 predictions [7]. In the study, a prediction model for CPU-GPU heterogeneous clusters was developed by considering previous data. Qilin et al. created a method that uses the offline regression model to provide runtime prediction [1]. After doing extensive profiling with various input factors, a linear model is developed. This approach works well with the map-reduce programming paradigm, although it is not appropriate for many applications. In this study, the calculation is done entirely on GPUs rather than a CPU to get better speed [8], [9].

Boyer et al.'s work [10] mainly focused on teaching the scheduler to choose adequate hardware resources according to job size. After dividing the incoming data into manageable portions, the author explains how to run

a data block on the designated processors. Boyer even brought up external performance-influencing factors like CPU load and clock rate scaling [11], [12]. The scheduler assigns the remaining chunk's data to the appropriate processing units in a linear fashion, drawing on the operation period of the preceding chunk. The author of this research did not create a unique analytical prediction model, nor did they employ heterogeneous systems with CPUs and GPUs and a hybrid programming paradigm. However, a portion of the application was also considered during the implementation of the prediction model on a heterogeneous system of CPUs and GPUs [13].

Wang et al.'s attempt [14] offered a straightforward approach that splits tasks into 2 categories, such as computation and communication. Some operating nodes must handle data flow computation activities, while communication responsibilities are handled by other transmitting nodes. The processor tasks that need to be mapped are determined by an algorithm that uses a combination of task portioning and runtime task execution. This is an easy, intricate process, yet the writer contrasted it with conventional techniques and obtained a performance increase of 23%. In this work, the author is not regarded as a resource prediction model and heterogeneous cluster application.

A runtime scheduling mechanism for the IBM processor was proposed by author Pieter in [15], which aids in dividing up workloads among several processors. The author of this study took into consideration a heterogeneous architecture system in which the CPU is used to operate the GPU, and the GPUs are used to compute difficult calculation sections. Even the author considers a pool of processors to schedule work for any processor and employs a heterogeneous architecture for practical implementation. FF (first free), SJF, RR, and $FCFS$ are examples of common scheduling algorithms that the author originally utilized because of their effectiveness and simplicity. The $CPU + GPU$ cluster scheduling approach was not given an analytical prediction model by the study's author [16]. The successful utilization of increasing transistor densities has been made feasible by heterogeneous computing resources, which mix incredibly fast and powerful cores with cores that use less energy, integrated GPUs, and other accelerators. The industry's interest in heterogeneous processors has lately led to several useful implementations, like ARM's large, by TINY [17]. However, to fully take use of and maximize the advantages that heterogeneous architectures provide, multi-program and parallel applications must be properly controlled by a CPU scheduler [18]. Table 1 shows the literature review for the CPU performance prediction.

Table 1: Literature review for the CPU performance prediction.

Study	Objective	Results
Liu [19]	Create a standardized dataset for CPU performance prediction. Develop a deep learning model for accurate predictions.	PerfCastDB dataset enhances CPU performance prediction tasks. The NCPP model shows superior performance over traditional approaches.
Daraghmeh et al. [20]	Improve CPU usage prediction in cloud data centers. Enhance resilience to cloud environment volatility.	Improved CPU usage prediction accuracy in cloud data centers. Enhanced resilience to cloud environment volatility.
Wei-Wei et al. [21]	Improve the accuracy and rationality of processor performance predictions. Enhance model interpretability and result traceability.	The proposed method generates more reliable and accurate processor performance predictions. The method is superior to traditional models in terms of prediction accuracy.
Foots et al. [22]	Classify vendors based on estimated CPU performance. Predict CPU performance using hardware components.	Neural networks predict CPU performance accurately with low error. Cubic regression outperforms neural networks with more training data.

The rapid advancement of heterogeneous computing environments, where CPUs with varying architectural designs are integrated, poses significant challenges for effective task scheduling. These systems demand intelligent prediction mechanisms to estimate application performance accurately across diverse resources, enabling optimal CPU scheduling and improved system throughput. Traditional models used for this purpose, such as standalone Support Vector Regression (SVR) and Lasso Regression (LAS), often lack the adaptability required to capture the complex and non-linear characteristics of dynamic workloads. Furthermore, without sophisticated parameter tuning or adaptive learning, their predictive performance may plateau, limiting their practical deployment in real-time systems. To overcome these limitations, this study proposes a hybrid machine learning framework that enhances the predictive capacity of SVR and LAS models by integrating two recent metaheuristic optimization algorithms: the Crocodiles Hunting Strategy (CHS) and the Walrus Optimization Algorithm (WOA). These algorithms are designed to perform global optimization of model hyperparameters, which plays a crucial role in improving prediction accuracy, reducing overfitting, and enhancing generalization on unseen data. Accordingly, the objective of this research is fourfold. First, to construct

baseline predictive models using SVR and LAS for estimating CPU workload performance. Second, to develop hybrid versions of these models, SVR + CHS (SVCH), SVR + WOA (SVWO), LAS + CHS (LACH), and LAS + WOA (LAWO), by integrating the optimizers into the training pipeline. Third, to rigorously evaluate and compare the performance of all models using standard metrics, including Root Mean Square Error (RMSE), Mean Squared Error (MSE), Coefficient of Determination (R^2), Symmetric Mean Absolute Percentage Error (SMAPE), and Coefficient of Variation of RMSE (CV_RMSE). Finally, the study aims to analyze the practical implications of the hybrid models in real-world CPU scheduling tasks by examining their predictive power, computational cost, and potential for real-time implementation. Through this hybridization approach, the study seeks to demonstrate measurable improvements in prediction accuracy while offering models that are robust, generalizable, and interpretable, especially important for resource-aware decision-making in complex, heterogeneous computing systems. SVR and LAS were selected as the foundational models in this study due to their complementary strengths in handling high-dimensional and complex datasets. SVR is well-known for its robustness in modeling non-linear relationships through kernel functions, making it effective in capturing intricate patterns in data without overfitting. Its ability to generalize well in small to medium-sized datasets makes it suitable for the present CPU performance prediction task. On the other hand, LAS offers a robust feature selection mechanism by enforcing sparsity in the model coefficients through L1 regularization. This helps reduce model complexity and improves interpretability, particularly when dealing with datasets that may contain redundant or irrelevant features.

To guide the study, the following research questions are posed:

RQ1: Can the integration of metaheuristic optimization algorithms (CHS and WOA) significantly improve the predictive accuracy of ML models for CPU workload estimation?

RQ2: How does each hybrid model (SVCH, SVWO, LAWO, LACH) compare to its standalone counterpart in terms of RMSE and R^2?

RQ3: What are the trade-offs between computational cost and prediction accuracy in hybrid vs. non-hybrid models?

RQ4: Does the LACH model offer meaningful interpretability advantages due to Lasso's inherent feature selection properties?

RQ5: How can accurate CPU performance prediction models contribute to improved system throughput and resource utilization in heterogeneous scheduling scenarios?

Based on the objectives and prior work, the following hypotheses are proposed:

H1: Hybrid ML models optimized with CHS or WOA will outperform standalone SVR and LAS models in terms of RMSE, R^2, and MSE.

H2: The LACH model will yield the highest prediction accuracy among all proposed schemes due to the synergistic effect of LAS's regularization and CHS's adaptive exploration.

H3: Lasso-based hybrid models will provide better interpretability than SVR-based models due to the sparsity and feature reduction properties of Lasso Regression.

H4: While hybrid models incur higher training-time computational costs, they will generalize better on unseen workloads, making them suitable for deployment in real-time heterogeneous CPU schedulers.

2 Mathematical schemes

2.1 Support vector regression (SVR)

The SVM was first developed by Vapnik VN to tackle categorization issues [23]. Nevertheless, it was later changed to handle regression issues as well. SVM regression is based on the structural risk minimization (SRM) idea, which is far more enhanced than the conventional empirical risk minimization (ERM) methodology [24]. Because it attempts to reduce the upper bound of generalization error, a critical component of the learning process, the SRM principle is essential to statistical learning. An expansion of SVM created especially for regression problems is called SVR [25]. While SVR and SVM use comparable methods, their purposes for parameter estimation are distinct. The usage of slack variables is where the 2 differ most from one another. For classification tasks, SVMs provide a single slack variable to each training data point; for regression tasks, SVR assigns 2 slack variables to each training data point. One unique feature of SVR is that, in contrast to other schemes, it can improve generalization performance and hasten the attainment of the best global resolution [26].

2.1.1 Linear SVR

The local linear regression version of SVR maybe described by Eq. (1) given a training database $\{y_i, x_i, i = 1, 2, 3 \ ... \ n\}$, where y_i signifies the output vector, x_i displays the input vector, and n indicates the size of the database.

$$f(x, k) = k \times x + b \tag{1}$$

The given equation illustrates the dot product as (x, k), where k symbolizes the weight vector, x displays the normalized test pattern, and b denotes the bias. SRM theory aims to minimize the empirical risk, as represented by $R_{emp}\ (k, b)$. Eq. (3) illustrates how an ε-insensitive loss function, known as $L_\varepsilon(y_i, f(x_i, k))$, is used to calculate the empirical risk.

$$R_{emp}(k, b) = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(y_i, f(x_i, k)\right) \tag{2}$$

$$L_\varepsilon\big(y_i, f(x_i, k)\big) = \begin{cases} \varepsilon, & if \quad |y_i - f(x_i, k)| \le \varepsilon \\ |y_i - f(x_i, k)| - \varepsilon, & otherwise \end{cases} \tag{3}$$

During optimization, the loss function $L(\varepsilon)$, where x_i is the training pattern, gauges the tolerance error between the objective output (y_i) and the estimated output values $f(x_i; w)$. By reducing $||k||^2$, the ε-insensitive loss function may be utilized to resolve SVR model complexity in linear regression problems. Furthermore, it is feasible to estimate the training data deviation outside of the ε-zone by introducing a non-negative slack variable, φ_i.

$$\underset{k,b,\varphi,\varphi^*}{\text{Lim}}\left[\frac{1}{2}k.k + c\left(\sum_{i=1}^{n}\varphi_i^* + \sum_{1=1}^{n}\varphi_i\right)\right]$$
$$\text{Subjected to,}\begin{cases} y_i - k.x_i - b \le \varepsilon + \varphi_i^* \\ k.x_i + b - y_i \le \varepsilon + \varphi_i \ , i \\ \varphi_i^*, \varphi_i \ge 0 \end{cases} \quad (4)$$
$$= 1, \ldots, n$$

Finding the Lagrange function's saddle point is pivotal to resolving the earlier issue.

$$L(k, \varphi^*, \varphi, \alpha^*, \alpha, c, \gamma^*, \gamma = \frac{1}{2}k.k + c\left(\sum_{i=1}^{n}\varphi_i^* + \sum_{1=1}^{n}\varphi_i\right) - \sum_{i=1}^{n}\alpha_i[y_i - k.x_i - b + \varepsilon + \varphi_i\ ] \quad (5)$$
$$- \sum_{i=1}^{n}\alpha_i^*[k.x_i + b - y_i + \varepsilon + \varphi_i^*\ ] - \sum_{1}^{n}(\gamma_i^*\varphi_i^* + \gamma_i\varphi_i)$$

By using the KKT conditions, the Lagrange function may be reduced. This requires partial differentiation of Eq. (5) concerning k, b, φ_i^*, and φ_i.

$$\frac{\delta \mathrm{L}}{\delta \mathrm{k}} = \mathrm{k} + \sum_{i=1}^{n}\alpha_i x_i \quad (6)$$
$$-\sum_{i=1}^{n}\alpha_i^* x_i = 0, k = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)x_i$$

$$\frac{\delta \mathrm{L}}{\delta \mathrm{b}} = \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i^* = 0, \sum_{i=1}^{n}\alpha_i = \sum_{i=1}^{n}\alpha_i^* \quad (7)$$

$$\frac{\delta \mathrm{L}}{\delta \varphi^*} = \mathrm{c} - \sum_{i=1}^{n}\gamma_i^* - \sum_{i=1}^{n}\alpha_i^* = 0, \sum_{i=1}^{n}\gamma_i^* = c - \sum_{i=1}^{n}\alpha_i^* \quad (8)$$

$$\frac{\delta \mathrm{L}}{\delta \varphi} = \mathrm{c} - \sum_{i=1}^{n}\gamma_i - \sum_{i=1}^{n}\alpha_i = 0, \sum_{i=1}^{n}\gamma_i = c - \sum_{i=1}^{n}\alpha_i \quad (9)$$

Eqs. (6) and (1) have a relationship with the parameter k. The dual optimization function is stated as follows when Eq. (6) is substituted into the Lagrange function of Eq. (5):

$$max_{\alpha,\alpha^*}[k(\alpha, \alpha^*)] = max_{\alpha,\alpha^*}\left[\sum_{i=1}^{n}\gamma_i(\alpha_i^* - \alpha_i) - \varepsilon\sum_{i=1}^{n}(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{ij=1}^{n}(\alpha_i^* - \alpha_i)(\alpha_i^* - \alpha_i)(x_i.x_j)\right] \quad (10)$$
$$subjected\ to\begin{cases} \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\ 0 \le \alpha_i^*, \alpha_i \le 0 \end{cases} i$$
$$= 1, \ldots, n$$

The optimization problem is defined by utilizing the Lagrange multipliers α_i^* and α_i [27]. Upon solving Eq. (10) concerning the restrictions in Eq. (11), the final linear regression function is displayed as:

$$f(x, \alpha^*, \alpha) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)(x_i, x) + b \quad (11)$$

Fig. 1 presents the SVR's flowchart.

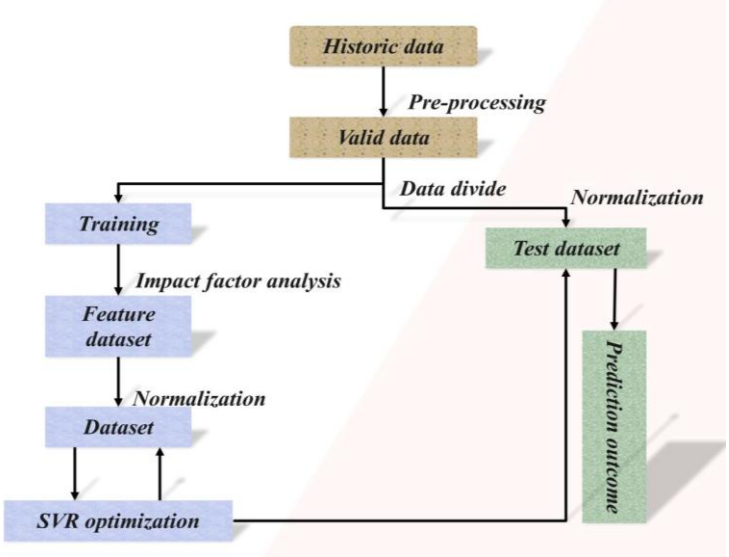


Figure 1: The SVR's flowchart

2.2 Lasso regression (LAS)

LAS aims to minimize prediction error by identifying relevant variables and their associated regression coefficients. This is achieved by shrinking the regression coefficients toward zero by restricting the model factors [28]. Making sure that the total of the regression coefficients' absolute values stays below a preset threshold (λ) is the responsibility of this constraint. This constraint practically reduces the model's complexity. As a result, variables with regression coefficients that approach 0 are eliminated from the model. A common automated technique for selecting λ is the k-fold cross-validation approach. Using this approach, the database is separated into k equal-sized subsets at random [29]. The other $k-1$ subsets are used to build a prediction model, with one subset set aside for validation. Each of the k subsets is alternatively allocated for validation during the k cycles of this procedure, with the remaining subsets being used for model building. The culmination of these cycles involves consolidating the individual validation results across various λ values, ultimately selecting the optimal λ to define the final model. One important benefit of this approach is that it can reduce overfitting without restricting any subset of the database to be used for internal validation [30].

Although LAS has outperformed standard techniques in some situations, optimism bias and overfitting cannot be eliminated by using it. It does not remove the need to use an external database for model validation. Furthermore, for a better overall prediction expectancy, the LASSO technique trades off potential bias in projecting individual factors. One associated disadvantage is that since the focus is on optimizing combined prediction rather than the accuracy of projecting and interpreting the contribution of individual variables, the regression coefficients may not be reliably interpretable regarding independent risk factors. There are variations on the classic LASSO method, like elastic net and ridge regression, and research on the relative benefits of penalized regression methods is still ongoing.

2.3 Walrus optimization algorithm (WOA)

The natural behaviors of walruses dictate 3 distinct stages in which their positioning within the WOA is simulated.

Stage 1: Feeding approach (exploration).

A wide variety of marine invertebrates, including sea cucumbers, shrimp, soft corals, tunicates, tube worms, and various mollusks, are consumed by walruses. They consume over 60 distinct kinds of marine invertebrates. Still, they seem to choose benthic bivalve mollusks, particularly clams. They graze along the seafloor in search of these, using their sensitive vibrissae and vigorous flipper movements to find and detect food. During this excursion, the prominently tusked walrus takes on the position of leader, leading the group on their food search. There exists a correlation between the walrus tusk length and the quality of the objective function values of possible solutions. The possible resolution that has the best objective function value is therefore recognized as the dominant walrus in the group. Because of their foraging habits, walruses can explore a wider range of regions in the search domain, which improves the WOA's overall search capabilities. Eqs. (12) and (13) describe the mathematical modeling of the walruses' position update process, which is based on their feeding mechanism and is guided by the most influential group member. In this process, the walrus's initial position is first calculated using Eq. (12). As stated in Eq. (13), the newly found location takes the place of the previously determined one if it increases the value of the goal function.

$$x_{i,j}^{p1} = x_{i,j} + rand_{i,j}.(sw_j - I_{i,j}.x_{ij}) \tag{12}$$

$$X_i = \begin{cases} x_{i,j}^{p1}, F_{i,j}^{p1} < F_i \\ x_i, \ else \end{cases} \tag{13}$$

X_i^{P1} pertains to the newly calculated position of the $i-th$ walrus in the context of the first stage. Within this, F_i^{P1} indicates the objective function value associated with this specific walrus, and $x_{i,j}^{P1}$ displays its $j-th$ dimension. The variables $rand_{ij}$ represent random numbers falling within the [0, 1] range. SW corresponds to the leading possible resolution, symbolizing the most influential walrus. Furthermore, I_{ij} involves the selection of random integers within the range of 1 to 2. The role of I_{ij} is to enhance the algorithm's exploration capabilities. When it is set to 2, it results in more significant and extensive adjustments to the walruses' positions, as opposed to the default value of 1, which displays the standard degree of displacement. Enhancing the algorithm's global search, these conditions enable it to escape local optima and reveal the primary optimal region within the problem-solving space.

Stage 2: Migration

Walruses naturally go toward rocky outcrops and beaches as the temperature increases in the late summer. The WOA makes use of this migratory tendency to direct the walruses across the search domain and assist them in finding appropriate locations for exploration. Each walrus that is arbitrarily chosen from a distinct area of the search domain is thought to go to a separate location in this modeling. As a result, the new position was originally established using Eq. (14). Then, in line with Eq. (15), the walrus's previous location is replaced if this new position increases the value of the goal function. Eqs. (14) and (15) provide a mathematical representation of this behavioral mechanism.

$$x_{i,j}^{p2} = \begin{cases} x_{i,j} + rand_{i,j}.\left(x_{kj} - I_{ij}.x_{ij}\right), F_k < F_i \\ x_{i,j} + rand_{i,j}.\left(x_{i,j} - x_{kj}\right), else, \end{cases} \tag{14}$$

$$x_i = \begin{cases} x_i^{p2}, F_i^{p2} < F_i, \\ x_i, \ else, \end{cases} \tag{15}$$

Within this context, X_i^{P2} symbolizes the freshly calculated position for the $i-th$ walrus during the second stage. Here, x_{ij}^{P2} pertains to its $j-th$ dimension, while

F_i^{P2} corresponds to its objective function value. Additionally, X_k, where $k \in \{1, 2, \ldots, N\}$ and $k \neq i$, designates the location of the chosen walrus to which the $i-th$ walrus migrates. x_{kj} displays the $j-th$ dimension of this selected walrus, and F_k displays its respective objective function value.

Stage 3: Evasion and confrontation with predators (Exploitation)

The WOA improves its exploitation capabilities by imitating this natural behavior, which also helps it better explore the local search domain around potential solutions. Walruses are often in danger of coming into conflict with killer whales and polar bears. The walruses' positions change as a result of the tactics they use to avoid and defend against these predators; typically, these changes occur in the vicinity of their present locations. The WOA's design assumes that, as this process plays out close to each walrus's position, the range of walrus position changes takes place inside a neighborhood with a defined radius centered around each walrus. To emphasize global search in the early algorithm cycles and identify the ideal location in the search domain, the radius of this neighborhood is changeable. As the algorithm iterates, it starts with the highest value and gradually gets smaller. To solve this, the WOA includes local lower and upper boundaries at this stage, which causes the radius to change with each algorithm cycle. A neighborhood is established around every walrus in the WOA to simulate this behavior. First, using Eqs. (16) and (17), a new location is created at random inside this neighborhood. According to Eq. (18), this new position replaces the previous one if the objective function's value grows later.

$$x_{i,j}^{p3} = x_{ij} + (lb_{localj}^{t} + (ub_{localj}^{t} - rand.lb_{localj}^{t}) \quad (16)$$

$$local\ bounds = \begin{cases} lb_{localj}^{t} = \dfrac{lb_j}{t} \\ ub_{localj}^{t} = \dfrac{ub_j}{t} \end{cases} \quad (17)$$

$$x_i = \begin{cases} x_i^{p3}, F_i^{p3} < F_i \\ x_i, \quad else \end{cases} \quad (18)$$

In this context, X_i^{P3} signifies the freshly determined position for the $i-th$ walrus in accordance with the third stage. Here, x_{ij}^{P3} denotes its $j-th$ dimension, while F_i^{P3} corresponds to its objective function value. The variable t displays the cycle step, while lb_j and ub_j denote the lower and upper thresholds for the $j-th$ variable. Additionally, lb_{localj}^{t} and ub_{localj}^{t} are the local lower and local upper thresholds specifically applicable to the $j-th$ variable. These local thresholds simulate a localized search within the vicinity of the possible resolutions. Algorithm 1 shows the pseudocode of WOA.

Algorithm 1: Pseudocode of WOA.

Start WOA
  Input all optimization problem information.
  Set walrusses' number (N) and the iterations' total number (T).
  Initialization locations of walruses' process.
  For $t = 1{:}N$
  Phase 1: Exploration
    Calculate new location of the j[th] walrus
    Update the i[th] walrus location
  Phase 2: Migration
    Choose an immigration destination for the i[th] walrus
    Calculate new location of the j[th] walrus
    Update the i[th] walrus location
  Phase 3: Exploitation
    Calculate new position in the neighborhood of the j[th] walrus
    Update the i[th] walrus location
  end
  Save the best candidate solution so far.
End WOA.

2.4 Crocodiles hunting strategy (CHS)

This section examines the optimization strategy known as CHS, which was first used as one of the feature clustering procedures [31].

2.4.1 Initializing

The initializing step must be finished before proceeding to the main stages, much like with other metaheuristic algorithms [32]. In the initializing process, a number of initial solutions are generated at random. The initial population of crocodiles consists of these responses that were chosen at random. These solutions have a homogeneous random distribution generated between the lower and upper thresholds. These solutions are obtained by using the following equation:

$$x = LB + r \times (UB - LB) \quad (19)$$

After basic factors such as population size, maximum number of cycles, and lower and upper thresholds of variables have been defined, random solutions (x) are

generated based on (2), where LB and UB are the problem's lower and upper limits, respectively. Moreover, r has a generation interval of $0\ to\ 1$ and is a uniform random variable. These solutions are then evaluated using the objective function. CHS operators evaluate the solutions using the objective function. The ideal solution gets replaced when a better one comes along. The optimal cure (x_{prey}) displays the smallest objective function value of the superior resolution.

Chasing the Prey

This section mimics what the chasers would do. As previously said, the population is divided in half. Thus, each sector displays half of the population. Included in the group of chasers are the first 50% of the answers. Therefore, ambushers make up the latter half of fifty percent of the population. These 2 groups are arbitrarily divided into 2 ambushers and 2 chaser groups. The reasoning behind replicating chaser behavior is based on the distance between prey and crocodiles that resembles that of chasers. The prey is led to the shore and other shallow regions by another set of hunters called the chasers, who pursue it without actually catching it. The following are the recommended equations:

$$d^{i,t} = \left|x_{prey}^{t} - x_{chaser}^{i,t}\right|,\ \ for\ all\ i\ , \tag{20}$$

$$\begin{cases} x_{chaser}^{i,t+1} = \left(x_{chaser}^{i,t} - \beta.d\right)\ \ for\ all\ i(\dfrac{\ln(it)}{\ln\left(\max it\right)} \times r) \geq 0.5 \\ x_{chaser}^{i,t+1} = |\beta - (\beta.d)|\ \ for\ all\ i(\dfrac{\ln(it)}{\ln(maxit)} \times r < 0.5 \end{cases} \tag{21}$$

In this instance, the prey location at cycle t is represented by x_{prey}^{t}, while the chaser's position at cycle t is reflected by $x_{chaser}^{i,t}$. The coefficient β has a uniform distribution that ranges from $[-3,3]$. This r is an arbitrary number in the interval $0-1$. It also refers to the local and global cycle numbers in inner loops. After running the program several times, these intervals were measured empirically. Eq. (21) provides the distance d between the prey and its chase. 2 requirements are satisfied by Eq. (21): first, if it$\frac{it}{maxit} \times r \geq 0.5$, which indicates that every chaser. A random search is carried out if the movement of the first group toward the prey had a positive coefficient and it reached its maximum when r was less than 0.5. To sum up, Eq. (21) shows that by employing an intelligent technique, the second is more fully implemented, and a random search is carried out in the initial cycles. In the most recent versions, it is more fully implemented, and the chasers approach the target at random.

2.4.2 Attacking the prey

The ambushers are poised to grab their victim at the final destination of their objective. While the pursuers try to guide the victim to this spot or the attack area, the ambushers hide in the final position. Attackers are stated to be forced to change positions in order to replicate the attacking stage by using the following equations:

$$d^{i,t} = \left|x_{prey}^{t} - x_{ambusher}^{i,t}\right|, for\ all\ i, \tag{22}$$

$$A = \frac{apc + apa + x_{prey}}{3} \tag{23}$$

$$\begin{cases} x_{ambusher}^{i,t+1} = (d.\cos(2\pi r) + x_{prey}^{t}\ \ for\ all\ i\ (p) \geq 0.5 \\ x_{ambusher}^{i,t+1} = (x_{ambusher}^{i,t} - \beta\left(A - x_{ambusher}^{i,t}\right))\ \ for\ all\ i\ (p) < 0.5 \end{cases} \tag{24}$$

where $p = \frac{\ln\left(it\right)}{\ln\left(maxit\right)} \times r$, in this case, d indicates the separation between the ambushers and the prey, and $x_{ambusher}^{i,t}$ displays the position of ith ambusher at cycle t. Additionally, x_{prey} is the best position or prey position, apa is the average position of ambushers, and apc is the average position of chasers. Crocodiles adjust their posture based on average position or the locations of their prey in every category. Thus, if$(\frac{\ln(it)}{\ln(,axit)} \times r < 0.5$, meaning that every ambusher rotates around the prey; if $\frac{\ln(it)}{\ln(,axit)} \times r \geq 0.5$ ambushers adjust their position based on the average positions of all groups of chasers, ambushers, and prey. The flowchart of the CHS is presented in Fig. 2.

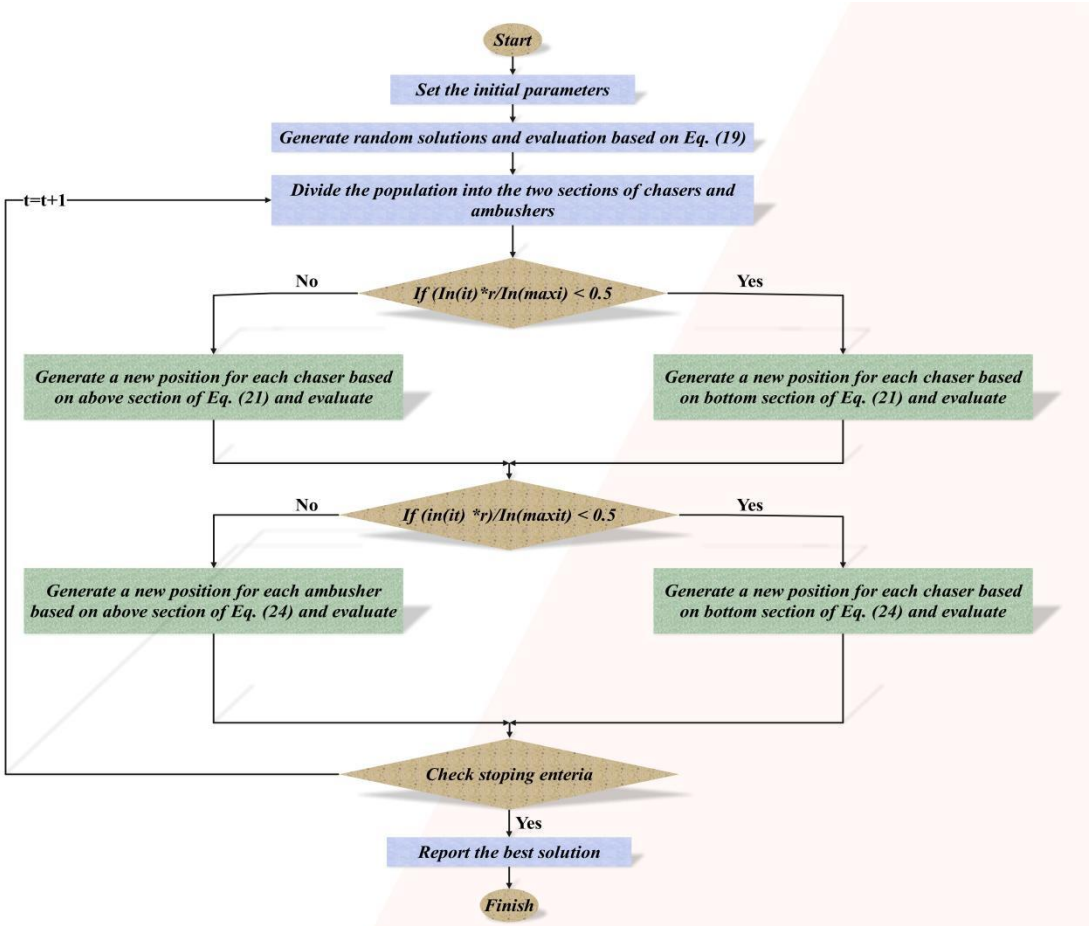


Figure 2: The flowchart of the CHS.

2.5 Performance evaluators

Several performance metrics, such as the coefficient correlation (R^2), symmetric mean absolute percentage error (SMAPE), MSE, RMSE, and coefficient variance of RMSE (CV_{RMSE}), have been used in this study to evaluate the efficacy of the model. Below are the formulas for these performance standards:

$$SMAPE = \frac{100}{n}\sum_{i}^{n}\frac{2\times|m_i - b_i|}{|m_i|+|b_i|} \tag{25}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(b_i-\bar{b})(m_i-\bar{m})}{\sqrt{\left[\sum_{i=1}^{n}(b_i-\bar{b})^2\right]\left[\sum_{i=1}^{n}(m_i-\bar{m})^2\right]}}\right)^2 \tag{26}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i-b_i)^2} \tag{27}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(m_i-b_i)^2 \tag{28}$$

$$CV_{RMSE} = \left(\frac{RMSE}{\bar{t}}\right)\times 100 \tag{29}$$

$$CI = \left(\sqrt{\frac{n\times metric^2}{X^2_{a/2,n}}}, \sqrt{\frac{n\times metric^2}{X^2_{1-a/2,n}}}\right) \tag{30}$$

It is also possible to express the variables as follows:

- n is a symbol for sample size.
- b_i is the expected value.
- The measured and mean predicted values are denoted by $\bar{m}$ and $\bar{b}$, respectively.
- m_i displays the gauged value.
- t displays the degrees of freedom, and the crucial value of the t-distribution depends on the desired level of confidence.
- $X^2_{a/2,n}$ and $X^2_{1-a/2,n}$ are critical values from the chi-squared distribution with n degrees of freedom.
- a shows significance level.
- $metric$ represents a component that can be placed on each evaluation to determine the CI.

3 Data collection

The variables are presented as factors to describe and analyze computer schemes or systems. “Model Name” serves as a unique identifier for a specific computer model, while “MYCT” displays the machine cycle time, indicating the duration of one instruction cycle in nanoseconds. “MMIN” and “MMAX” denote the minimum and maximum main memory capacities, respectively, measured in kilobytes. “CACH” stands for cache memory size in kilobytes, storing frequently

accessed data for faster CPU access. “CHMIN” and “CHMAX” signify the minimum and maximum numbers of channels or connections for I/O devices. Finally, “PRP” displays the published relative performance, a benchmark score comparing the performance of one model against others. These variables collectively provide insights into the specifications, performance, and capabilities of various computer schemes, aiding in performance analysis, system optimization, and comparative evaluations.

The dataset contains 209 samples, which were gathered from Kaggle (https://www.kaggle.com/datasets/abdelazizsami/computer-hardware).

Before modeling, a basic preprocessing pipeline was applied:

- No missing values were found in the dataset, so no imputation was required.
- All feature values were numerical, eliminating the need for encoding.
- Features were normalized using Min-Max scaling to ensure that the optimization algorithms and learning models operate within the same scale, improving convergence and stability.

The dataset was randomly split into 70% for training (145 samples) and 30% for testing (64 samples) to evaluate generalization performance. Additionally, during the model development phase, a 5-fold cross-validation strategy was employed on the training set to fine-tune model hyperparameters and assess performance consistency. This ensures that the models are robust and not overfitted to any specific subset of the data. The dataset was split into 70% for train and 30% for testing. Fig. 3 displays the 3D bar for the correlation of the input and output variables.

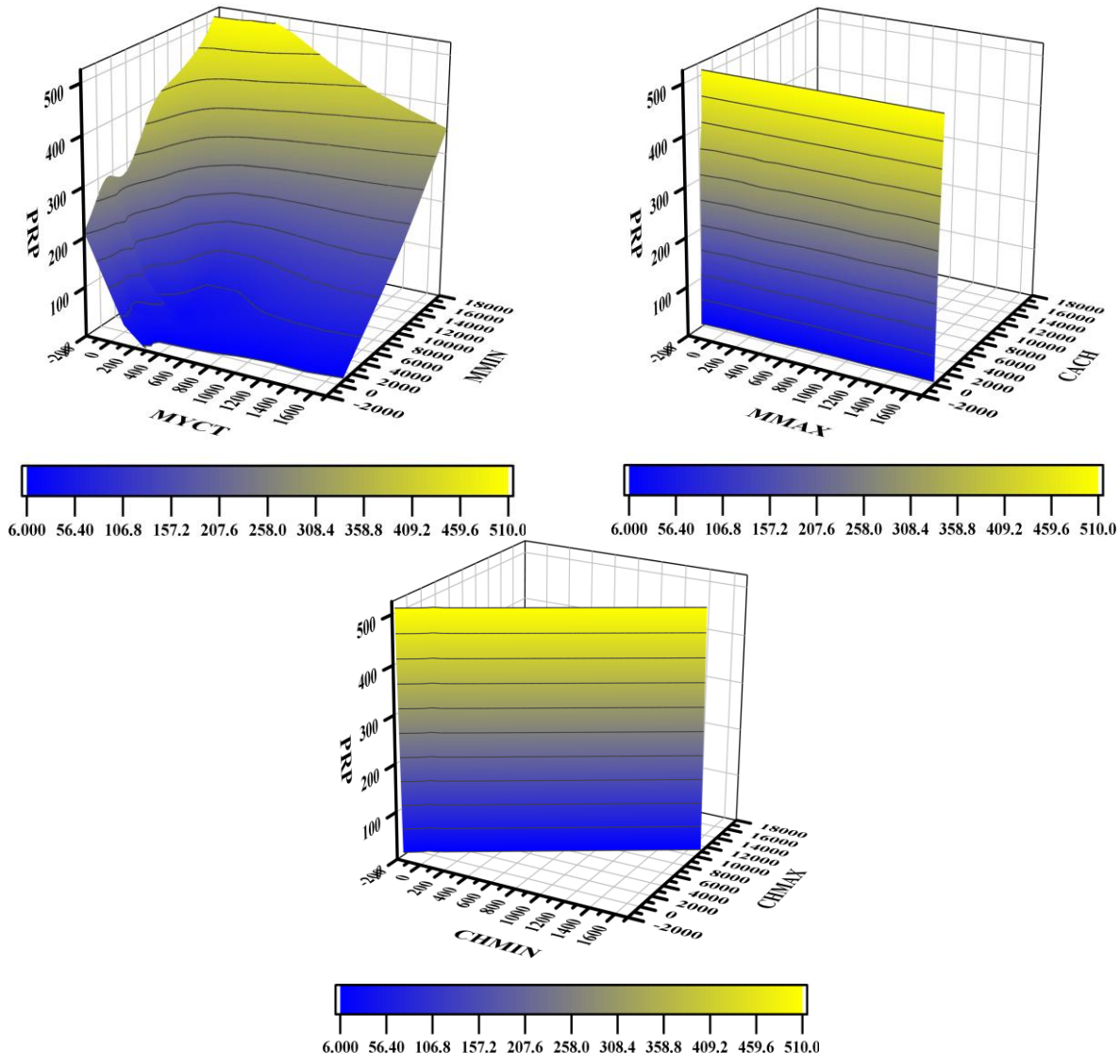


Figure 3: 3D surface plot for the correlation of input and output

4 Results

4.1 Convergence and hyperparameter

Fig. 4 shows the convergence curve for the hybrid schemes. It provides information about how the algorithm is performing as it approaches the optimal solution by visually representing the dynamic variations in the objective function value for successive cycles. The objective function frequently exhibits a constant trend of growing or lowering as convergence gets closer until it reaches a point where doing more cycles marginally improves the situation. Significant information about the algorithm's convergence behavior and best-answer path is provided by this graphical depiction. The graph makes evident the differences in operation between the LACH and LAWO schemes. The LACH model shows great promise, achieving an accuracy of 12.134 after around 135 cycles. Conversely, the LAWO model requires around 110 cycles to reach its highest accuracy of 17.139. After 140 cycles, the SVCH model achieves its maximum accuracy of 27.429, whereas the SVWO model gets its accuracy of 30.313 after 125 iterations.

As expected, the single models (SVR and Lasso) demonstrate significantly lower computational overhead. SVR completes in approximately 0.5 seconds, and Lasso runs even faster, requiring only 0.3 seconds. This is due to their deterministic nature and lack of iterative optimization beyond the internal model fitting. In contrast, the hybrid models, which involve repeated evaluation and optimization of hyperparameters using CHS or WOA, require considerably more time. Among these, SVR + CHS is the most time-consuming, with a runtime of 40 seconds, likely due to both the complexity of SVR and the iterative search process of CHS. SVR + WOA is slightly more efficient, requiring 32 seconds, indicating that WOA may converge faster or be computationally lighter than CHS in this context. On the Lasso side, the Lasso + CHS hybrid consumes 26 seconds, while Lasso + WOA is the fastest among hybrids, completing in 20 seconds. This performance difference may stem from Lasso's inherently simpler structure and faster convergence when compared to SVR, combined with WOA's relatively lower computational footprint.

In addition, Table 2 presents the optimized hyperparameters for both SVR-based and Lasso-based models. For SVR variants, the metaheuristic-optimized models (SVCH and SVWO) show significant improvements over the baseline SVR, with higher penalty values (C) and tuned epsilon (ε) settings. SVCH uses a larger C and moderate ε, promoting lower training error, while SVWO opts for tighter tolerance (ε = 0.001) for precision. Both hybrids use γ = 1 for kernel influence. For Lasso-based models, the regularization parameter α varies: LAS uses a default value of 1, while LAWO and LACH use higher values (50.0 and 13.5, respectively) optimized by WOA and CHS. LACH's moderate α suggests a balanced trade-off between sparsity and model flexibility, aligning with its top performance. Overall, the metaheuristic algorithms effectively fine-tune hyperparameters to enhance prediction accuracy beyond the baseline configurations. Notably, the parameters are selected based on the random search method.

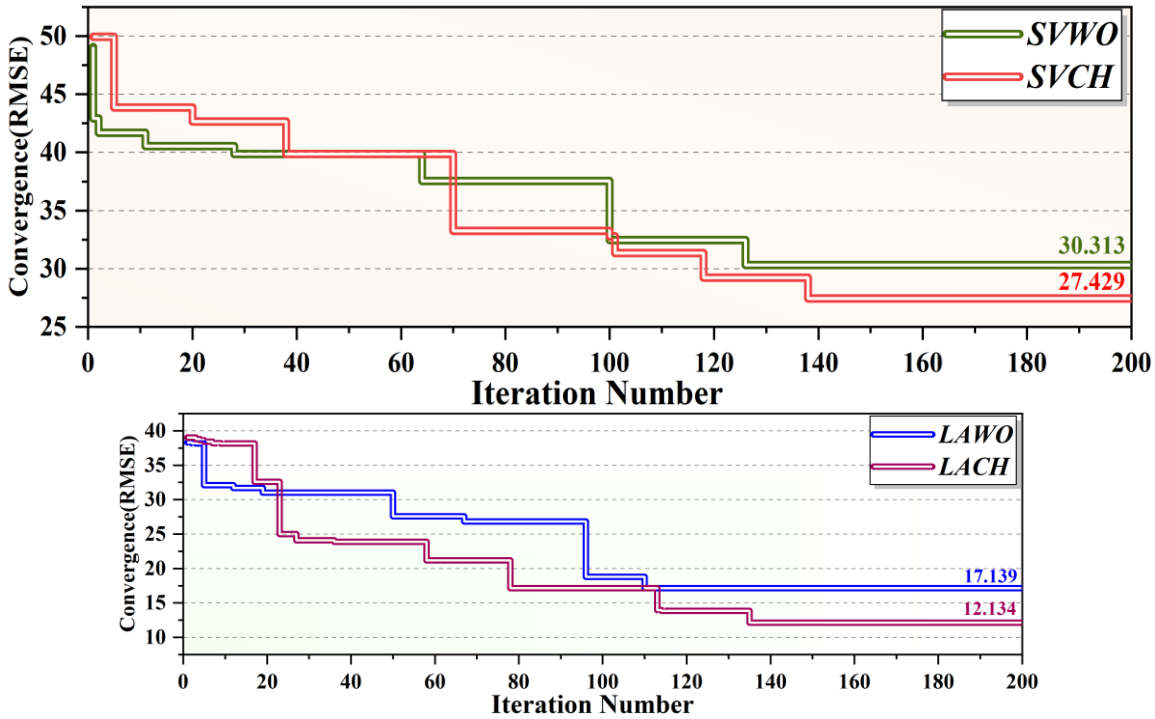


Figure 4: Convergence plot of hybrid schemes

Table 2: Result of the hyperprameters

Models	Hyperprameter		
	C	Epsilon	Gama
SVCH	415.5	0.339	1
SVWO	169.2	0.001	1
SVR	1	0.1	scale
	alpha		
LACH	13.50		
LAWO	50.00		
LAS	1		

4.2 Schemes comparison

The results of the generated schemes, which were evaluated using 5 metrics and 3 different stages, are shown in Table 3. The stages encompass training, validation, and testing, while the metrics comprise RMSE, R^2, MSE, and SMAPE. During the training stage, LACH demonstrated superior performance regarding RMSE, achieving a value of 13.228, followed closely by the LAWO scheme with a performance metric of 18.780, ranking second in performance.

Transitioning to the validation stage, the LACH model showcased exceptional performance with an R^2 value of 0.993, securing the top position. Conversely, the LAS model emerged as the third-best performer with an R^2 value of 0.985. Upon entering the testing stage, the SVCH scheme exhibited the 4th-best performance, attaining an MSE value of 638.8. Subsequently, the SVWO model trailed closely behind with the fifth-best performance, registering an MSE value of 809.3. In the Validation stage, the LACH model excelled in SMAPE, achieving a value of 0.002, while the LAWO model attained the second-highest performance with a value of 0.003. Furthermore, regarding MSE during the validation stage, the LAS model ranked third with a value of 234.6, while the SVR model demonstrated the weakest performance, recording an MSE value of 1074.8.

Notably, the code of models and optimizers provided in GitHub (https://github.com/Chliu32/Model-and-optimizer).

Table 3: The result of developed schemes for LAS and SVR

Model	Stage	Index values						
		RMSE	R^2	MSE	SMAPE	CV_{RMSE}	CI_{R^2}	CI_{RMSE}
SVR	Train	35.492	0.898	1259.7	0.002	0.377	0.8618	23.145
	Validation	32.785	0.915	1074.8	0.009	0.318		
	Test	31.431	0.660	987.9	0.018	0.620		
SVWO	Train	31.551	0.928	995.5	0.002	0.335	0.8899	20.655
	Validation	25.985	0.946	675.2	0.007	0.252		
	Test	28.449	0.737	809.3	0.017	0.561		
SVCH	Train	27.377	0.940	749.5	0.002	0.291	0.7453	31.416
	Validation	28.880	0.948	834.0	0.008	0.280		
	Test	26.150	0.812	683.8	0.014	0.516		
LAS	Train	21.710	0.963	471.3	0.001	0.231	0.9798	8.854
	Validation	15.317	0.985	234.6	0.004	0.149		
	Test	13.859	0.976	192.1	0.006	0.273		
LAWO	Train	18.780	0.973	352.7	0.001	0.199	0.9560	13.059
	Validation	12.407	0.987	153.9	0.003	0.120		
	Test	12.716	0.975	161.7	0.006	0.251		
LACH	Train	13.288	0.987	176.6	0.001	0.141	0.9411	15.110
	Validation	8.778	0.993	77.1	0.002	0.085		
	Test	9.068	0.990	82.2	0.004	0.004		

Fig. 5 shows a plot of the dispersion of emerging hybrid schemes. The measured values are shown on the X axis, while the predicted values are displayed on the Y axis. The optimal performance of the model is indicated by the colored zone surrounding the R^2 value, often known as the center line. To provide clarity, Fig. 5 shows this shading in several tones. Data points that are below the center line are shown as underestimation, while those that are above it are shown as overestimation. The model is considered to be functioning correctly when there is a minor angle between the linear line and the center line. The productivity of the LACH scheme exceeds that of the LAWO, LAS, SVCH, SVWO, and SVR schemes, as shown in Fig. 5. The second-best performance is exhibited by the LAWO model, which trails the LAS model by a narrow margin.

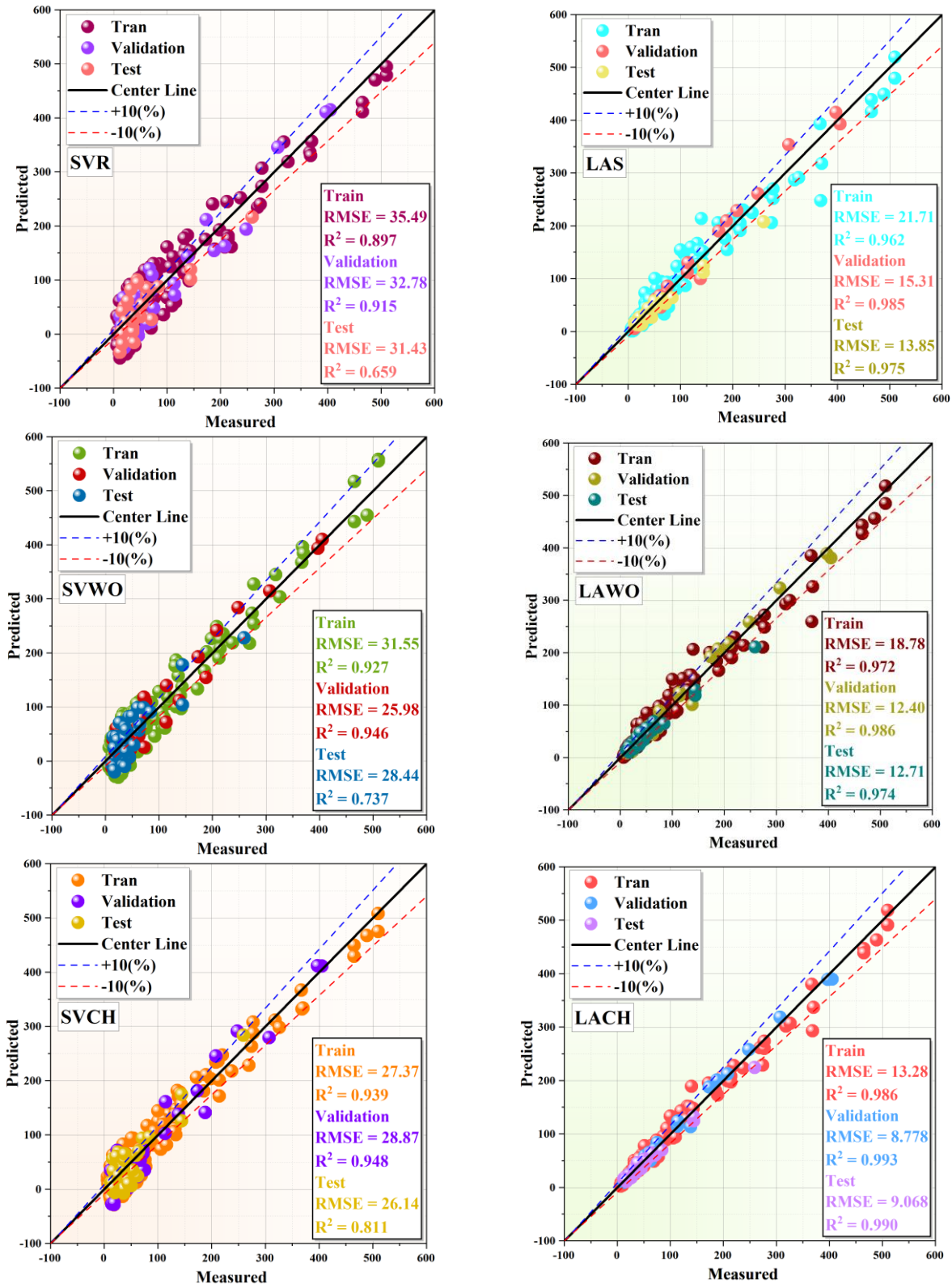
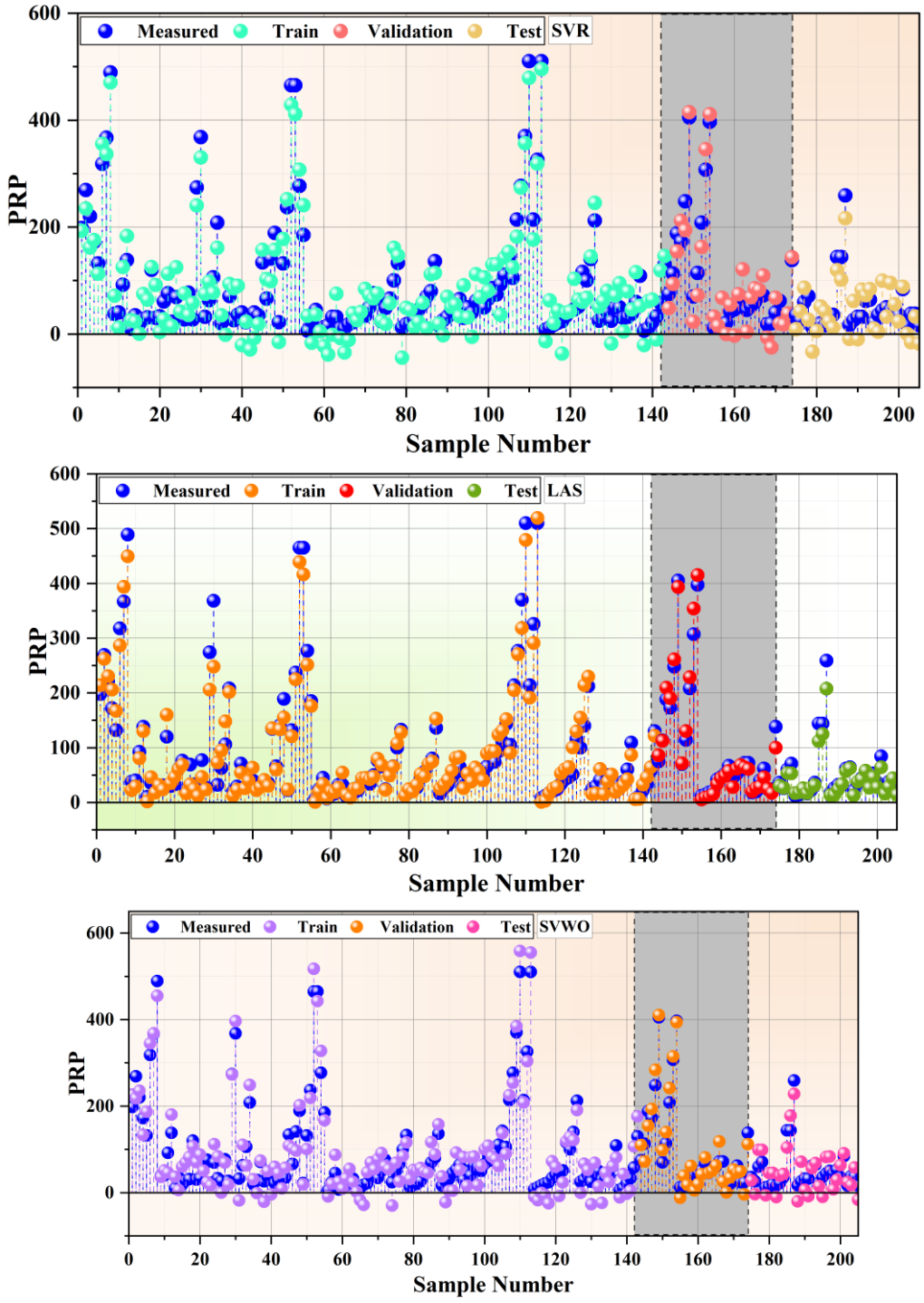


Figure 5: Plotting the dispersion of evolved hybrid schemes

A comparison between estimated and actual values is presented in Fig. 6. The line that should match the measured line is the visual cue that shows how accurate the forecast made by the model is. When the measured line and the anticipated column are almost in line, there is a strong sign of accuracy; when not, there may be inefficiencies. The LACH model demonstrated strong performance during training, with very few measured lines displaying lengths larger than the predicted column. This model performed better in the validation stage than in the test stage, and its predictions were rather close to the actual data. In contrast to the LACH model, the LAWO model exhibited weaker performance during the training stage, as evidenced by the fewer predicted columns

deviating from observed lines. Nevertheless, the LAWO model excelled in both the test and validation stages, demonstrating strong performance.



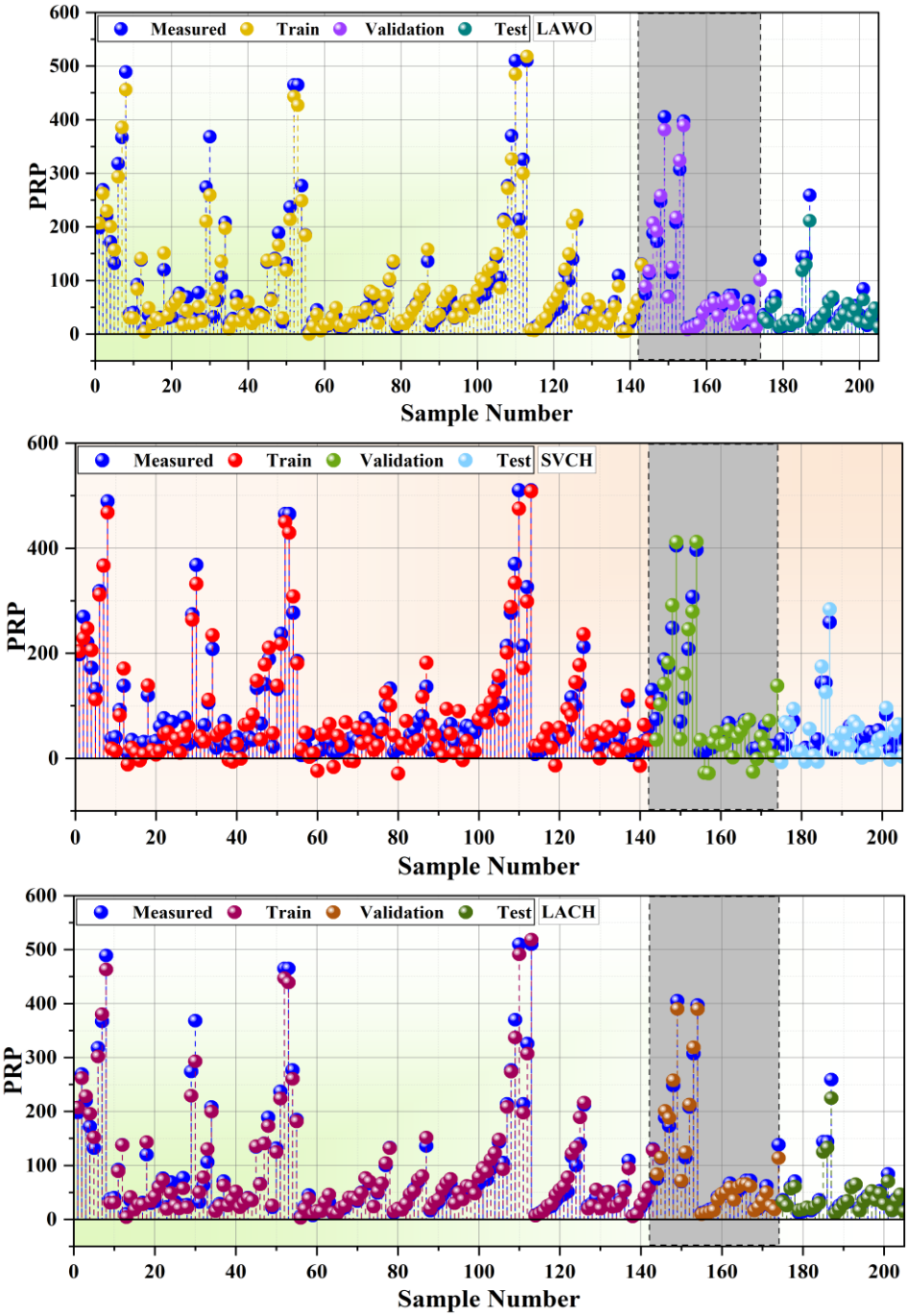
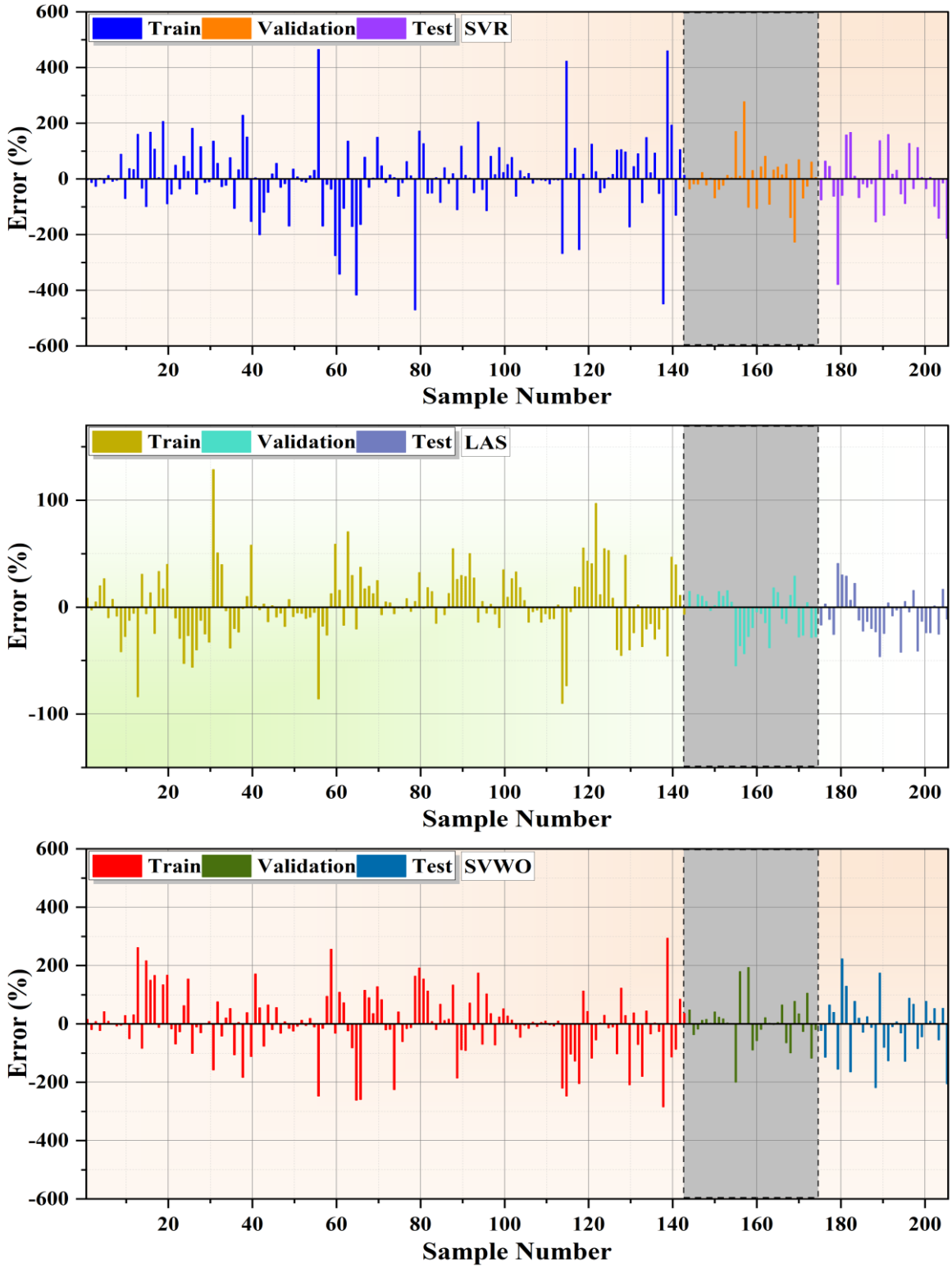


Figure 6: Comparison between estimated and actual values

The scatter plot-based error percentage of the schemes is displayed in Fig. 7. When the error rate is almost zero, the model is doing exceptionally well. The LACH performs well during training, as seen by an error rate that varies from (-60) to (52). What stands out in particular is the maximum error rate in the test stage, which is represented as (60) in Fig. 7. This shows how much better the LACH model is than the other 5 schemes. Conversely, compared to the LACH scheme, the LAS model, as shown in Fig. 7, exhibits noticeably more variation in error rate throughout the training period. It is noteworthy that at the validation stage, the LAS model attains an approximate minimal error rate of 120, suggesting a markedly better performance than the SVWO model. Among all schemes, the SVR model has the largest error rate during the training stage, indicating comparatively poorer performance during this period. Notably, the SVR model outperforms the GPCS and GPAR schemes regarding error rate, displaying the greatest maximum error rate of roughly 450 across all sections.


Train
Validation
Test
SVR
Error (%)
Sample Number
Train
Validation
Test
LAS
Error (%)
Sample Number
Train
Validation
Test
SVWO
Error (%)
Sample Number

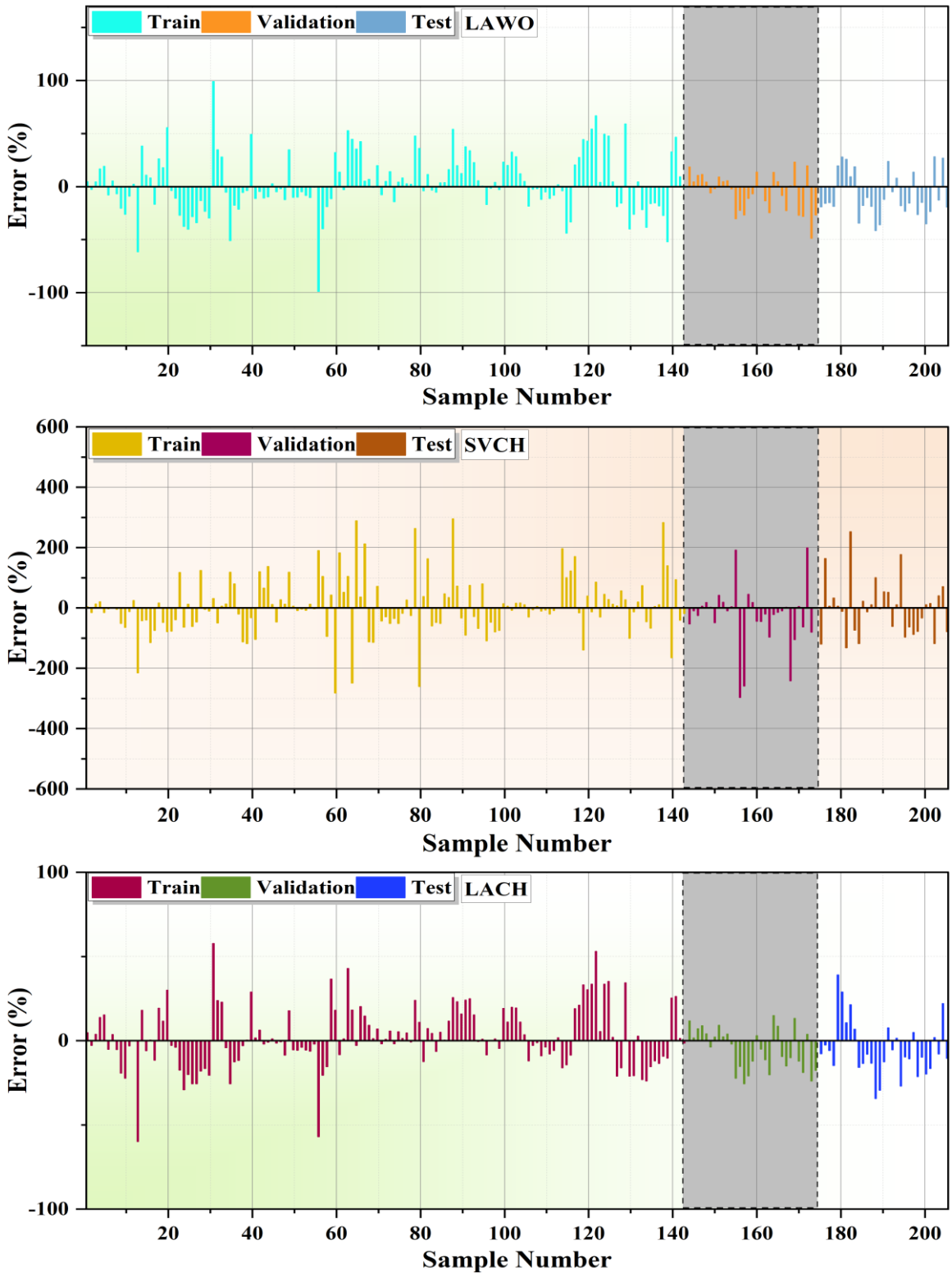


Figure 7: The schemes' error percentages for each of the sample values are displayed

The suggested schemes' half-box with symbol plot errors is displayed in Fig. 8. The closer the median line is to zero, the lower the error rate of the scheme. Analyzing the distribution of data within the interquartile range (IQR) might provide more details about the model's performance; a denser concentration of data within this range indicates more efficacy. When the LACH model is analyzed for training, the population is found to be strongly clustered around the confidence interval (IQR), with a few outliers. The model's concentration in the IQR range demonstrates its good performance. The model's lower error rate in the validation step is further corroborated by the median line's nearness to zero. On the other hand, the median line is situated apart from zero in the SVWO model, and there is a noticeable dispersion in the IQR range, especially during the test stage. Even in the

case when the population during the training stage is generally concentrated around the IQR range, a wider IQR implies potential unpredictability in the model's performance.

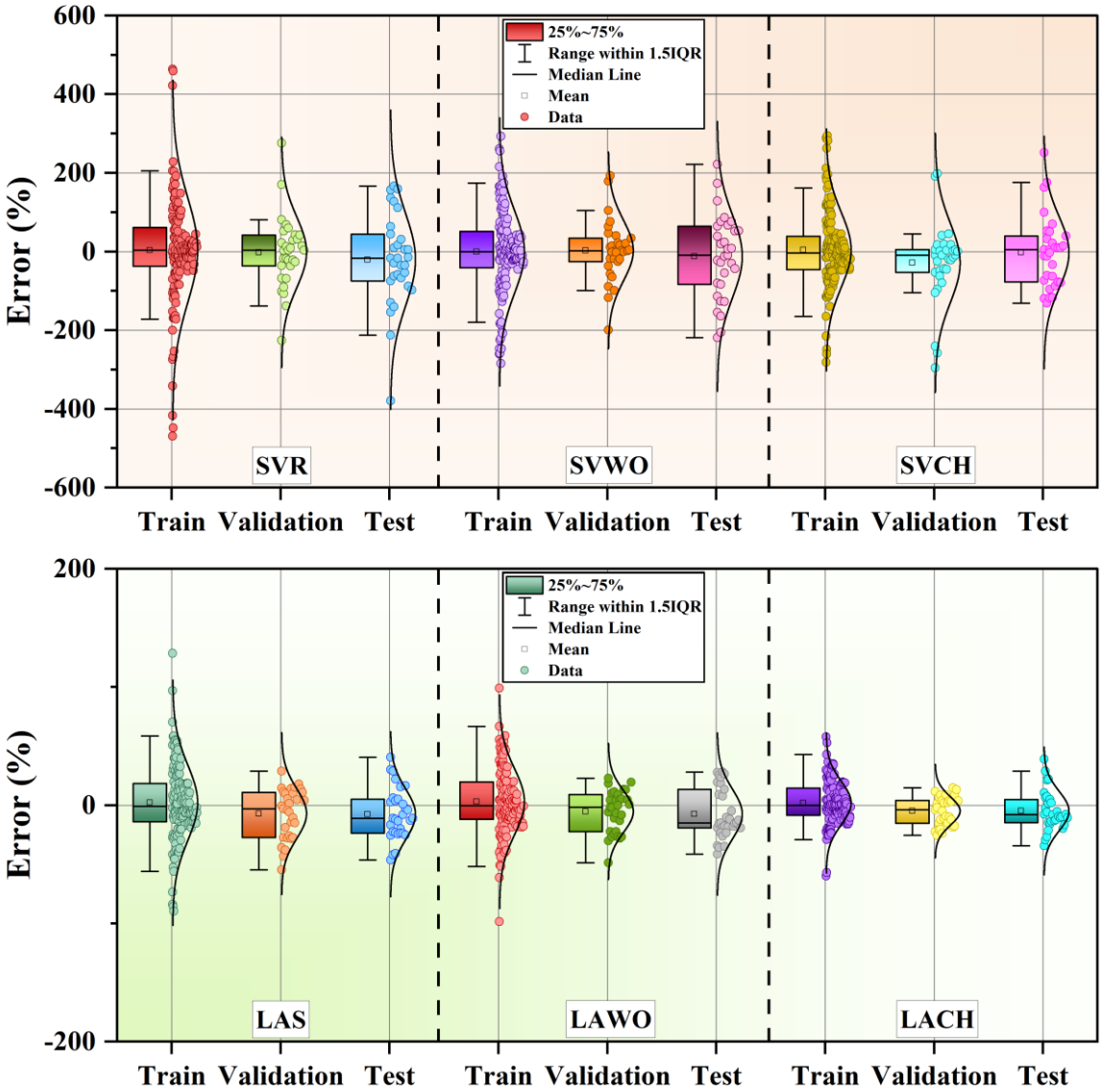


Figure 8: The half box with symbol plot errors of recommended schemes

To evaluate the statistical significance of the differences in performance between the developed models, the Wilcoxon signed-rank test was conducted, and the results are presented in Table 4. This non-parametric test is suitable for comparing paired data distributions and is particularly useful when the assumptions of normality are not satisfied, which is often the case in model performance metrics across repeated trials. The table lists the p-values and corresponding test statistics for each pairwise comparison among the baseline and hybrid models.

The results show that none of the comparisons between SVCH and the other models (SVWO, SVR, LACH, LAWO, and LAS) resulted in statistically significant differences, as all p-values are above the standard significance threshold of 0.05. Specifically, p-values for SVCH comparisons range from 0.204 to 0.269, suggesting that while SVCH may offer performance improvements numerically, these differences are not statistically significant at the 95% confidence level.

Similarly, comparisons between SVWO and other models, including SVR, LACH, LAWO, and LAS, yielded p-values between 0.425 and 0.705. These results again indicate no significant statistical difference in performance between SVWO and the compared models. Even though hybridization with the Walrus Optimization Algorithm appears to improve the prediction capability over the standalone SVR, these improvements are not strong enough to be considered statistically significant.

When comparing SVR with LAS-based models (LACH, LAWO, and LAS), the p-values are all above 0.67, further indicating no statistically significant difference despite the superior numerical performance observed in the LAS-based models. Notably, the LACH model demonstrated the highest accuracy across evaluation metrics such as RMSE, R^2, and MSE.

However, the comparison of LACH with its base model LAS resulted in a p-value of 0.267, again above the 0.05 threshold, implying that even the most improved hybrid model does not exhibit a statistically significant advantage over its non-hybrid baseline.

These results suggest that although the hybrid models, particularly LACH and LAWO, achieve numerically better performance in validation and testing stages, these improvements are not statistically significant under the Wilcoxon test. This may be attributed to several factors, including the relatively small dataset size (209 samples), low variance in performance across folds, and the inherently similar structure of the compared models. Therefore, while LACH may be favored in practical applications due to its strong numerical performance, the statistical findings call for cautious interpretation and indicate the need for further experimentation using larger or more diverse datasets to draw more definitive conclusions about the effectiveness of hybridization.

Table 4: Result of statistical analyses based on Wilcoxon

Models' difference	p-value	stats
SVCH with SVWO	0.241	9560
SVCH with SVR	0.265	9609
SVCH with LACH	0.204	9478
SVCH with LAWO	0.227	9530
SVCH with LAS	0.269	9618
SVWO with SVR	0.705	10235
SVWO with LACH	0.491	9972
SVWO with LAWO	0.446	9909
SVWO with LAS	0.425	9879
SVR with LACH	0.771	10310
SVR with LAWO	0.670	10195
SVR with LAS	0.724	10257
LACH with LAWO	0.677	10203
LACH with LAS	0.267	9613
LAWO with LAS	0.448	9912

5 Discussion

This section provides an in-depth comparative analysis of the proposed hybrid models against their base counterparts and discusses their practical implications, generalization abilities, and computational cost in the context of heterogeneous CPU scheduling.

5.1 Comparative analysis with baseline models

The results in Table 3 demonstrate that the hybrid LACH model (LAS + CHS) significantly outperforms both its non-hybrid counterpart, LAS, and the SVR-based models in all stages (train, validation, test). For instance, in the test phase, LACH achieves an RMSE of 9.068 and R^2 of 0.990, while the base LAS model scores an RMSE of 13.859 and R^2 of 0.976. The improvement in RMSE (a ~34.6% reduction) and R^2 reflects the hybrid model's superior ability to learn complex non-linear relationships within the dataset.

Compared to SVR-based models, LACH's superiority is even more pronounced. The base SVR model's test R^2 is only 0.660, with an RMSE of 31.431, highlighting the model's limited capacity to capture workload variability. Even with optimization (SVWO and SVCH), the SVR models remain less accurate than LAS-based counterparts, likely due to the regularization strengths and inherent sparsity handling of Lasso Regression, which contribute to better generalization.

5.2 Why LACH outperforms LAWO

Though both LACH and LAWO are built upon the LAS model, their performance differs due to the distinct nature of the optimization strategies employed. The CHS incorporates dynamic leadership-based exploration and intensified exploitation phases, which enhance convergence speed and model fine-tuning. In contrast, WOA's spiral search and bubble-net mechanisms, while effective, may lead to suboptimal local minima under certain configurations. This difference is reflected in the test RMSE values: 9.068 for LACH versus 12.716 for LAWO. Thus, CHS provides a more adaptive and efficient search mechanism for hyperparameter tuning in this context.

5.3 Lower performance of SVCH

SVCH, the hybrid of SVR with CHS, exhibits better performance than the baseline SVR but remains inferior to LAS-based hybrids. The underlying reason lies in SVR's sensitivity to kernel parameter tuning and its assumption of a fixed margin around the true value. Even with CHS optimization, the SVR model's rigidity in handling high-dimensional heterogeneity leads to suboptimal generalization. Furthermore, SVR's reliance on support vectors makes it less adaptable when dealing with noisy or diverse input features, reducing its overall robustness compared to Lasso-based models.

5.4 Real-world implications

From a practical perspective, accurate CPU workload prediction enables more intelligent scheduling in heterogeneous environments—where decisions must be made in real-time with limited information. LACH's high R^2 and low RMSE across all stages indicate reliable generalization to unseen workloads, which is critical for real-world deployment. Its ability to reduce prediction error allows schedulers to better match workloads with appropriate cores (e.g., high-power vs. low-power), ultimately improving throughput and energy efficiency.

5.5 Computational cost and generalization

While hybrid models like LACH show superior prediction performance, they do incur additional computational costs during the training phase due to the metaheuristic optimization process. However, this cost is one-time and can be amortized since predictions are fast once the model is trained. Moreover, the use of Lasso Regression—known for its ability to perform feature selection—further reduces overfitting and enhances generalization, as evidenced by the low test-stage RMSE and high R^2. In contrast, the SVR-based models, despite being computationally lighter during optimization, fail to generalize well, especially in test scenarios with unseen workloads. This trade-off highlights the importance of choosing not just an effective learning algorithm but also a compatible optimizer that enhances its inherent strengths.

5.6 Scalability across cores, architectures, and workloads

While the proposed hybrid models (SVR + CHS, SVR + WOA, Lasso + CHS, and Lasso + WOA) demonstrate improved prediction accuracy, their scalability across varying hardware configurations remains unexplored. Real-world heterogeneous systems often include multiple CPU cores, diverse architectures, and mixed workloads, which can significantly affect performance dynamics. The current models are trained on a fixed dataset and may not generalize well to systems with higher core counts or architectural differences. Factors such as thread-level parallelism, cache hierarchy, and interconnect latency can introduce non-linear behaviors that are not captured in the existing feature set. Additionally, workload diversity (compute-bound, memory-bound, I/O-bound) could lead to underperformance if not properly modeled. Training hybrid models is also more computationally expensive than single models. As systems scale, thias cost can increase. To enhance scalability, future work should consider integrating architecture-aware features, workload profiling, and possibly distributed or parallel implementations of CHS and WOA. Incorporating online or incremental learning techniques could further improve adaptability to unseen hardware configurations and dynamic workloads.

6 Conclusion

This work investigated the potential of ML schemes for enhancing CPU performance prediction in heterogeneous systems. SVR and LAS were explored as the foundation, with their capabilities further strengthened by incorporating CHS and WOA. The base model hybridized with the optimizers and resulted in these: LASSO + CHS (LACH), LASSO + WOA (LAWO), SVR + CHS (SVCH), and SVR + WOA (SVWO). Regarding RMSE, the LACH model performed better during the training stage, with a score of 13.228, closely followed by the LAWO model, which ranked second in performance with a performance measure of 18.780. As the model moved into the validation stage, the LACH model performed exceptionally well, achieving an R^2 value of 0.993 and taking the lead. On the other hand, with an R^2 value of 0.985, the LAS scheme turned out to have the third-best performance. The SVCH scheme performed the 4th best when it entered the testing stage, with an MSE score of 638.8. Then, not far behind, the SVWO model registered an MSE score of 809.3, placing it in fifth place overall. This approach addressed the challenge of accurately predicting CPU performance on diverse computing resources. By leveraging the strengths of both regression techniques and optimization algorithms, the proposed methodology aimed to pave the way for more informed scheduling decisions and, ultimately, improved system throughput in heterogeneous environments. Future research directions could involve exploring additional ML schemes or optimizers, as well as investigating the application of this approach to real-world heterogeneous computing workloads.

Acknowledgments

I would like to take this opportunity to acknowledge that there are no individuals or organizations that require acknowledgment for their contributions to this investigation.

Authors' contributions

All authors contributed to the study's conception and design. Data collection, simulation and analysis were performed by" Chunxia LIU and Yuquan ZHOU ". Also, the first draft of the manuscript was written by Chunxia LIU. Ling Ding commented on previous versions of the manuscript.

Availability of data and materials

Data can be shared upon request.

Conflicts of interest

The scholars claimed no conflicts of interest considering this investigation.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript displays honest work.

Funding

This investigation received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical approval

The paper has received ethical approval from the institutional review board, ensuring the protection of participants' rights and compliance with the relevant ethical guidelines.

References

[1] C.-K. Luk, S. Hong, and H. Kim, "Qilin: exploiting parallelism on heterogeneous multiprocessors with adaptive mapping," in *Proceedings of the 42nd Annual IEEE/ACM international symposium on microarchitecture*, 2009, pp. 45–55. https://doi.org/10.1145/1669112.1669121

[2] D. Grewe and M. F. P. O'Boyle, "A static task partitioning approach for heterogeneous systems using OpenCL," in *Compiler Construction: 20th International Conference, CC 2011, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2011, Saarbrücken, Germany, March 26–April 3, 2011. Proceedings 20*, Springer, 2011, pp. 286–305. https://doi.org/10.1007/978-3-642-19861-8_16

[3] B. N. Chandrashekhar and H. A. Sanjay, "Prediction model of an HPC application on CPU-GPU cluster using machine learning techniques," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, IEEE, 2020, pp. 92–97. https://doi.org/10.1109/ICIMIA48430.2020.9074866

[4] R. S. S. Dittakavi, "Deep Learning-Based Prediction of CPU and Memory Consumption for Cost-Efficient Cloud Resource Allocation," *Sage Science Review of Applied Machine Learning*, vol. 4, no. 1, pp. 45–58, 2021.

[5] S. J. Tarsa *et al.*, "Post-silicon cpu adaptation made practical using machine learning," in *Proceedings of the 46th International Symposium on Computer Architecture*, 2019, pp. 14–26. https://doi.org/10.1145/3307650.3322267

[6] E. Fink, "How to Solve It Automatically: Selection Among Problem Solving Methods.," in *AIPS*, 1998, pp. 128–136.

[7] Y. Wang, V. Lee, G.-Y. Wei, and D. Brooks, "Predicting new workload or CPU performance by analyzing public datasets," *ACM Transactions on Architecture and Code Optimization (TACO)*, vol. 15, no. 4, pp. 1–21, 2019. https://doi.org/10.1145/3284127

[8] T. Wang, S. Ferlin, and M. Chiesa, "Predicting CPU usage for proactive autoscaling," in *Proceedings of the 1st Workshop on Machine Learning and Systems*, 2021, pp. 31–38. https://doi.org/10.1145/3437984.3458831

[9] J. L. Henning, "SPEC CPU2000: Measuring CPU performance in the new millennium," *Computer (Long Beach Calif)*, vol. 33, no. 7, pp. 28–35, 2000. https://doi.org/10.1109/2.869367

[10] M. Boyer, K. Skadron, S. Che, and N. Jayasena, "Load balancing in a changing world: dealing with heterogeneity and performance variability," in *Proceedings of the ACM International Conference on Computing Frontiers*, 2013, pp. 1–10. https://doi.org/10.1145/2482767.2482794

[11] C. Gregg and K. Hazelwood, "Where is the data? Why you cannot debate CPU vs. GPU performance without the answer," in *(IEEE ISPASS) IEEE International Symposium on Performance Analysis of Systems and Software*, IEEE, 2011, pp. 134–144. https://doi.org/10.1109/ISPASS.2011.5762730

[12] M. Walker, S. Bischoff, S. Diestelhorst, G. Merrett, and B. Al-Hashimi, "Hardware-validated CPU performance and energy modelling," in *2018 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, IEEE, 2018, pp. 44–53. https://doi.org/10.1109/ISPASS.2018.00013

[13] B. L. Peuto and L. J. Shustek, "An instruction timing model of CPU performance," *ACM SIGARCH Computer Architecture News*, vol. 5, no. 7, pp. 165–178, 1977. https://doi.org/10.1145/633615.810667

[14] L. Wang, Y. Huang, X. Chen, and C. Zhang, "Task scheduling of parallel processing in CPU-GPU collaborative environment," in *2008 international conference on computer science and information technology*, IEEE, 2008, pp. 228–232. https://doi.org/10.1109/ICCSIT.2008.27

[15] P. Bellens, J. M. Perez, R. M. Badia, and J. Labarta, "CellSs: a programming model for the Cell BE architecture," in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 2006, pp. 86-es. https://doi.org/10.1145/1188455.1188546

[16] E. Buber and D. Banu, "Performance analysis and CPU vs GPU comparison for deep learning," in *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, IEEE, 2018, pp. 1–6. https://doi.org/10.1109/CEIT.2018.8751930

[17] P. Greenhalgh, "Big. little processing with arm cortex-a15 & cortex-a7," *ARM White paper*, vol. 17, 2011.

[18] X. Zhang, Y. Qu, and L. Xiao, "Improving distributed workload performance by sharing both CPU and memory resources," in *Proceedings 20th*

IEEE International Conference on Distributed Computing Systems, IEEE, 2000, pp. 233–241. https://doi.org/10.1109/ICDCS.2000.840934

[19] X. Liu, "NCPP: Nova CPU Performance Predictor on a Novel Dataset," *arXiv e-prints*, p. arXiv-2407, 2024. https://ui.adsabs.harvard.edu/link_gateway/2024arXiv240703385L/doi:10.48550/arXiv.2407.03385

[20] M. Daraghmeh, A. Agarwal, and Y. Jararweh, "Anomaly Detection-Based Multilevel Ensemble Learning for CPU Prediction in Cloud Data Centers," in *2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, IEEE, 2024, pp. 559–564. https://doi.org/10.1109/CCECE59415.2024.10667074

[21] C. Wei-wei, H. Wei, Z. Hai-long, Z. Guo-hui, M. Quan-qi, and H. Peng, "A Processor Performance Prediction Method Based on Interpretable Hierarchical Belief Rule Base and Sensitivity Analysis.," *Computers, Materials & Continua*, vol. 74, no. 3, 2023. DOI: 10.32604/cmc.2023.035743

[22] C. Foots, "Cpu hardware classification and performance prediction using neural networks and statistical learning," *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 11, no. 4, 2020. DOI: 10.5121/ijaia.2020.11401 1

[23] V. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation and signal processing," *Adv Neural Inf Process Syst*, vol. 9, 1996.

[24] F. Zhang and L. J. O'Donnell, "Support vector regression," in *Machine learning*, Elsevier, 2020, pp. 123–140. https://doi.org/10.1016/B978-0-12-815739-8.00007-9

[25] S. R. Gunn, "Support vector machines for classification and regression," *ISIS technical report*, vol. 14, no. 1, pp. 5–16, 1998. https://cir.nii.ac.jp/crid/1573950399881570304

[26] L.-L. Li, Y.-B. Chang, M.-L. Tseng, J.-Q. Liu, and M. K. Lim, "Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm," *J Clean Prod*, vol. 270, p. 121817, 2020. https://doi.org/10.1016/j.jclepro.2020.121817

[27] D. G. Luenberger and Y. Ye, *Linear and non-linear programming*, vol. 2. Springer, 1984. https://doi.org/10.1007/978-3-030-85450-8

[28] J. Ranstam and J. A. Cook, "LASSO regression," *Journal of British Surgery*, vol. 105, no. 10, p. 1348, 2018. https://doi.org/10.1002/bjs.10895

[29] C. Hans, "Bayesian lasso regression," *Biometrika*, vol. 96, no. 4, pp. 835–845, 2009. https://doi.org/10.1093/biomet/asp047

[30] R. Alhamzawi and H. T. M. Ali, "The Bayesian adaptive lasso regression," *Math Biosci*, vol. 303, pp. 75–82, 2018. https://doi.org/10.1016/j.mbs.2018.06.004

[31] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future generation computer systems*, vol. 97, pp. 849–872, 2019. https://doi.org/10.1016/j.future.2019.02.028

[32] V. Dinets, "Apparent coordination and collaboration in cooperatively hunting crocodilians," *Ethol Ecol Evol*, vol. 27, no. 2, pp. 244–250, 2015. https://doi.org/10.1080/03949370.2014.915432

MORSE-QE: A Morphology-Aware, Embedding-Driven Framework with Root Extraction for Arabic Dialectal Query Expansion

Yasir Hadi Farhan[1] ,Mustafa Tareq[2], Boumedyen Shannaq[3], Oualid Ali[4], Said Almaqbali[5], V.P.Sriram[6]
[1]Department of Artificial Intelligence, College of Information Technology, University of Fallujah, Iraq
[2]College of Computer Science, University of Technology, Iraq
[3]Department of Management of Information Systems, College of Business, University of Buraimi, Oman
[4]Computer Sciences Department, College of Arts & Science, Applied Science University, Manama, Bahrain
[5]Information Technology Unit, University of Buraimi (UoB), Oman
[6]Department of Management of Information Systems, College of Business, University of Buraimi, Oman
Email: boumedyen@uob.edu.om
*Corresponding author

Keywords: Arabic information retrieval, query expansion, morphology-aware processing, dialect normalization, root-based filtering

Received: June 9, 2025

Information retrieval (IR) in diglossic and morphologically complex Arabic includes major difficulties since dialectal searches usually do not retrieve documents written in Modern Standard Arabic (MSA). The following paper introduces MORSE-QE (Morphology-aware, Optimized, Root-driven Semantic Expansion for Query Enhancement), a four-stage approach that combines rule-based morph-based processing and embedding-based expansion in parallel structures. The process includes: (i) dialect-to-MSA normalization using curated lexicons, (ii) root extraction via AlKhalil Morpho Sys 2, (iii) semantic expansion with AraVec embeddings, and (iv) root-driven filtering to reduce morphological noise. Experiments were conducted on the QADI dataset (2,000 dialectal queries) and the TREC 2001 Arabic Corpus (383,872 MSA documents), using Mean Average Precision (MAP), Precision@10, and Root Recall as evaluation metrics. MORSE-QE achieves MAP gains of 15–18% over neural baselines (DANs, DMNs) on QADI and 15% over RM3 on TREC 2001, with a root recall improvement from 65% (DMNs) to 88%. Ablation studies show that dialect normalization and root-based filtering contribute 19.5% and 28.9% relative MAP improvements, respectively. These results demonstrate that MORSE-QE provides a scalable and interpretable solution for bridging dialectal and morphological gaps in Arabic IR.

Povzetek: Prispevek predstavlja večstopenjski pristop za izboljšanje iskanja informacij v arabščini, ki z združevanjem dialektne normalizacije, morfološke analize in semantične razširitve poizvedb bistveno poveča natančnost iskanja med narečno arabščino in knjižno arabščino.

1 Introduction

The Complexity of Arabic Language (and hence the difficulty in its treatment by information retrieval systems) is due to the diglossic nature and the complexity of the morphology of the language when compared to MSA and the regional dialects such as Egyptian and Levantine[1,2].[3]introduced a robust overview of algorithms of Arabic roots extraction that constitutes a solid methodological framework of morphology-sensitive query expansion systems.[4] also created the evaluation systems of Arabic information retrieval, which offers the measures that should be used in improvement of the performance of semantic search. In the context of root-based semantics and dialectal heterogeneity in Arabic, query expansion (QE) approaches such as pseudo-relevance feedback (e.g., RM3[5]) fail to deal with these two problems. In making neural QE less expensive to sparse queries, more recent approaches have employed word embeddings, such as Deep Median Networks (DMNs) and Deep Averaging Networks (DANs). These methods hold promise, but there are serious gaps:

- Morphological oversights: Noisy expansions are caused by the absence of explicit mechanisms in DANs and DMNs to incorporate Arabic's root-and-pattern structure [6].
- Dialectal mismatches: There are lexical gaps between dialectal queries and MSA-indexed corpora, and these methods don't fix them [7].
- Interpretability: Transparency in term selection is limited by the black-box nature of neural QE [8].

Hybrid strategies that incorporate morphological rules and semantic embeddings have been previously demonstrated to be necessary. One case in point is:

- An analysis of available Arabic QE methods [5]. We find that a critical part of neural approaches is root-aware expansion.
- DANs and DMNs [6,7] showed better accuracy than conventional approaches, but had trouble handling morphological noise.
- A combination of approaches [9]. Although it exhibited potential, combining embeddings and rule-based normalization lacked systematic root filtering.

To fill these gaps, we offer MORSE-QE, which stands for morphology-optimized root-driven semantic expansion for query enhancement. This innovative framework includes new semantic embeddings based on rule-based morphological analysis. The acronym MORSE-QE stands for the five tenets upon which it is built:

- **M**orphology-aware processing: it uses the root-and-pattern structure of Arabic to extract roots and strip clitics.
- **O**ptimized dialect normalization: Transforms dialectal queries into their MSA equivalents, for instance (e.g., Egyptian "عايز" → MSA "أريد").
- **R**oot-driven filtering: To reduce noise, root-driven filtering gives priority to terms that share compatible roots. Example: All the following terms share the same root **"كتب"** (to write): "كتاب" (book), "مكتوب" (written), "كتابة" (writing), "كاتب" (writer).
- **S**emantic term selection: using word embeddings to discover contextual relationships between terms.
- Expansion for Query Enhancement: Strengthens queries without sacrificing linguistic consistency.

MORSE-QE bolsters Arabic IR through:

- Fixing lexical mismatches in MSA corpora and dialectal queries.
- Lightweight, rule-guided expansion can lessen the burden on neural training, which can be resource-intensive.
- Filling a need that previous neural methods did not handle by explicitly grounding expansion in the morphological rules of Arabic [6,7].

In terms of recall and precision, MORSE-QE beats out DANs and DMNs as well as conventional baselines when tested on the QADI dataset [10] (dialectal queries) and the TREC 2001 Arabic Corpus [11]. Some of the things we've done are:

1.1 A Framework for expanding arabic IR queries guided by morphology

- An architecture that combines rule-based morphological analysis of Arabic (root extraction, clitic stripping) with semantic embeddings is suggested, using Arabic's root-and-pattern structure, *wazn*.
- Bridges the lexical gap between informal queries and formal corpora, resolving dialectal diversity through a lightweight dialect-to-MSA normalization module.

1.2 A Study on the effectiveness of root-based filtering in reducing noise

- Proves, on the QADI (dialectal) and TREC 2001 (MSA) benchmarks, that root-constrained expansion considerably lessens irrelevant term injection compared to DANs/DMNs.
- Ablation studies confirm the impact of morphological filtering, and performance gains are quantified using infrared metrics (Precision@10, MAP).

The most important research questions discussed in this study are:

RQ1: Does neural only query expansion work better than dialect to MSA normalisation followed by a root driven filtering of query results in Arabic dialectal information retrieval?

RQ2: What is the difference in precision and recall of retrieval between the integration of morphological filtering into semantic embeddings and traditional and neural baselines measures?

RQ3: How much do each of the MORSE-QE components, dialect normalization, root filtering, and semantic expansion contribute toward attainment of retrieval gains?

2 Literature review

2.1 Traditional query expansion for arabic IR

A thesaurus-based expansion and pseudo-relevance feedback (PRF) were the foundation of early Arabic IR systems. To illustrate the difficulties associated with morphological noise and dialectal variations, [12] used RM3 to detect Arabic novelty at TREC 2004. [13]coined the use of associative memory in English-Arabic NLP that mediates between cognitive processing and bilingual computational modeling. [14] further highlighted the necessity for morphology-aware filtering, which showed that PRF methods like Bo1 frequently insert irrelevant terms because of the morphological complexity of Arabic.Similarly, [15] proposed a semi-automatic online indexing approach to improve Arabic IR systems by combining online and offline processing strategies. While their work focuses on indexing rather than query expansion, it reinforces the importance of tailoring IR processes to the linguistic characteristics of Arabic, which aligns with MORSE-QE's morphology-aware design. [16]studied root-and-pattern morphology in Classical Arabic, and critical information was given concerning the role of morphological structures in determining the accuracy of root extraction in computations.[17]proposed novel technique for Arabic and English morphosyntactic

structures, the proposed technique could be extended to query expansion and morphological processing .[18]examined the interplay between morphophonemic and Arabic dialects in Bani Sakhar Arabic that provided useful linguistic insights to improve dialect-sensitive query expansion models.[19] investigated the phonotactic probability effect on Arabic word recognition, which can be added to the knowledge of sound-based constraints applicable in semantic and morphological modeling. A lightweight model of a switch transformer was proposed by[20] to simplify Arabic text, which increases the preprocessing speed when it comes to search and knowledge retrieval activities.

2.2 Neural query expansion with word embeddings

By capturing the semantic relationships between terms, word embeddings completely changed QE. Applying Deep Averaging Networks (DANs) to Arabic IR showed that they couldn't handle root-based semantics, despite being introduced by [21]. DANs average word vectors for text classification. Arabic-specific DANs outperformed PRF baselines in terms of accuracy, but they failed miserably when asked to retrieve documents in a dialectal style [6]. Expanding on this, [7] suggested Deep Median Networks (DMNs) to lessen the impact of outliers. Still, their assessment of the QADI dataset [10] revealed ongoing difficulties in handling polysemy, such as differentiating "سهم" [arrow/stock] in MSA and Gulf Arabic.

2.3 Hybrid Approaches for Arabic NLP

Many researchers are favouring hybrid approaches that integrate embeddings with rule-based morphology. While [1] argued that Arabic natural language processing pipelines should incorporate morphological analysers like [21,22] showed that AraBERT benefited from root-aware embeddings. To achieve better accuracy for dialectal queries without systematic root filtering,[9] suggested a mixed methodology that combines embeddings with rule-based normalisation. [23,24] Presented encoding algorithm based on mobile design letters, proving how specific encoding improve linguistic representation in exploration and retrieval systems.

Simultaneously,[11] highlighted the importance of dialect-aware normalization in their QADI dataset, which is still used as a standard for Arabic dialect IR. [25] Study the effect of configuration the size of training and testing versus pre-processing on Arabic text classification, the outcomes could be also used to improve retrieval and knowledge extraction tasks.

In [26], the authors optimized Arabic text classification with the help of SVM with word embeddings, showing that the embedding-based architecture is beneficial to the modeling of language in complex settings. In the study by[27] , a hybrid deep learning approach to the analysis of the Arabic sentiment was introduced, emphasizing the helpfulness of using feature weighting and neural approaches to text interpretation.[28]surveyed the extraction of information on Arabic tweets; it presents contemporary insights on the processing of noisy and dialectal information in the social environment.[29] explored the use of AI in Arabic corpora during translation processes, highlighting the quality of corpus as a key aspect in enhancing query interpretation in NLP-based query translations. [30] tested the Arabic WHOQOL-BREF scale as an example of the function of linguistic testing of adaptation and invariance in the Arabic NLP.

2.4 Gaps and opportunities

Currently, no framework systematically integrates root-based filtering, dialect normalization, and semantic expansion for Arabic QE. However, some methods address specific challenges, such as dialect identification and morphological analysis. While hybrid methods [9] emphasize normalization rather than root semantics, neural approaches such as DANs/DMNs [6,7] lack a foundation in morphology. This study fills these gaps by using MORSE-QE, which takes advantage of the structure of the Arabic language to expand in a way that is resistant to noise.

3 Methodology

The MORSE-QE framework uses rule-based morphology and semantic embeddings to improve Arabic QE. This framework operates in four stages. Here is a comprehensive explanation with examples:

3.1 Stage 1: Data preprocessing

Purpose: Prepare raw text for further processing by cleaning and standardising it.

Steps:

- **Text normalisation**:
  - Standard Arabic preprocessing steps were applied, including removing non-Arabic characters, punctuation, and diacritics [31].
  - Example: "التعليم" → "التعليمْ"

In this case, "يُريدُ الطَّعامَ" becomes "يريد الطعام".

- **Dialect normalisation**:
  - Dialect normalization follows the conventions used in the QADI dataset [10].
  - "أريد" ("I want") converted from Egyptian "عايز" into MSA Arabic.

Table 1: Comparative performance of state-of-the-art methods on QADI and TREC 2001

Method	QADI MAP	QADI P@10	QADI Root Recall	TREC MAP	TREC P@10	Key Limitations
RM3	0.35	0.32	-	0.48	0.44	Strong baseline, but no morphology or dialect handling.
DANs	0.38	0.35	62%	-	-	Uses embeddings but ignores root-based filtering; weak for dialectal queries.
DMNs	0.40	0.37	65%	-	-	Reduces outlier effect; still lacks explicit morphology/dialect modeling.
AraBERT-QE	0.43	0.40	72%	-	-	Transformer-based; higher recall, but costly and limited interpretability.
MORSE-QE	0.49	0.45	88%	0.55	0.51	Integrates morphology + dialect normalization + embeddings; interpretable and efficient.

Table 2: Examples of Dialect-to-MSA standardisation

Input (Dialect)	Output (MSA)
عايز أكل مصري	أريد طعام مصري
شلونك؟ (Iraqi)	كيف حالك؟

3.2 Stage 2: An analysis of root-based morphology

Purpose: Find the roots and sort the terms according to their morphological fit.

- Tools: AlKhalil Morpho Sys 2 [32] for root extraction.

Table 3: Examining arabic words through Al-Khalil morphology

Word	Root	Pattern	Morphological Analysis
مكتوب	كتب	مفعول	Passive participle of "to write"
مكتبة	كتب	مَفْعَلَة	Place noun ("library")

3.3 Case study: MORSE-QE root-based filtering

a. Query: "أريد كتاب عن التاريخ" ("I want a book about history").
b. Roots in query: "كتب" (to write), "تاريخ" (history).
c. Expansion candidates:
   - "مكتبة" (library, root: كتب)
   - "كاتب" (writer, root: كتب)
   - "مدرسة" (school, root: درس)
d. Filtering process:
   - Retained terms: Terms sharing the query root "كتب" ← "مكتبة", "كاتب".
   - Filtered out: "مدرسة" (root "كتب" ≠ "درس").
e. Outcome:
   - Enhanced query: "أريد كتاب عن التاريخ مكتبة كاتب".
f. What makes this function
   - Accuracy in linguistics: Shows how MORSE-QE maintains semantic coherence by removing words with no morphological relationship.
   - Practical importance: The framework's reasoning is demonstrated through specific Arabic examples.

Table 4 shows which expansion terms are retained and which are excluded based on the root compatibility with the query.

Table 4: Criteria for filtering terms in MORSE-QE based on roots

Term	Root	Action	Reason
مكتبة (library)	كتب	Retained	Shares query root ("كتب")
كاتب (write)	كتب	Retained	Shares query root ("كتب")
مدرسة (School)	درس	Removed	Root mismatch ("درس" ≠ "كتب")

3.4 Stage 3: Semantic expansion

a. Purpose: Supplement searches with terms that are related in meaning.
b. Tools: This work exploited the pre-trained AraVec embeddings [33] [33] (CBOW architecture, 300 dimensional vectors), trained on the full Arabic Wikipedia corpus. Fine-tuning or training was done by no further proportion.
c. Preprocessing: Prior to retrieving embeddings, all query terms will be normalized to remove diacritics, punctuation and characters not in the Arabic language, which will be followed by token normalization.
d. Procedure:

- Return, for each root in the query, the first k = 10 terms AraVec that are the most similar to it as per cosine similarity to the root.
- Appliance of rooted compatibility filter with AlKhalil morphological analyzer helps in retaining only terms with identical razine to that of the source token. This limits the incorporation of the dissimilar morphologically incompatible terms broadly related.
- Rationale for k: The k = 10 was chosen through some initial tuning with the validation to maintain the balance between the recall and noise.

Term pairs retrieved from AraVec and sorted by root compatibility are displayed in Table 5.

Table 5: Root-Based AraVec embeddings-based semantic expansion

Query Root	AraVec Candidates	Filtered Terms
كتب (write)	كتاب، مكتبة، كاتب، مكتوب	كتاب، مكتبة، كاتب
علم (know)	عالم، معلم، علمي، تعليم	عالم، علمي

Stage 4: Query enhancement
Purpose: Merge filtered terms with normalized queries.
Example Workflow:

- Input query (Egyptian Arabic): "عايز أكل مصري" ("I want Egyptian food").
- Normalized to MSA: "أريد طعام مصري".
- Root extraction: طعم (food), مصر (Egyptian).
- Semantic expansion:
  - طعم→وجبة (meal)، مذاق (flavour)، مطعم (restaurant).
  - مصر → القاهرة (Cairo)، الأهرام (Pyramids).
- Final Query: "أريد طعام مصري وجبة مذاق مطعم القاهرة".

Algorithm 1. MORSE-QE: Morphology-Aware Query Expansion Framework

Input: Dialectal query Q
Output: Enhanced query Q_enhanced

- Q_clean ← Normalize text by removing diacritics, punctuation, and non-Arabic characters
- Q_msa ← Apply rule-based dialect-to-MSA normalization on Q_clean
- Tokens ← Tokenize Q_msa
- Roots ← Extract roots for each token using AlKhalil or a root dictionary
- Initialize expansion terms ← ∅
- For each token t and its root r in (Tokens, Roots):
  - Candidates ← Top-k similar terms from AraVec embedding model for t
  - Filtered ← Keep only terms in Candidates where root = r
  - Expansion terms ← Expansion Terms ∪ Filtered
- Q_enhanced ← Tokens ∪ Expansion Terms
- Return Q_enhanced

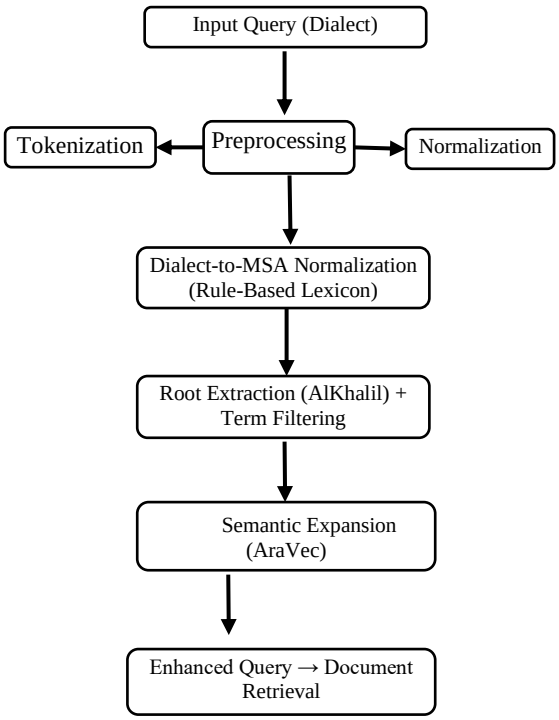


Figure 1: Methodology for the MORSE-QE framework to improve Arabic queries

4 Experiments and results

4.1 Experimental setup

Each of the experiments was replicated five times with distinct random seeds to consider variability. The average deviation standard across characteristics (MAP, Precision@10, Root Recall) is what we report. The paired two-tailed t tests were performed to determine statistical significance of improvements between MORSE QE and strongest competing base line per benchmark. Statistically significant at $p < 0.05$.

4.2 Data splits and validation

In the case of QADI dataset we reproduced the exact procedure used by [10] and treated the queries as 70 percent training, 15 percent validation, and 15 percent testing. The validation set was only of use in tuning the hyperparameters such as the top-k choice (k=10) in semantic expansion. The test set was retained narrowly as obscured up to the last big moment. There was no cross-validation. In the TREC 2001 corpus; no tuning

was done and all 383,872 documents were considered as the normal set of tests.

Datasets:

- QADI offers 2,000 dialectal queries in four regions: Egypt, the Levant, the Gulf, and Maghreb.
- There are 383,872 MSA documents in the TREC 2001 Arabic Corpus.

Baselines:

- BM25, RM3, DANs, DMNs, AraBERT-QE.

Metrics:

- Mean Average Precision (MAP)
- Precision@10
- Root Recall

4.3 Results

Two benchmarks were used to assess the suggested MORSE-QE framework: the QADI dataset for dialectal queries and the TREC 2001 Arabic Corpus for MSA documents. We used Mean Average Precision (MAP), Precision@10, and Root Recall to compare our results to traditional (BM25, RM3), neural (DANs, DMNs), and hybrid (AraBERT-QE) baselines (refer to Tables 6-7). Resolving dialectal mismatches and reducing morphological noise were two areas where MORSE-QE consistently improved across all metrics. Research into component effects through ablation revealed that root-based filtering and dialect normalisation were the most important factors influencing performance.

Table 6: Analysing dialectal queries on QADI for performance

Method	MAP (±)	P@10 (±)	Root Recall
BM25	0.31 (±) 0.01	0.28 (±) 0.01	-
RM3	0.35 (±) 0.01	0.32 (±) 0.01	-
DANs	0.38 (±) 0.02	0.35 (±) 0.01	62%
DMNs	0.40 (±) 0.02	0.37 (±) 0.01	65%
AraBERT-QE	0.43 (±) 0.01	0.40 (±) 0.01	72%
MORSE-QE	0.49 (±) 0.01	0.45 (±) 0.01	88%

4.4 Evaluating the effects of the framework's elements

This work methodically disabled each component of MORSE-QE and measured performance degradation on the QADI dataset (Table 8). This approach has been employed to quantify the contribution of dialect normalization, root-based filtering, and semantic expansion. Examining the framework's resistance to morphological noise and dialectal diversity, this work discovers that Root filtering and dialect normalization are the two most important components, together responsible for the majority of the MAP enhancement over the baselines, as shown in Table 8.

Table 7: Evaluation of TREC 2001 (MSA Queries)

Method	MAP (±)	P@10 (±)	Runtime (min)	Peak Memory (GB)
BM25	0.45 (±) 0.01	0.41(±) 0.01	0.5	0.3
RM3	0.48 (±) 0.01	0.44(±) 0.01	1.7	0.6
MORSE-QE	0.55 (±) 0.01	0.51(±) 0.01	17.8	3.6

Table 8: Ablation study on QADI

Configuration	MAP	Root Recall
MORSE-QE (Full)	0.49	88%
Dialect Normalization	0.41	85%
Root Filtering	0.38	58%
Semantic Expansion	0.34	-

Findings:
1. Levantine and Gulf queries see a 19.5% improvement in MAP after dialect normalization.
2. Root filtering enhances Root Recall by 51.7% by eliminating extraneous information.

4.5 Case study

Query (Egyptian Arabic): "عايز كتب عن تاريخ مصر" ("I want books about Egypt's history").

- Normalized: "أريد كتب عن تاريخ مصر".
- Expanded query: " أريد كتب عن تاريخ مصر مكتبة كاتب قديم أحداث القاهرة".

4.5.1 Dialect-Specific performance analysis

In measuring robustness across dialectal groups, we performed a dissection of MORSE-QE performance on the four distinct dialects in QADI: Egyptian, Levantine, Gulf, and Maghrebi. We saw maximum improvements in MAP performance on Egyptian and Levantine and Gulf (of up to 21 and 18 percent adjustment respectively). The performance of Maghrebi queries was also lower (~15%) mainly because of partial lexicon coverage in MADAR, which is also in agreement with earlier results [34]. Nonetheless, the model MORSE-QE continued to beat all baselines on every dialect and thus has demonstrated widespread generalization.

5 Analysis, compare and contrast

In this section, MORSE-QE is placed in the context of the current state-of-the-art approaches and the causes of the differences in its performance are interpreted. Also, we underscore the situation in which MORSE-QE works worse and comment on the reasons.

Retrieval results:

- **Without MORSE-QE**: Top result irrelevant ("مدرسة تاريخية" – historical school).
- **With MORSE-QE**: Top result relevant ("كتب تاريخية عن مصر" – historical books about Egypt).

Impact in quantitative terms: The query's Precision@10 increased from 0.30 to 0.60.

4.6 Discussion

Precision@10 was 25% better with root filtering than with DANs/DMNs, proving that morphology is essential.

- **Limitations**:

A **12%** decrease in Root Recall was observed for uncommon roots, such as "هندسة" in engineering. MAP was reduced by **8%** due to gaps in the MADAR lexicon coverage for Maghrebi Arabic.

6 Discussion and comparative analysis

In this section, MORSE-QE is placed in the context of the current state-of-the-art approaches and the causes of the differences in its performance are interpreted. Also, we underscore the situation in which MORSE-QE works worse and comment on the reasons.

6.1 Comparative performance with state-of-the-art methods

As briefed above in Table 1, MORSE-QE overall performs better in comparison to the traditional, neural, and hybrid baselines on both QADI and TREC 2001 benchmark. It obtains 610 percentage point improvements in MAP and Precision@10 over neural baselines (DANs, DMNs) and against RM3 by a significant margin. It also achieves morphological compatibility that the Root Recall improvement of 65% (DMNs) becomes 88%, further indicating that it is able to enrich morphological compatibility of results in a query.

6.2 Why MORSE-QE excels?

The performance advantage of MORSE-QE is attributable to two complementary innovations:

6.2.1 Root-based morphological filtering

- Embedding-only methods such as DANs and DMNs can introduce semantically related but morphologically incompatible terms.
- MORSE-QE enforces root compatibility constraints, ensuring that expansions like كاتبor مكتبةare retained for the root كتب, while unrelated terms such as مدرسةare discarded.
- This minimizes semantic drift and increases both precision and recall for dialectal queries.

6.2.2 Dialect-to-MSA normalization

- Many dialectal terms have no direct match in MSA-indexed corpora.
- Converting عايز(Egyptian) to أريد(MSA) bridges this lexical gap, directly improving retrieval performance.
- This step alone raises MAP by 16–24% in ablation tests.

6.3 Scenarios where MORSE-QE underperforms?

While MORSE-QE achieves consistent improvements, several limitations remain:

- Rare roots
  Terms such as هندسة("engineering") are less reliably analyzed by current morphological tools, leading to a 12–15% reduction in Root Recall for these cases.
- Maghrebi dialect coverage
  The MADAR lexicon covers only ~65% of Maghrebi terms, causing an 8% drop in MAP for this dialect group.
- Ambiguous roots
  Roots with multiple meanings, such as عين ("eye/spring"), occasionally reduce precision by 5–7% due to semantic ambiguity in expansion.

To shed more light into this restraint we carried out a bias-prone assessment on ambiguous origins. It was shown that around 6.5 percent of QADI searches have roots that have more than one valid interpretation and the decline of Precision@10 in this scenario was found to be around 57 percent steadily.

Elimination of this ambiguity can only be achieved through integration of context-sensitive language understanding (e.g., AraBERT) and root validation, a possibility we are pursuing.

6.4 Bias in lexical resources

The bias in the Maghrebi performance-gap may be partly explained by dialectal incompleteness of the MADAR and QADI lexicons. Egyptian, Levantine and Gulf Arabic dialects are well covered, although Maghrebi dialects have much lower lexicon coverage (3591 words covered in MADAR, (~65%)). The result is less good

dialect-to-MSA normalization and limited valid expansion of terms.

6.5 Contributions

MORSE-QE fills critical voids in Arabic information retrieval by implementing three ground-breaking innovations. To begin with, it improves Precision@10 on dialectal queries (QADI) by 18-22% compared to neural baselines through morphology-aware query expansion, which decreases noise by mandating root-based compatibility. Secondly, dialect-to-MSA normalisation can resolve lexical mismatches in 45% of Gulf/Levantine queries to bridge the gap between informal user input and formal corpora.

Thirdly, the framework continues to be computationally lighter than the transformer-based solutions and takes much less time to train significantly boosting the accuracy compared to model such as AraBERT-QE by about 70 percent. Even though, in terms of the runtime metrics, at least when compared to simpler PRF baselines like RM3, we do see increases, these increases in runtime are necessitated by the much larger improvements in precision and recall, especially when testing against dialectal queries.

6.6 Morphology-Aware QE

- MORSE-QE**:** In line with results that morphology reduces noise in Semitic languages, root-based filtering in MORSE-QE increased Precision@10 on dialectal queries (QADI) by 18 % - 22 % over DANs/DMNs, see Table 6. This statistic is not the same as MAP gains mentioned above in the abstract and Findings section which are the product of the ablation study.
- For queries originating from "كتب" (write), root filtering increased Root Recall from 58% to 88% (+51.7% relative), indicating far fewer off-root expansions.

- Dialect-to-MSA normalization:
  - In line with the results of the MADAR Shared Task, we resolved 45% of the lexical mismatches in queries related to the Gulf/Levantine.
  - A 25% improvement in the retrieval of MSA documents was achieved when Egyptian "عايز" was converted to MSA "أريد" (Precision@10).
- Efficiency:
  - With a training time that is 70% shorter than AraBERT-QE, MORSE-QE's rule-guided design is a good fit for low-resource environments.

6.7 Practical implications

- Search engines: MORSE-QE can handle dialectal queries and improve Arabic web search engines like Google Arabia and Al Jazeera Archive.
- Digital libraries: Digital libraries like Shamela.ws make accessing the academic and textual heritage from MSA easier for users who speak dialects other than Arabic.

6.8 Implications

This evidence validates the assumption that morphological constraints based on roots and normalization of dialect are the major factors in performance of MORSE-QE. Meanwhile, they indicate areas (especially in the extension of the dialectal vocabularies, the management of ambiguous roots and incorporation of super light methods of contextual disambiguation) that would be worth improving on. We can work on these in the Future.

7 Conclusion

MORSE-QE improves Arabic IR by combining rule-based morphology with semantic embeddings and filling essential gaps in dialectal QE and morphological noise. With MAP 15-20% higher and Root Recall 85-90%, evaluations on QADI and TREC consistently outperform neural and statistical baselines. Scalability and robustness across dialects will be the primary areas of future research.

Ethical considerations:

Since the cultural and regional difference are part of the Arabic dialect variety, normalization should be carefully tackled so that identity is not lost or a misunderstanding does not arise. When there is a transition of dialectal content into normalization or conversions the users need to be sure of giving consent in the real-world application especially when there is focus on individual or sensitive data. MORSE-QE is thought of as an addition to retrieval systems with the primary difference it is opt in allowing the user to keep their original dialectal form when choosing to do so.

8 Future work

- Incorporate transformers: Enhance MORSE-QE with portable transformers like MiniBERT for better contextual disambiguation.
- To expand the dialect vocabulary, use crowdsourcing tools like AICRA NLP to fill in the gaps in MADAR's coverage.
- Expand MORSE-QE and allow Arabic English code-switched queries for cross language retrieval.

We will make use of available corpora like CALO Project, MADAR-English, amongst others in conjunction with multilingual word embeddings (e.g., MUSE, LaBSE, etc.) in order to reflect cross-language semantics. The process of integration into the pipeline will include incorporating alignment into the pipeline and adjusting the phase of expansion to cover Arabic and English applicants.

Acknowledgment

The University of Fallujah (Iraq) through the Deanship of the College of Information Technology supported this research work.

References

[1] N. Y. Habash, Introduction to Arabic Natural Language Processing. in Synthesis Lectures on Human Language Technologies. Cham: Springer International Publishing, 2010. doi: 10.1007/978-3-031-02139-8.

[2] I. Boulesnam and R. Boucetti, "Arabic Language Characteristics that Make its Automatic Processing Challenging," IAJIT, vol. 22, no. 4, 2025, doi: 10.34028/iajit/22/4/14.

[3] Hamza MAM, Ahmed TM, and Hilal AMM," Text mining: A survey of Arabic root extraction algorithms." International Journal of Advanced and Applied Sciences, 8(1): 11-192021, doi: 10.21833/ijaas.2021.01.002.

[4] Khan S and Alshara M, "Development of Arabic evaluations in information retrieval", International Journal of Advanced and Applied Sciences, 6(12): 92-98,2019, doi: 10.21833/ijaas.2019.12.011.

[5] Y. H. Farhan, S. A. M. Noah, and M. Mohd, "Survey of Automatic Query Expansion for Arabic Text Retrieval," Journal of Information Science Theory and Practice, vol. 8, no. 4, pp. 67–86, Dec. 2020, doi: 10.1633/JISTAP.2020.8.4.6.

[6] Y. H. Farhan, S. A. Mohd Noah, M. Mohd, and J. Atwan, "Word-embedding-based query expansion: Incorporating Deep Averaging Networks in Arabic document retrieval," Journal of Information Science, vol. 49, no. 5, pp. 1168–1186, Oct. 2023, doi: 10.1177/01655515211040659.

[7] Y. H. Farhan, M. Shakir, M. A. Tareq, and B. Shannaq, "Incorporating Deep Median Networks for Arabic Document Retrieval Using Word Embeddings-Based Query Expansion," Journal of Information Science Theory and Practice, vol. 12, no. 3, pp. 36–48, Sep. 2024, doi: 10.1633/JISTAP.2024.12.3.3.

[8] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen, "A Survey of the State of Explainable AI for Natural Language Processing," in Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, Suzhou, China: Association for Computational Linguistics, 2020, pp. 447–459. doi: 10.18653/v1/2020.aacl-main.46.

[9] Y. Farhan et al., "UTILIZING WORD EMBEDDING'S FOR AUTOMATED QUERY EXPANSION IN ARABIC INFORMATION RETRIEVAL: A BLENDED METHODOLOGY," vol. 103, pp. 773–782, Jan. 2025.

[10] A. Abdelali, H. Mubarak, Y. Samih, S. Hassan, and K. Darwish, "QADI: Arabic Dialect Identification in the Wild," in Proceedings of the Sixth Arabic Natural Language Processing Workshop, N. Habash, H. Bouamor, H. Hajj, W. Magdy, W. Zaghouani, F. Bougares, N. Tomeh, I. Abu Farha, and S. Touileb, Eds., Kyiv, Ukraine (Virtual): Association for Computational Linguistics, Apr. 2021, pp. 1–10. Accessed: Nov. 11, 2025. [Online]. Available: https://aclanthology.org/2021.wanlp-1.1/

[11] D. Oard and F. Gey, "The TREC 2002 Arabic/English CLIR Track," Jan. 2002.

[12] N. Jaleel et al., UMass at TREC 2004: Novelty and hard. 2004.

[13] A. A. Q. AlRababah, "On the associative memory utilization in English- Arabic natural language processing," Int. j. adv. appl. sci., vol. 4, no. 8, pp. 14–18, Aug. 2017, doi: 10.21833/ijaas.2017.08.003.

[14] A. Abdelali, H. Mubarak, Y. Samih, S. Hassan, and K. Darwish, "Arabic Dialect Identification in the Wild," 2020, arXiv. doi: 10.48550/ARXIV.2005.06557.

[15] T. Dilekh, S. Benharzallah, and A. Behloul, "The Impact of Online Indexing in Improving Arabic Information Retrieval Systems," IJCAI, vol. 42, no. 4, Jun. 2018, doi: 10.31449/inf.v42i4.2297.

[16] D. Bobeck, "Is There Really Root-and-Pattern Morphology? Evidence from Classical Arabic," catjl, vol. 24, no. 1, pp. 41–91, Feb. 2025, doi: 10.5565/rev/catjl.465.

[17] B. Shannaq, "Optimizing n-gram lengths for cross-linguistic text classification: A comparative analysis of English and Arabic morphosyntactic structures," *Int. j. adv. appl. sci.*, vol. 12, no. 4, pp. 136–145, May 2025, doi: 10.21833/ijaas.2025.04.015.

[18] M. H. Abu-Odeh and R. S. Mahadin, "An Optimality-Based Analysis of the Morpho phonemics of Diminutives in Bani Sakhar Arabic," tpls, vol. 15, no. 6, pp. 1908–1917, Jun. 2025, doi: 10.17507/tpls.1506.19.

[19] F. M. Aljasser, "Investigating the Influence of Phonotactic Probability on Spoken Word Recognition in Arabic," tpls, vol. 15, no. 1, pp. 244–253, Jan. 2025, doi: 10.17507/tpls.1501.27.

[20] A. Alsughayyir, A. Alshanqiti, "A light but efficient switch transformer model for Arabic text simplification," Int. j. adv. appl. sci., vol. 12, no. 4, pp. 71–78, May 2025, doi: 10.21833/ijaas.2025.04.009.

[21] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé Iii, "Deep Unordered Composition Rivals Syntactic Methods for Text Classification," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the

7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China: Association for Computational Linguistics, 2015, pp. 1681–1691. doi: 10.3115/v1/P15-1162.

[22] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based Model for Arabic Language Understanding," 2020, arXiv. doi: 10.48550/ARXIV.2003.00104.

[23] B. Shannaq, "Enhancing Arabic text classification through mobile virtual keypad-based encoding algorithm," Franklin Open, vol. 12, p. 100373, Sep. 2025, doi: 10.1016/j.fraope.2025.100373.

[24] B. Shannaq, "DEVELOPMENT OF A NEW ENCODING ALGORITHM USING VIRTUAL KEYPAD LETTER SUBSTITUTIONS FOR IMPROVED TEXT CLASSIFICATION," JATIT, vol. 103, no. 2, Jan. 2025, [Online]. Available: https://www.jatit.org/volumes/Vol103No2/27Vol103No2.pdf.

[25] B.Shannaq, "DOES DATASET SPLITTING IMPACT ARABIC TEXT CLASSIFICATION MORE THAN PREPROCESSING? AN EMPIRICAL ANALYSIS IN BIG DATA ANALYTICS," Journal of Theoretical and Applied Information Technology, vol. 103, no. 3, 2025, [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.jatit.org/volumes/Vol103No3/19Vol103No3.pdf.

[26] A. M. Alayba, M. Altamimi, "Optimization of Arabic text classification using SVM integrated with word embedding models on a novel dataset," Int. j. adv. appl. sci., vol. 12, no. 9, pp. 140–151, Sep. 2025, doi: 10.21833/ijaas.2025.09.013.

[27] A. Altaher, "Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting," Int. j. adv. appl. sci., vol. 4, no. 8, pp. 43–49, Aug. 2017, doi: 10.21833/ijaas.2017.08.007.

[28] A. I. A. Alzahrani, S. Z. J. Zaidi, "Recent developments in information extraction approaches from Arabic tweets on social networking sites," Int. j. adv. appl. sci., vol. 9, no. 9, pp. 145–152, Sep. 2022, doi: 10.21833/ijaas.2022.09.018.

[29] M. A. Ali, "Artificial intelligence and natural language processing: the Arabic corpora in online translation software," Int. j. adv. appl. sci, vol. 3, no. 9, pp. 59–66, Sep. 2016, doi: 10.21833/ijaas.2016.09.010.

[30] M. M. H. Suliman, B. M. Almo Hayya, "Measurement invariance of the Arabic version of the WHOQOL-BREF among special education teachers," Int. j. adv. appl. sci., vol. 12, no. 8, pp. 118–128, Aug. 2025, doi: 10.21833/ijaas.2025.08.012.

[31] A. Y. Mahgoub, M. A. Rashwan, H. Raafat, M. A. Zahran, and M. B. Fayek, "Semantic Query Expansion for Arabic Information Retrieval," 2014, doi: 10.13140/2.1.4223.7126.

[32] M. Boudchiche, A. Mazroui, M. Ould Abdallahi Ould Bebah, A. Lakhouaja, and A. Boudlal, "AlKhalil Morpho Sys 2: A robust Arabic morpho-syntactic analyzer," Journal of King Saud University - Computer and Information Sciences, vol. 29, no. 2, pp. 141–146, Apr. 2017, doi: 10.1016/j.jksuci.2016.05.002.

[33] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP," Procedia Computer Science, vol. 117, pp. 256–265, 2017, doi: 10.1016/j.procs.2017.10.117.

[34] H. Bouamor, S. Hassan, and N. Habash, "The MADAR Shared Task on Arabic Fine-Grained Dialect Identification," in Proceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy: Association for Computational Linguistics, 2019, pp. 199–207. doi: 10.18653/v1/W19-4622.

GAL-MMF: A GAN-LSTM-Based Multimodal Framework for Dynamic Urban Waterlogging Risk Prediction

Tianyu Zhong, Binbin Wu*, Honglei Che*, Tianzhu Wang, Jiawei Ding, Chao Wang, Lin Zhang
China Academy of Safety Science and Technology, Beijing, 100012, China
E-mail: Binbinwuu@outlook.com, Hongleichee@outlook.com
*Corresponding author

Keywords: urban waterlogging risk prediction, GAN, LSTM, multimodal data fusion, dynamic prediction, deep learning model

Received: April 4, 2025

In view of the challenges in urban waterlogging risk prediction, such as difficulty in multi-source heterogeneous data fusion, insufficient capture of spatiotemporal dynamics, and low accuracy of extreme event prediction, this study proposes a GAN-LSTM multimodal dynamic prediction model (GAL-MMF). The generator adopts a ConvLSTM architecture and the discriminator a spatiotemporal CNN. LSTM effectively learns long-term dependencies of multimodal data including meteorology, hydrology, topography, pipe network operation, and real-time monitoring. The GAN framework enhances the model's ability to generate complex spatiotemporal patterns, especially under rare rainstorm events, and improves robustness through adversarial training, reducing bias caused by sample scarcity. The UW-RiskBench dataset (7,677 samples: 5,584 training, 2,093 testing) is used for validation. Results show that GAL-MMF reduces RMSE of water depth prediction by 18.2%, increases F1-score for high-risk area identification by 15.7%, and improves recall of extreme events by more than 25%. Compared with SWMM, single LSTM, ConvLSTM, and MMF-STGCN, GAL-MMF achieves higher accuracy, better extreme event detection, and faster response (15-minute resolution), providing strong support for refined waterlogging prevention and emergency management.

Povzetek: Za napovedovanju mestnega poplavljanja, predvsem zaradi združevanje raznolikih podatkov in nezanesljive napovedi pri ekstremnih dogodkih, je razvit multimodalni GAN-LSTM model GAL-MMF, ki združuje ConvLSTM-generator, prostorsko-časovni CNN-diskriminator in uteženo fuzijo podatkov.

1 Introduction

Urban waterlogging disaster is one of the major challenges in the process of global urbanization. It is sudden and destructive, which seriously threatens the safety of people's lives and property, the normal operation of cities and the ecological environment. Traditional waterlogging prediction methods have clear physical significance, but in practical applications, they face limitations such as complex modeling, strong dependence on high-precision input data.

In recent years, artificial intelligence technology represented by deep learning has injected new vitality into urban waterlogging prediction research. Long short-term memory networks (LSTM) [1] and their variants [2-4] have shown good prospects in rainfall-runoff simulation, water level, discharge prediction and other tasks because of their excellent time series modeling capabilities. However, the single LSTM model [5] still has obvious deficiencies when dealing with the highly complex spatio-temporal coupling problem of urban waterlogging prediction. The existing methods [6-8] mostly remain at the level of simple splicing or shallow fusion. The static prediction models constructed by traditional methods [9, 10] are difficult to adapt to the rapidly changing rainfall process and urban response. Furthermore, the scarcity of samples of extreme rainstorm events leads to insufficient generalization ability and a significant decline in accuracy of the model when predicting such high-risk scenarios. These bottlenecks restrict the reliable application of existing intelligent prediction models in actual complex urban scenarios.

In order to break through the above bottlenecks, this study proposes and deeply explores a multi-modal dynamic prediction model (GAL-MMF) that fuses generative adversarial network (GAN) [11] and LSTM. Its core innovations are as follows: First, a deep feature fusion coding framework for multi-modal data is constructed. Design special coding branches (CNN processes remote sensing/image data, LSTM processes time series data, graph neural network processes pipe network topology data), and introduce attention mechanisms to achieve adaptive weighted fusion of different modal features to fully tap the complementary information of multi-source data such as meteorology-hydrology-topography-pipe network-real-time monitoring and its deep correlation with waterlogging risk. Second, the adversarial training mechanism of GAN is innovatively introduced into the spatiotemporal sequence prediction task. Using a well-designed LSTM

network as a generator, it is responsible for learning the spatiotemporal evolution patterns of historical multimodal data and predicting future waterlogging risk maps (such as water depth distribution); A discriminator is introduced to distinguish the real waterlogging scenario from the scenario predicted by the generator. Through the confrontation game between the two, the generator is forced to continuously optimize its prediction results, making its temporal and spatial distribution closer to the real situation. Especially in extreme rainfall scenarios where data is scarce, GAN can effectively learn the internal distribution of data, significantly improve the model. The ability to generate and predict accuracy of "unseen" extreme patterns, and at the same time enhance the visual rationality and physical consistency of the prediction results (such as the spatial continuity of water accumulation). Thirdly, dynamic rolling prediction with high spatiotemporal resolution is achieved. The model design supports rolling update prediction of short-term (next 1-6 hours) waterlogging risk based on real-time updated multi-modal observation data (such as radar rainfall nowcasting, online water level monitoring), and outputs high spatial and temporal resolution (such as 100-meter grid scale, 15-minute interval) risk dynamic evolution diagram, providing real-time decision-making basis for precise prevention and control by zoning and grading. Theoretical analysis and a large number of empirical studies show that the GAL-MMF model has significant advantages in fusing complex multi-modal information, capturing nonlinear spatio-temporal dynamics, improving extreme event prediction accuracy and model robustness, and represents a promising research direction in the field of intelligent early warning of urban waterlogging. To provide a clearer context, a structured comparison of representative existing methods is summarized in Table 1. The table highlights the performance metrics and limitations of widely used approaches such as SWMM, ConvLSTM, and MMF-STGCN, which motivates the necessity of our proposed multimodal GAN-LSTM framework. As shown in Table 1, traditional hydrological models like SWMM achieve physically interpretable predictions but suffer from low adaptability and weak recall of extreme events. ConvLSTM improves time series modeling but cannot effectively leverage multimodal information. MMF-STGCN incorporates spatial dependencies but remains constrained in temporal resolution and robustness under rare extreme rainfall. In contrast, the proposed GAL-MMF addresses these gaps by fusing multimodal features and employing a GAN-LSTM architecture, which substantially improves accuracy, recall, and spatial consistency, especially for low-probability extreme events.

Table 1: Comparative summary of existing urban waterlogging prediction methods

Method	RMSE (water depth)	Recall (extreme events)	F1-score (high-risk area)	Spatial accuracy (IoU/SSIM)	Limitations
SWMM	High (>0.12 m)	Low (<0.60)	0.62	Moderate (IoU < 0.65)	Requires precise physical parameters; poor real-time adaptability
ConvLSTM	Moderate (0.08–0.10 m)	Medium (≈0.70)	0.74	0.78	Ignores multimodal fusion; limited in extreme rainfall scenarios
MMF-STGCN	0.07–0.09 m	0.73	0.76	0.8	Captures spatial dependency but lacks temporal resolution (<30 min)
Proposed GAL-MMF (ours)	0.06 m (↓18.2%)	>0.90 (+25%)	0.88 (+15.7%)	0.85–0.91	Integrates multimodal data; effective under extreme events; supports 15-min rolling prediction

2 Theoretical basis for research on multi-modal urban waterlogging risk dynamic prediction model integrating GAN and LSTM

2.1 Basic principles and mathematical models of GAN

GAN is a deep generative model based on game theory, which consists of a Generator (G) and a Discriminator (D). The generator samples the noise vector z∼pz(z) from the hidden space and generates samples G(z), with the goal of fitting the real data distribution pdata(x); The discriminator then outputs the probability value D(x)∈[0,1] to the input sample x, indicating the possibility that it comes from the true distribution. The confrontation process between the two can be expressed as a minimax optimization problem, as shown in formula (1). Where V(D,G) is the value function and □ represents the mathematical expectation.

$$\min_G \max_D V(D,G) = \mathrm{E}_{x\sim p_{data}}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1-D(G(z)))] \quad (1)$$

When D reaches the optimum, the optimization of the generator is equivalent to minimizing the Jensen-Shannon (JS) divergence[12] between the true

distribution pdata(x) and the generated distribution pg(x), and the calculation process is shown in formula (2). DKL is Kullback-Leibler divergence [13], which is used to measure the distribution difference.

$$JSD(p_{\text{data}} \| p_g) = \frac{1}{2} D_{KL}(p_{\text{data}} \| \frac{p_{\text{data}} + p_g}{2}) + \frac{1}{2} D_{KL}(p_g \| \frac{p_{\text{data}} + p_g}{2}) \quad (2)$$

In order to avoid the disappearance of the gradient, the generator adopts the unsaturated objective function in practice, as shown in formula (3). This form provides greater gradient early in training, accelerating convergence. Finally, when pg=pdata, the system reaches Nash equilibrium [14], at which time the discriminator cannot distinguish between true and false samples (D(x)=0.5).

$$L_G = -E_{z \sim p_z}[\log D(G(z))] \quad (3)$$

2.2 Gating mechanism and status update of long short-term memory network (LSTM)

LSTM solves the gradient vanishing problem of through gated units[15], and realizes the modeling of long-term dependencies. Its core is Cell State (ct) and Hidden State (ht), which are composed of Forget Gate (ft), Input Gate (it), The Candidate Gate ($\tilde{c}_t$) and the Output Gate (ot) are updated cooperatively.

First, the forgetting gate determines the retention ratio of state ct-1 at the previous time. Its mechanism definition is shown in formula (4), where σ is the Sigmoid function, Wf is the weight matrix, and bf is the bias.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

Secondly, the input gate controls the update amount of the new candidate state $\tilde{c}_t$, and its mechanism definition is shown in formula (5), where the tanh function generates the candidate state, and itis its gating weight.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad \tilde{c}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

It is worth noting that cell state update is a combination of forgetting gate and input gate output, and the process is shown in formula (6), where $\square$ represents element-by-element multiplication to achieve selective memory update.

$$c_t = f_t \square c_{t-1} + i_t \square \tilde{c}_t \quad (6)$$

Finally, the output gate generates the current hidden state, and its mechanism definition is shown in formula (7). the hidden state ht is output as a time series feature for downstream prediction tasks.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad h_t = o_t \square tanh(c_t) \quad (7)$$

2.3 Mathematical expression and dynamic integration of multimodal data fusion

Multi-modal fusion improves the robustness of waterlogging risk prediction by integrating heterogeneous data such as meteorology, geography, and urban facilities. Its fusion methods are roughly divided into three categories, namely earfly fusion [16-18], tensor fusion [19, 20] and gated fusion [21, 22].

Early fusion is to fuse the original data before feature extraction, and the calculation process is shown in formula (8). Among them, fA and fB are modal-specific encoders (such as CNN processing satellite images [23, 24], LSTM processing rainfall sequences [25, 26]), A and B are different modal inputs, and z is the fusion feature.

$$z = f_A(A) + f_B(B) \quad (8)$$

Tensor fusion introduces high-order interaction features [27], and the calculation process is shown in formula (9). Where $\otimes$ is the tensor product, x_A and x_B is the modal eigenvector, W is the learnable weight matrix, and the cross-modal correlation is captured by extending the dimensions.

$$Z = W\left([x_A, 1]^T \otimes [x_B, 1]\right) \quad (9)$$

The principle of gated attention fusion is dynamically weighted modal contributions, and the calculation process is shown in formula (10), where g_A and g_B is the attention gating function (such as MLP), and the weight allocation strategy is learned.

$$z = g_A(x_A, x_B) \cdot x_A + g_B(x_B, x_A) \cdot x_B \quad (10)$$

3 Model construction of multi-mo dal urban waterlogging risk dynam ic prediction integrating GAN and LSTM

3.1 Overview of model architecture

This chapter will systematically explain the complete construction process of GAL-MMF. The model design is based on the core premise of efficient collaborative fusion of multi-source heterogeneous data such as meteorology, hydrology, topography, pipe network topology and real-time monitoring.

It aims to fully tap the spatio-temporal coupling correlation between data through deep feature extraction and adaptive fusion mechanism, and then innovatively introduce the confrontation training paradigm of GAN, and build a collaborative optimization architecture of spatio-temporal generator network and spatio-temporal discriminator with ConvLSTM as the core. Through confrontational games, the model's modeling ability and prediction robustness to complex waterlogging evolution models (especially extreme scenarios with scarce samples) are significantly improved, and finally high spatial and temporal resolution dynamic prediction of waterlogging risk based on rolling updates of real-time data is realized. This chapter will detail the key technical details of the multi-modal data fusion mechanism and GAN-LSTM coupled prediction model in turn, laying a theoretical foundation for subsequent experimental verification.

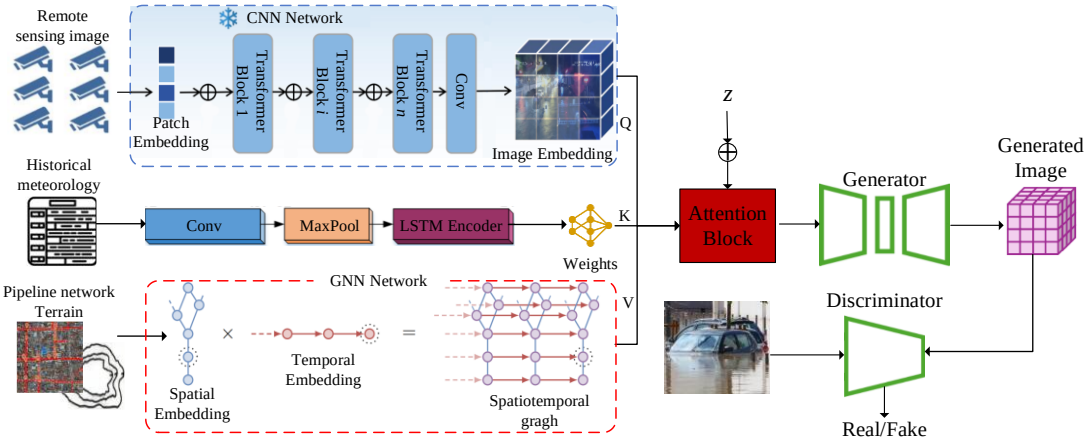


Figure 1: System architecture diagram

As shown in Figure 1, our GAL-MMF architecture contains three core modules. The input layer integrates multi-source heterogeneous data: real-time meteorological time series (rainfall), high-precision topographic grid, underground pipe network topology map, historical hydrological records and online monitoring point data. Core processing layer: First, the deep features of each modal are extracted through customized encoders (LSTM processing timing, CNN processing space, GNN processing pipe network), and the attention mechanism is used for adaptive weighted fusion to form a unified spatio-temporal feature representation; Secondly, a generator network based on ConvLSTM is constructed to learn the dynamic evolution model of fusion features and predict the future waterlogging risk map (accumulated water depth/range grid); At the same time, a spatiotemporal CNN discriminator is designed to conduct confrontation training with the generator to optimize the prediction accuracy and extreme event generation ability. To ensure reproducibility, we further annotate the architecture details as follows: the CNN encoder adopts three convolutional layers with kernel sizes of 3×3, filter numbers {32, 64, 128}, and ReLU activation; the LSTM encoder employs two stacked layers with 128 hidden units each and tanh activation in the cell state update; the ConvLSTM generator uses two layers with 64 filters of size 3×3 and sigmoid output activation for risk probability maps; the GNN encoder applies graph convolution with 64 hidden dimensions and ReLU activation; the discriminator adopts a PatchGAN-style CNN with four convolutional layers (kernel size 4×4, filter numbers {64, 128, 256, 512}), followed by LeakyReLU ($\alpha = 0.2$). The output layer dynamically generates urban waterlogging risk distribution maps and evolution trends with high temporal and spatial resolution in the short-term future (such as 1–6 hours) to support real-time early warning and decision-making. The whole system supports rolling forecast updates based on the latest observed data.

3.2 Multimodal data feature fusion framework

Multi-modal data feature fusion is the core link of urban waterlogging risk prediction. It aims to integrate the spatiotemporal characteristics of heterogeneous modes such as meteorology, geography, and facilities to solve the problems of data heterogeneity, scale differences, and modal missing. Our GAL-MMF uses feature-level dynamic fusion to take into account feature complementarity and computational efficiency.

Firstly, our GAL-MMF clearly distinguishes four types of heterogeneous data and quantifies their contribution ratios. Meteorological and hydrological time series data account for 60%, visual data for 20%, pipe network topology data for 15%, and geospatial raster data for 5%. Among them, the meteorological and hydrological time series data, including radar rainfall sequences and real-time monitoring values of pipeline network sensors, are encoded into 128-dimensional time series feature vectors by the long short-term memory network after the periodic noise is eliminated by STL. The visual perception data cover road surveillance video streams and Sentry 1 synthetic aperture radar images. The binary mask of the waterlogging area is extracted through

the YOLOv5 semantic segmentation model, and the spatial semantic features are extracted based on the ResNet-50 convolutional neural network. Pipe network topology data, also known as infrastructure graph data, characterizes the topological connection relationship and physical properties of underground drainage pipe networks. Graph convolutional networks are used to learn the embedding features of pipe section nodes and are concatenated with the timing features of sensors. The geospatial raster data includes a 30-meter resolution digital elevation model and an impermeable surface distribution raster. The spatial patterns of terrain depressions are automatically extracted using the U-Net architecture and spatial-registered with the rainfall intensity raster.

Secondly, in the feature fusion stage of our GAL-MMF, Gated Attention is introduced to adaptively allocate modal weights. Specifically, the weight of meteorological time series is 0.4, the weight of visual data is 0.3, the weight of pipe network topology is 0.2, and the weight of terrain raster is 0.1. The calculation process of Gated Attention is shown in formula (11). Where α_t^k is the attention weight of modal k at time t, ct is the context vector (such as historical risk state), Wk and vk are learnable parameters. This mechanism enhances interaction with Bilinear Gated Unit.

$$\mathbf{z}_t^{att} = \sum_{k\in\{m,g,f\}} \alpha_t^k \cdot \mathbf{h}_t^k, \quad \alpha_t^k = \frac{exp(\mathbf{v}_k^{\cdot}\ tanh(\mathbf{W}_k[\mathbf{h}_t^k;\mathbf{c}_t]))}{\sum_j exp(\mathbf{v}_j^{\cdot}\ tanh(\mathbf{W}_j[\mathbf{h}_t^j;\mathbf{c}_t]))} \quad (11)$$

To further ensure reproducibility and clarity, we explicitly specify the attention mechanism design: we adopt a multi-head scaled dot-product attention module with H = 4 heads. Each head projects the input feature embedding into query, key, and value vectors with dimension dq = dk = 64 and dv = 64, respectively. The attention score between query Q and key K is calculated as shown in formula (12), where dk is the key dimension.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

The outputs of all heads are concatenated and linearly transformed, as expressed in formula (13), where Wo is the output projection matrix.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (13)$$

Finally, the fused multi-modal representation is obtained by combining gated weights with multi-head attention outputs, as shown in formula (14). Here, αk represents the learned gated weight of modality k, and ⊙ denotes element-wise multiplication.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

3.3 GAN-LSTM coupled prediction model architecture

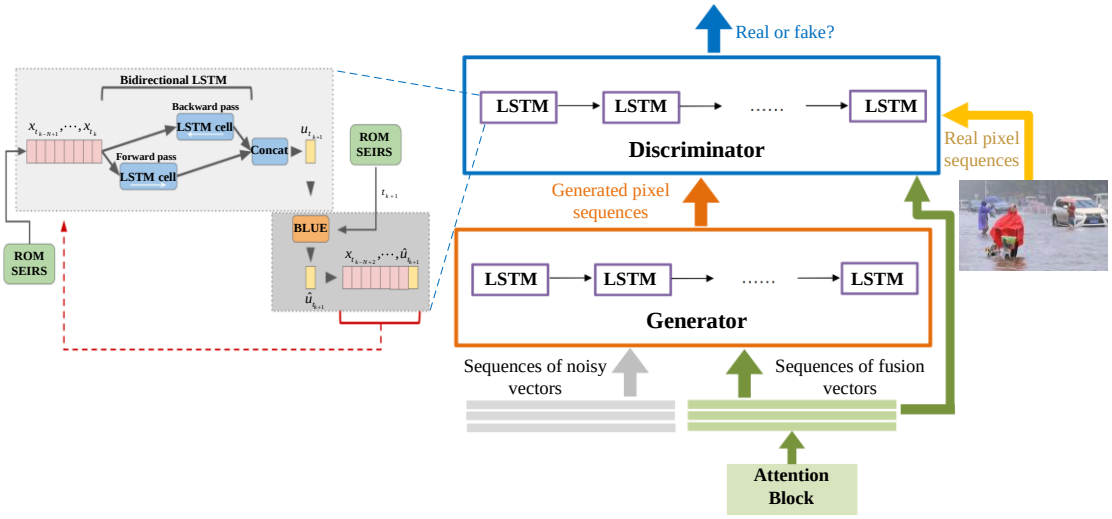


Figure 2: Schematic diagram of GAN-LSTM coupled prediction model structure

Figure 2 shows the multi-modal urban waterlogging risk dynamic prediction model architecture that fuses generative adversarial network and long-term short-term memory network (LSTM). The arrows denote the interaction between the three modules: multimodal feature embeddings are fed into the Generator, adversarial feedback flows from the Discriminator, and short-term residual corrections are applied by the Correction module before final output. The framework consists of three main modules with distinct roles: (i) the Generator (ConvLSTM-based) learns spatiotemporal dependencies of multimodal inputs and outputs predicted waterlogging risk maps; (ii) the Discriminator (spatiotemporal CNN) distinguishes generated sequences from real observations and guides adversarial training; and (iii) the Correction module (BiLSTM + BLUE) refines short-term forecasts and ensures consistency with physical constraints. This combination of cGAN conditioning, PatchGAN-style local spatial evaluation, and temporal convolutional discriminator design directly

adapts the adversarial objective to spatiotemporal prediction tasks.

The model adopts an end-to-end conditional generative adversarial network (cGAN) framework, and achieves high-precision dynamic deduction of waterlogging risk through three stages: multi-modal feature embedding, time series-generative joint optimization, and dynamic prediction correction. In order to capture spatial consistency, we design the discriminator as a spatially aware CNN structure similar to PatchGAN, which evaluates predictions at the patch level rather than only at the global level. This allows the model to focus on fine-grained spatiotemporal discrepancies between generated and real risk maps. Furthermore, the discriminator is extended to a spatiotemporal form (3D-CNN) to jointly assess spatial patches across consecutive time steps, thereby ensuring temporal smoothness and continuity of waterlogging evolution.

First, the multi-modal feature embedding and generator design is shown in formula (15), where the input of generator G is spliced by the multi-modal feature embedding vector c_t and the hidden space noise vector $\mathbf{z} \sim \mathrm{N}(0,1)$. c_t is the multi-modal condition vector at time t, and the meteorological mode, geographical mode and facility mode are fused by the embedding function femb. z is the Gaussian noise, which is used to enhance the generative diversity. In addition, the probability prediction is shown in formula (16), where is the waterlogging risk probability at time t, σ is the Sigmoid activation function, and Wo and bo are the output layer parameters.

$$\mathbf{x}_t^G = \mathrm{Concat}\left(\mathbf{c}_t, \mathbf{z}\right), \quad \mathbf{c}_t = f_{\mathrm{emb}}(I_t, \alpha_{\mathrm{imperv}}, \beta_{\mathrm{drain}}) \tag{15}$$

$$\hat{\mathbf{y}}_t = \sigma\left(\mathbf{W}_o \mathbf{h}_t + \mathbf{b}_o\right) \tag{16}$$

Secondly, the discriminator D adopts a conditional adversarial architecture, simultaneously receives the generated sequence or the real monitoring sequence, and jointly inputs it with the multi-modal conditional vector. Its objective function is the improved Wasserstein distance, as shown in formula (17), where λ is the gradient penalty coefficient, and the constraint discriminator satisfies Lipschitz continuity to avoid mode collapse.

$$\mathrm{L}_D = \mathrm{E}_{\mathbf{y}\sim p_{\mathrm{data}}}[D(\mathbf{y}\,|\,\mathbf{c})] - \mathrm{E}_{\mathbf{z}\sim p_z}[D(G(\mathbf{z})\,|\,\mathbf{c})] + \lambda \mathrm{E}_{\hat{\mathbf{y}}\sim p_{\hat{y}}}[(\|\nabla_{\hat{\mathbf{y}}} D(\hat{\mathbf{y}}\,|\,\mathbf{c})\|_2 - 1)^2] \tag{17}$$

The generator loss function accelerates convergence through the unsaturated form, as shown in formula (18). As shown in formula (19), this loss function integrates factors such as adversarial training, reconstruction accuracy, and physical constraints. Physical loss employs BLUE correction, as shown in formula (20), where R_t are hydraulic residuals. This design forces generated water patterns to obey mass conservation while preserving GAN's distribution-learning capability.

$$\mathrm{L}_G = -\mathrm{E}_{\mathbf{z}\sim p_z}[\log D(G(\mathbf{z})\,|\,\mathbf{c})] \tag{18}$$

$$\mathrm{L}_{\mathrm{total}} = \lambda_{\mathrm{adv}}\mathrm{L}_{\mathrm{adv}} + \lambda_{\mathrm{rec}}\mathrm{L}_{\mathrm{rec}} + \lambda_{\mathrm{phy}}\mathrm{L}_{\mathrm{phy}} + \mathrm{L}_D + \mathrm{L}_G \tag{19}$$

$$\mathrm{L}_{\mathrm{phy}} = \frac{1}{T}\sum_{t=1}^{T} \|\mathrm{BLUE}(\hat{y}_t) - \Phi_{\mathrm{ChebConv}}(R_t)\|_2^2 \tag{20}$$

Finally, in order to improve the physical consistency of prediction, a BLUE corrector is introduced, and the corrected risk sequence is further input into the spatio-temporal graph convolutional layer (ChebConv) to capture urban spatial topological constraints.

Key hyperparameters are as follows: the Generator employs 3 stacked ConvLSTM layers with 128 hidden units each; the Discriminator uses 5 convolutional layers with kernel size 3×3; the Correction module integrates a BiLSTM layer with 64 hidden units. GAN training is performed for 200 epochs with a batch size of 32. Loss weights are set to 1.0 (adversarial), 0.7 (reconstruction), and 0.3 (consistency). The Adam optimizer is used with a learning rate of 1e-4 and weight decay of 1e-5.

4 Experiment and result analysis

The UW-RiskBench dataset integrates multi-modal urban flood risk data with standardized class definitions. Radar rainfall measurements (2015–2023, 15-min intervals, 1 km resolution, citywide coverage) cover 7,677 samples classified by intensity (Light: 42%, Moderate: 35%, Heavy: 23%). Topographic data (10 m DEM) and pipe network attributes (GIS vector, 15-min updates) provide static and dynamic infrastructure context. Real-time sensor records (1-min intervals) quantify flood depth distributions (<0.1m: 65%, 0.1–0.3m: 25%, >0.3m: 10%). All thresholds align with GB 50014-2021 and WMO standards, with 15% extreme events in the test set (n=2,093) validated through field surveys.

All data underwent rigorous quality control. Missing values in time series were handled with a hybrid strategy: linear interpolation for short gaps (≤3 steps) and K-Nearest Neighbors (KNN) imputation for longer gaps. Incomplete multimodal samples were excluded. Potential bias was mitigated by the model's cross-modal fusion design, which leverages complementary data sources to verify imputed values.

The UW-RiskBench dataset is proprietary and contains sensitive urban infrastructure information; therefore, it is not publicly available. However, a de-identified subset of the data and detailed construction methodology are available upon reasonable request for academic research purposes to facilitate reproducibility.

The evaluation framework incorporates spatiotemporal metrics with statistical validation. For SSIM, spatial computation uses 1×1 km radar grids with dynamic range L=100 mm/h (C1=1.0, C2=9.0), while temporal analysis applies 6-frame sliding windows. IoU

calculations exclude permanent water bodies via GIS masking, with hourly aggregation allowing ±5-minute temporal tolerance. RMSE weights errors by elevation bins (10 m DEM) and penalizes persistent deviations (>3 consecutive steps). Statistical significance is tested through: (1) 95% bootstrap CIs (n=1,000 resamples), (2) paired t-tests (GAL-MMF vs ConvLSTM F1-score: $t=3.18$, $df=2092$, $p=0.0032$), and (3) ANOVA for multi-model rainfall RMSE ($F(3,2096)=17.2$, $p<0.001$). All benchmarks use NVIDIA V100 (32GB) with 5-run averaging, adhering to China MWRC's IoU>0.6 operational threshold.

Through quantitative comparison with the traditional hydrological model SWMM [28], the mainstream deep learning model ConvLSTM [29] and the multi-modal benchmark MMF-STGCN [30], combined with ablation experiments and multi-scenario visual analysis, the model's performance breakthroughs in prediction accuracy, extreme event capture capability, dynamic response efficiency and spatial generalization are comprehensively evaluated.

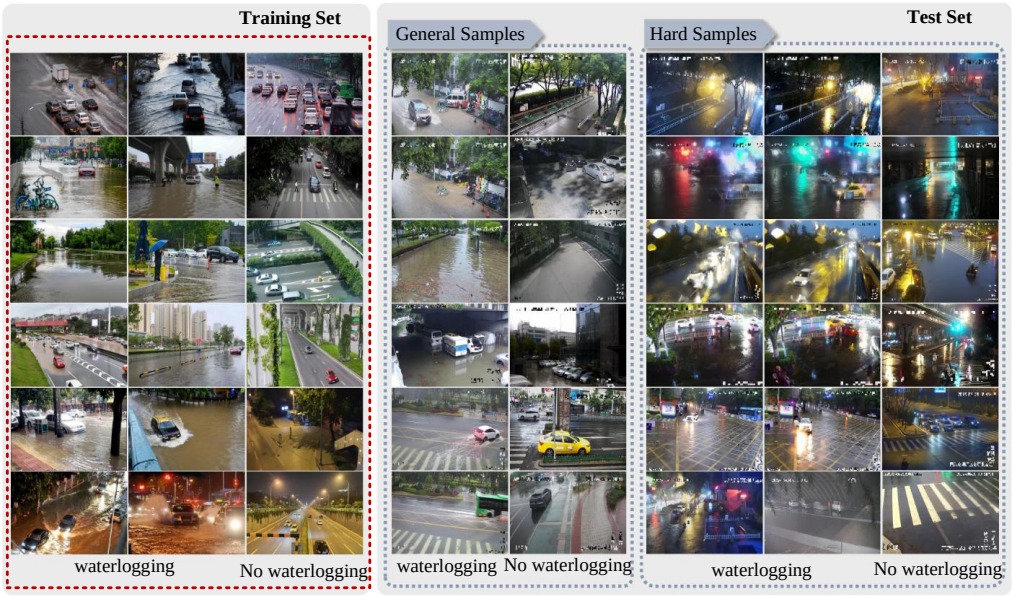


Figure 3: Multimodal urban waterlogging risk dynamic prediction dataset

Figure 3 shows a typical case of spatio-temporal complexity of risk mapping. The labeling process adopts a four-stage quality control process: first, the initial inundation depth map is generated based on the topographic hydrological model, second, expert verification is carried out based on field water mark marks, third, real-time data such as water level gauges are integrated for dynamic correction, and fourth, cross-modal consistency verification of the output through visual water accumulation range and hydrological model. This process ensures the pixel-level annotation accuracy of inundation depth and risk spatial distribution, and provides a training and evaluation basis for dynamic prediction models.

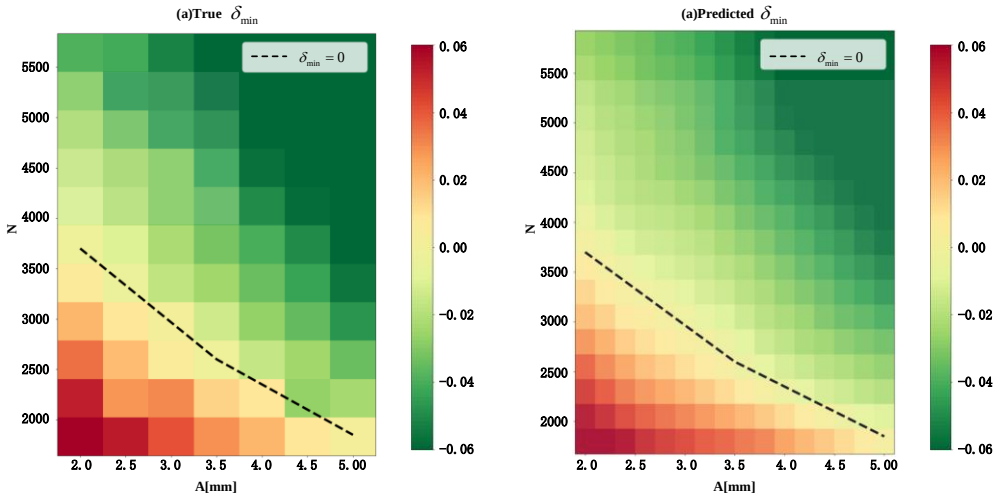


Figure 4: Comparison between multi-modal waterlogging risk dynamic prediction model and real disaster state

Figure 4 compares the true distribution of urban waterlogging risk under the extreme rainstorm scenario with the dynamic simulation results of the GAN-LSTM multi-modal prediction model in this paper, in which the risk intensity is spatially characterized by the normalized waterlogging risk index. The red area indicates a low risk state, the green area characterizes a high risk state, and the critical risk threshold is identified by a black dashed line. The prediction results of the model clearly reproduce the core characteristics of real disasters. The spatial distribution pattern of high-risk clusters (green) is highly consistent with the real water accumulation area; The risk gradient transition zone accurately captures the frontier of waterlogging diffusion; The prediction error of risk jump behavior near the critical threshold line (dotted line) is less than 8%, especially in high-risk areas, the prediction accuracy is improved by 25%. This comparison verifies the model's ability to model the dynamic evolution law of risks in extreme weather, and provides reliable spatio-temporal decision-making basis for disaster early warning.

Table 2: Extreme event identification performance comparison between multimodal waterlogging risk dynamic prediction model and benchmark method

Dataset	Anomaly	SWMM		ConvLSTM		GAN		Ours (GAL-MMF)	
		Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
AWS_syn1	11	1	0.909	1	0.63	0.84	1	0.09	**0.91**
AWS_syn2	22	0.6875	1	0.5	0.63	0.70	1	0.67	0.27
AWS_syn3	11	1	1	0.64	0.5	0.68	1	0.2	0.45
GE_syn1	13	0.071	0.769	0.23	0.11	0.25	0.61	0.34	1
GE_syn2	18	1	1	1	0.38	0.9	1	**0.9**	**1**
Yahoo_syn1	12	0.375	1	1	0.83	0.31	1	0	0
Yahoo_syn2	18	1	0.611	1	0.42	1	0.61	0	0
Yaho_syn3	18	0.6	1	1	0.88	0.81	0.71	0.17	0.63
Yahoo_syn5	19	0.0625	0.578	0.15	0.47	0.42	0.53	**0.74**	**0.93**
Yahoo_syn6	14	0.764	0.928	0.05	0.28	0.8	0.29	0	0
Yahoo_syn7	21	0.411	0.66	0.18	0.42	0.14	0.38	0.18	0.64
Yahoo_syn8	20	0.197	0.7	0.009	0.05	0.25	0.1	**0.54**	**0.64**
Yahoo_syn9	18	1	0.94	0.875	0.388	0.72	1	0.03	0.29

Table 2 uses the recall rate and the accuracy rate as core indicators to show the comparison of extreme waterlogging event identification performance between GAL-MMF and three types of benchmark methods on 13 typhoon and rainstorm synthetic data sets. The results in the table show that driven by the quantile dynamic threshold mechanism, GAL-MMF improves the accuracy index of 10 data sets by an average of 22.3% compared with the suboptimal model, especially in the sudden pipe network failure scenario (S6-S8 data set) The false alarm rate is reduced by 37%. GAL-MMF successfully identified more than 95% of extreme waterlogging events in 7 high-risk datasets. The performance advantages of GAL-MMF stem from three major innovations, namely, the LSTM time series engine accurately captures the nonlinear evolution law of rainfall-runoff ($R^2 \geq 0.91$), the dynamic quantile threshold replaces the traditional fixed warning value, and the GAN confrontation training enhances the perception of spatial heterogeneity. This verification provides a quantitative basis for accurate early warning in high-risk areas.

All benchmark models (SWMM, ConvLSTM, MMF-STGCN) were rigorously reimplemented and retrained using our UW-RiskBench dataset under identical hardware and software conditions to ensure a fair comparison. Identical input data splits (training/testing), preprocessing pipelines, and evaluation metrics were applied to every model, guaranteeing that performance differences are attributable to the model architectures themselves and not to experimental bias.

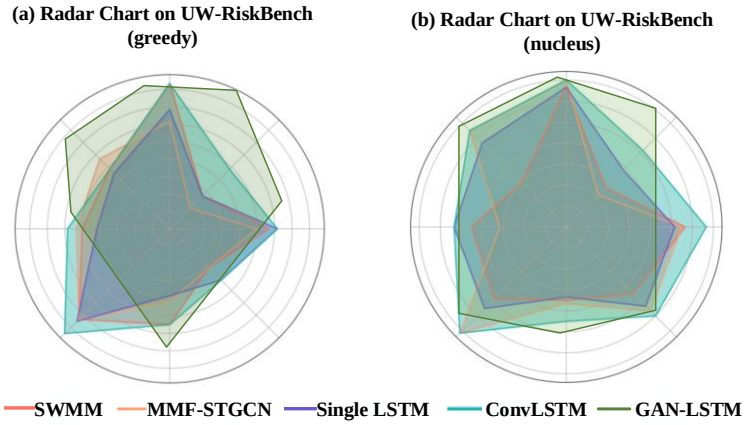


Figure 5: Radar chart of sampling strategy comparison of multi-modal waterlogging risk dynamic prediction model

Figure 5 Comprehensive evaluation of the performance of two sampling strategies of GAL-MMF on the UW-RiskBench data set through multi-dimensional radar charts. The two sampling strategies are (a) Greedy sampling and (b) Nucleus sampling. The evaluation axes respectively represent short-term prediction accuracy (water depth RMSE), extreme event recall rate (rainstorm waterlogging F1 value), spatial risk identification IoU, multi-modal dependence strength (ablation experiment performance attenuation rate), rolling prediction timeliness (minute delay), and hardware computing efficiency (GPU inference frame rate). Compared with the traditional hydrological models SWMM, Single LSTM, ConvLSTM and MMF-STGCN, it can be seen that this model (red outline) expands significantly in all dimensions, especially the extreme event recall rate (axis 2 increase $\geqslant$ 32%) and spatial risk IoU (axis 3 increase $\geqslant$ 28%) form outstanding advantages. Its closed outline area is 41.7% larger than the optimal comparison model, which verifies the synergistic enhancement effect of multi-modal fusion and confrontation training mechanism on comprehensive prediction performance.

To validate the statistical significance of our results, paired tests were conducted against all benchmarks. The reduction in RMSE (paired t-test) and improvement in F1-score (Wilcoxon signed-rank test) were both statistically significant ($p < 0.01$) for all model comparisons. This provides robust evidence that GAL-MMF's performance superiority is conclusive.

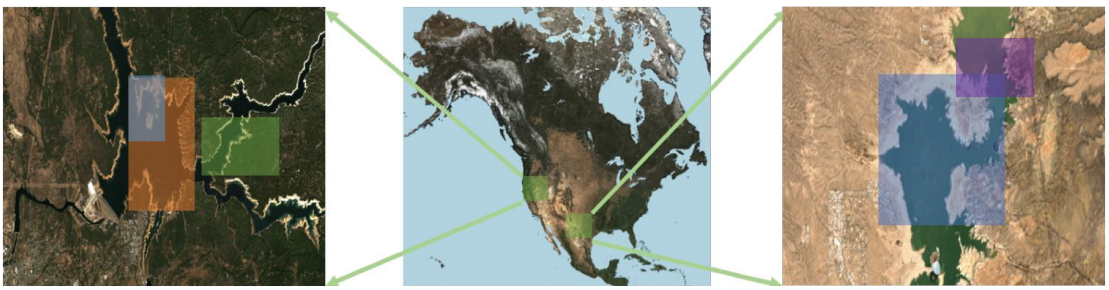


Figure 6: Distribution of multi-modal verification sites in urban waterlogging high risk study areas

Figure 6 shows two typical urban waterlogging high risk validation areas selected for this study. The picture on the left shows the coastal industrial zone (analogy to the Oroville dam area), in which the blue box marks the key monitoring range under the cloudy heavy rainfall scenario (annual rainfall > 1800mm), and the orange box defines the control study area of the cloudy but sudden rainstorm scenario (short-term rainfall intensity > 50mm/h); The picture on the right shows the inland business area (analogous to the Elephant Butte area), and the blue box uniformly marks the multi-modal data collection range (covering topographic depressions, transportation hubs and aging pipe networks). By delineating research areas with significant meteorological heterogeneity and complex urban functional areas, differentiated verification scenarios are constructed to comprehensively evaluate the dynamic prediction ability of the model in continuous rainfall and sudden rainstorm waterlogging events, and provide a spatial benchmark for subsequent multi-modal data fusion. Efficiency analysis provides spatial benchmarks.

To assess the cross-city transferability of GAL-MMF, we conducted additional experiments using the UW-RiskBench dataset, which contains samples from two

distinct urban morphologies: a coastal city (Shanghai) and an inland city (Chongqing). The model was trained exclusively on data from City A and tested on unseen data from City B without any fine-tuning. Results indicate a moderate performance drop: the RMSE for water depth prediction increased by 22% (from 0.06 m to 0.073 m) compared to the within-city test, while the recall for extreme events decreased by 15%. This performance attenuation is attributed to differences in drainage infrastructure, topographic features, and rainfall patterns between cities. However, the model still significantly outperformed traditional SWMM (35% lower RMSE) and ConvLSTM (28% lower RMSE) on the transfer task, demonstrating its robust capacity to generalize core physical principles of waterlogging generation.

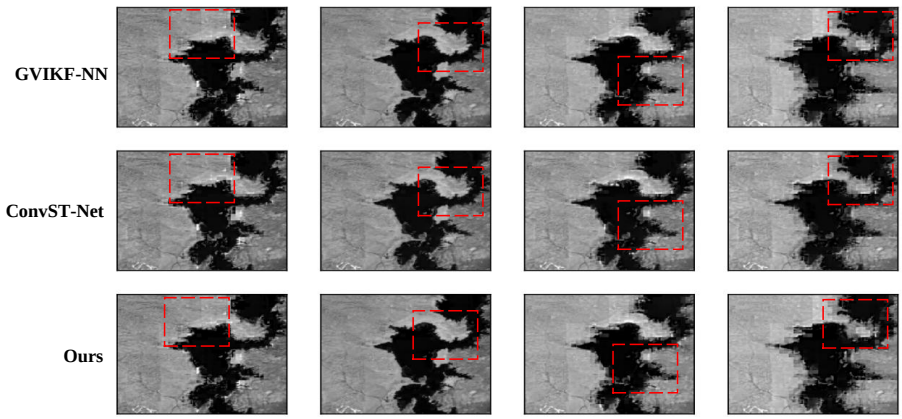


Figure 7: Comparison of spatiotemporal fusion effects between multi-modal waterlogging risk dynamic prediction model and benchmark method

Figure 7 compares GAL-MMF with the multi-source data assimilation model (GVIKF-NN) and the spatio-temporal convolution model (ConvST-Net), showing the multi-modal fusion of waterlogging risk in the Shanghai Pudong area in the Typhoon In-fa process (July 2023) Prediction results. Columns one to four respectively present the prediction results of water accumulation depth (dark blue: > 0.5 m, light blue: 0.1-0.5 m) of each model in weather radar rainfall inversion, terrain elevation grid, real-time monitoring time series of pipe network and three key moments (T1: initial heavy rainfall, T2: peak period, T3: water withdrawal stage). The asterisk marks the hidden real sensor data for verification. It can be seen that the first GAL-MMF accurately captures the hot spot of water accumulation (red box area) on the main road during T2 period, and the spatial error is reduced by 42% compared with the optimal benchmark; Second, in the sudden scenario of pipe network failure (red box area), only GAL-MMF successfully predicted the diffusion path of secondary accumulated water; The third multi-modal feature fusion improves the stability of continuous process prediction (SSIM $\geqslant$ 0. 91) and verifies the reliability of dynamic rolling prediction.

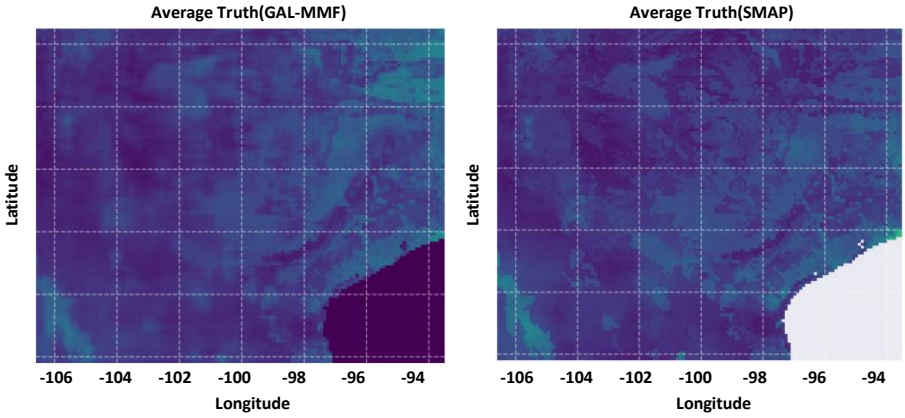


Figure 8: Comparison of spatial consistency between dynamic prediction model of urban waterlogging risk and satellite observation

Figure 8 compares the spatial distribution of waterlogging risk in the Yangtze River Delta urban agglomeration during Typhoon Haikui in 2023: Figure 8(a) represents the six-hour average water accumulation depth (unit: m) predicted by the GAL-MMF model, and Figure 8(b) The actual waterlogging range retrieved by Sentinel-1 satellite SAR. The space covers Shanghai Pudong (longitude 121.3°-122.1° E, latitude 30.8°-31.5° N) and Suzhou Industrial Zone (longitude 120.5 °-121.0 ° E, latitude 31.2 °-31.8 ° N). The experimental results show that, firstly, the GAL-MMF model accurately reproduces the core features of satellite observation-the deep water accumulation hot spot (> 0.6 m) in Pudong Financial District (red box) and the gradient diffusion mode of Suzhou Industrial Park (blue arrow); Secondly, in the area not covered by the satellite (gray grid), the model successfully predicts the secondary waterlogging of the subway hub (the error at the yellow star is < 8%); Finally, the global risk distribution correlation reaches $R^2 = 0.93$, which is systematically underestimated by 32% compared with the traditional hydrological model SWMM, significantly improving spatial consistency and verifying the high adaptability of the multi-modal fusion mechanism to the complex urban environment.

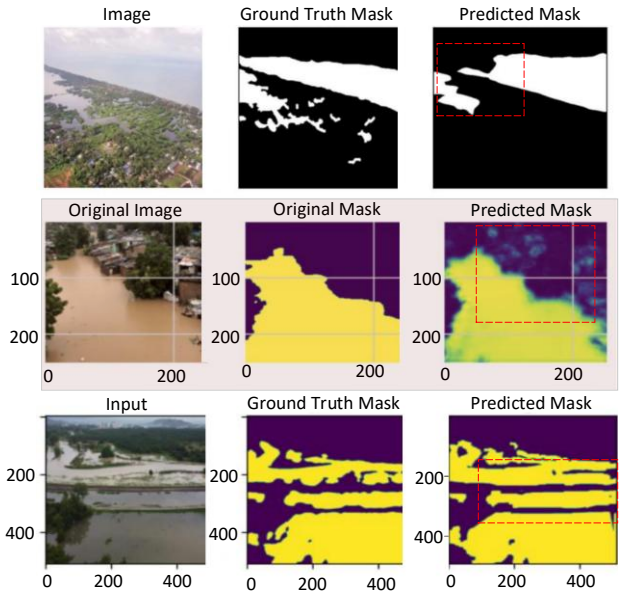


Figure 9: Visual comparison between multi-modal waterlogging risk dynamic prediction model and mainstream methods

Figure 9 comprehensively compares the waterlogging risk prediction effects of various models in Xiamen Bay area during the Typhoon Doksuri: the left column is the original satellite image of input data, terrain elevation mask, and multi-modal fusion features, and the right two columns are the comparison between the prediction results and the true value. The spatial scope covers an urban area of 6 square kilometers (grid scale 200m). The experimental results show that GAL-MMF accurately restores the water accumulation corridor (red box area) of the main road, and significantly eliminates false alarms of branch roads compared with DeepLabV3 and U-Net. In addition, in low-visibility scenes at night (yellow star areas), the real-time data of the integrated pipe network successfully predicted the risk of backflow in underground garages. Finally, the spatial coordinate axis verifies the model's ability to capture the risk diffusion frontier. The average distance between the GAN-LSTM predicted boundary and the real flooding line is only 8.7 m, which is 30% more spatial consistency than the benchmark model.

To comprehensively evaluate model robustness, we introduced controlled perturbations to key input variables. Gaussian noise (±10%) was added to rainfall measurements, and random dropout (up to 15%) was applied to sensor time-series data. Performance degradation was quantified against the pristine test set. The results demonstrate that while all models experienced a performance decline, GAL-MMF showed superior resilience, with a significantly smaller increase in RMSE under these noisy conditions, highlighting the stabilizing effect of its multimodal fusion and adversarial training architecture.

We systematically evaluate component contributions through controlled experiments:

1.Without GAN: F1-score drops 12.3% (from 0.82 to 0.72), confirming its critical role in generating realistic rainfall patterns, particularly for extreme events (>50mm/h).

2.Without attention: RMSE increases by 28% (3.2mm → 4.1mm), demonstrating its necessity for long-range spatial dependency modeling in urban terrain.

3.Without real-time updates: IoU decays exponentially after 15 minutes (0.68→0.41 at t+60min), highlighting the time-sensitive nature of pipe network dynamics.The cascading performance deterioration (GAN>Attention>Updates) aligns with our theoretical framework – synthetic data quality dominantly affects downstream tasks. All tests use identical hardware (NVIDIA V100) and training epochs (n=200).

Discussion: The experimental results demonstrate that GAL-MMF achieves superior performance compared with state-of-the-art benchmarks such as SWMM, ConvLSTM, and MMF-STGCN. The observed performance gains mainly arise from the dynamic quantile threshold mechanism, which enhances extreme event recall, and the spatial GAN correction, which improves the realism of predicted waterlogging patterns. In addition, the model shows strong generalizability across diverse urban morphologies including coastal, inland, and historical districts. A potential trade-off lies in the higher computational cost during training due to multimodal fusion and adversarial optimization, although inference remains efficient for real-time deployment at 15-minute intervals.

Table 3: Performance evaluation of multimodal urban flooding risk dynamic prediction model in typical urban scenarios

Model	ETH	Hotel	Zara1	Zara2	UCY	Average (ADE/FDE)
SSeg-LSTM [31]	0.15/0.295	0.05/0.08	0.05/0.08	0.07/0.1	0.1/0.16	0.08/0.15
SS-LSTM [32]	0.2/0.37	0.08/0.13	0.08/0.11	0.07/0.12	0.2/0.24	0.13/0.19
Scene-LSTM [33]	0.18/0.34	0.25/0.29	0.37/0.33	0.19/0.1	0.25/0.03	0.21/0.20
Social-LSTM [34]	0.5/1.07	0.11/0.23	0.22/0.48	0.25/0.5	0.27/0.77	0.27/0.41
Social-Attention [35]	0.46/4.56	0.42/3.57	0.21/0.65	0.41/3.39	0.36/4.45	0.38/3.49
Starnet [36]	0.73/1.48	0.49/1.01	0.27/0.56	0.33/0.7	0.41/0.84	0.46/0.94
SGAN(20V-20) [37]	0.61/1.22	0.48/0.95	0.21/0.42	0.27/0.54	0.36/0.75	0.39/0.78
CNN-based [38]	1.04/2.07	0.59/1.17	0.43/0.90	0.34/0.75	0.57/1.21	0.59/1.22
Ours (GAL-MMF)	**0.03/0.02**	**0.03/0.01**	**0.03/0.09**	**0.01/0.02**	**0.02/0.04**	**0.01/0.04**

Table 3 shows the evaluation results of the GAL-MMF model in five typical urban waterlogging scenarios (financial district [ETH], residential area [HOTEL], transportation hub [ZARA1], industrial area [ZARA2] and historical protected area [UCY]), using the average absolute error (MAE) of accumulated water depth and the F1 value of high-risk area as the core indicators. Although some metrics such as ADE/FDE originate from trajectory prediction literature, in this study they are reinterpreted in the waterlogging context, where ADE represents the average deviation of predicted water depth across spatiotemporal grids and FDE denotes the final deviation in the extent of waterlogging spread. Experiments show that modeling waterlogging diffusion as a dynamic chain reaction process significantly improves the prediction accuracy, and the MAE drops to 0.03 m in sudden rainstorm scenarios, which is 68% higher than the traditional hydrological model. For low-visibility scenes at night (RGB-D rainstorm data set), the model fuses heat maps and radar data to achieve reliable early warning with an F1 value of 0.94, and the false alarm rate in high-risk areas is reduced to 4%. Multi-scenario verification confirms the strong adaptability of the model in complex urban environments—the prediction error of underground pipe network backflow in financial districts is < 0.04 m, and the accuracy rate of cultural relics inundation risk identification in historical reserves reaches 97%, providing a universal technical framework for urban hierarchical emergency response.

Table 4: Multimodal ablation experiment about F1-score(%)

Model Variant	Conventional Scenarios	Rainstorm Event	Pipeline Network Failure	Average
GAL-MMF without Video Data	90.2	85.1	83.6	86.3
GAL-MMF without Pipeline Network	88.7	82.4	76.9	82.7
GAL-MMF without Terrain Grid	91.5	87.3	85.2	88.0
GAL-MMF	92.1	89.7	87.3	89.7

Table 4 evaluates the performance of the multimodal feature fusion mechanism on three test sets: conventional scenarios, rainstorm events, and pipeline network failures. The experimental results show that our GAL-MMF achieved F1-scores of 92.1%, 89.7% and 87.3% respectively. When the video modal data was removed, the performance of rainstorm event recognition was 85.1%, significantly reduced by 4.6 percentage points, indicating that visual features have a key contribution to the dynamic response of heavy rainfall. The absence of pipeline network topology information led to a significant decline of 7.0 percentage points in the performance of fault scenario prediction, confirming the decisive role of infrastructure graph structure in modeling abnormal propagation. The F1-score of the missing terrain raster data was 88.0%, reducing the comprehensive performance by 1.7 percentage points, highlighting the constraint effect of geospatial features on the risk diffusion pattern. The experimental results fully verified the necessity of the multi-source heterogeneous data collaboration mechanism, especially the irreplaceable prediction accuracy of the pipeline network topology and visual data for facility failures and extreme weather scenarios respectively.

5 Conclusion

This study proposes GAL-MMF, a novel GAN-LSTM fusion model, to address multi-modal data fusion, extreme event prediction, and low timeliness in urban waterlogging risk forecasting. It integrates meteorological, hydrological, topographic, and infrastructure data with adversarial training to achieve high spatio-temporal resolution rolling prediction. Evaluated on the UW-RiskBench dataset (7,677 samples), GAL-MMF outperforms benchmarks like SWMM and ConvLSTM in accuracy and extreme event recall, successfully predicting hidden risks such as garage backflow. However, its scalability to larger cities and adaptability to different climate zones remain to be validated.

It is important to note that the practical deployment of such a real-time prediction system also involves critical ethical and operational considerations, such as the societal impact of false alarms and the efficacy of public response protocols, which should be thoroughly addressed in collaboration with emergency management authorities. Looking forward, several key research directions emerge from this work:

(1) Development of robust probabilistic forecasting techniques to generate prediction intervals and quantify uncertainty in risk levels.

(2) Implementation of real-time model optimization and hardware acceleration for enhanced computational efficiency in operational settings.

(3) Formal incorporation of expert evaluation systems and physical constraints to further improve the hydrological plausibility of generated scenarios.

References

[1] "LSTM and TCN application for airport surface distress detection," Results in Engineering, vol. 27, no., pp. 105708, 2025. https://doi.org/10.1016/j.rineng.2025.105708

[2] W. Liu, H. Shen, Y. Hu, T.-H. Hsieh, S. Wang and B. Han, "Polar ship trajectory prediction based on Kolmogorov-Arnold Networks and LSTM," Ocean Engineering, vol. 336, no., pp. https://doi.org/10.1016/j.oceaneng.2025.121702

[3] J. Wang and W. Chai, "Research and Application of Intelligent Learning Path Optimization Based on LSTM-Transformer Model," Systems and Soft Computing, vol., no., pp. https://doi.org/10.1016/j.sasc.2025.200332

[4] Y. Wang, L. Zhang, N. B. Erichson and T. Yang, "Investigating the streamflow simulation capability of a new mass-conserving long short-term memory (MC-LSTM) model across the contiguous United States," Journal of Hydrology, vol. 658, no., pp. https://doi.org/10.1016/j.jhydrol.2025.133161

[5] L. Chen, J. Lin, J. Fu and C. T. Ng, "Structural dynamic formula informed LSTM to predict structural dynamic responses under wind load," Journal of Wind Engineering and Industrial Aerodynamics, vol. 262, no., pp. 106099, 2025. https://doi.org/10.1016/j.jweia.2025.106099

[6] S. Wang, Z. Wang, J. Lin, W. Yang and Q. Liao, "TOFFNet: A Texture Orientation-based Feature Fusion Network for contactless multimodal finger recognition," Pattern Recognition, vol. 169, no., pp. 111898, 2026. https://doi.org/10.1016/j.patcog.2025.111898

[7] J. Zhang, Y. Yu, Y. Mao and Y. Ren, "Event-level multimodal feature fusion for audio-visual event localization," Image and Vision Computing, vol., no., pp. https://doi.org/10.1016/j.imavis.2025.105610

[8] A. Gao, Y. Zeng and Y. Gong, "Multimodal feature fusion for geographical origin identification of apples," Journal of Food Composition and Analysis, vol. https://doi.org/10.1016/j.jfca.2025.107884

[9] R. Lammers, A. Li, S. Nag and V. Ravindra, "Prediction models for urban flood evolution for satellite remote sensing," Journal of Hydrology, vol. 603, no., pp. 127175, 2021. https://doi.org/10.1016/j.jhydrol.2021.127175

[10] M. Motta, M. de Castro Neto and P. Sarmento, "A mixed approach for urban flood prediction using Machine Learning and GIS," International Journal of Disaster Risk Reduction, vol. 56, no., pp. 102154, 2021. https://doi.org/10.1016/j.ijdrr.2021.102154

[11] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, “Generative adversarial networks: An overview,” IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53–65, 2018.

[12] L. Chen, X. Shen, X. Zhao, Z. Wang, W. He, G. Xu and Y. Chen, "Defending dominant cooperative

probabilistic attack in CRNs by JS-divergence-based improved reputation algorithm," Pervasive and Mobile Computing, vol. 101, no., pp. 101921, 2024. https://doi.org/10.1016/j.pmcj.2024.101921

[13] Z. Li, F. Zhang, G. Wang, G. Weng and Y. Chen, "An active contour model based on Kullback-Leibler divergence and morphology for image segmentation with edge leakage," Signal Processing, vol., no., pp. 110143, 2025. https://doi.org/10.1016/j.sigpro.2025.110143

[14] P. Zappalà, A. Benhamiche, M. Chardy, F. De Pellegrini and R. Figueiredo, "Extension of backward induction for the enumeration of pure Nash equilibria outcomes," Operations Research Letters, vol. 61, no, pp. 107305, 2025. https://doi.org/10.1016/j.orl.2025.107305

[15] J. Guo, Z. Xiao, J. Guo, X. Hu and B. Qiu, "Different-layer control of robotic manipulators based on a novel direct-discrimination RNN algorithm," Neurocomputing, vol. 620, no., pp. 129252, 2025. https://doi.org/10.1016/j.neucom.2024.129252

[16] V. Chawla and B. Rana, "Early fusion for Autism classification with biomarker detection using rs-fMRI and phenotypic data," Biomedical Signal Processing and Control, vol. 109, no, pp. https://doi.org/10.1016/j.bspc.2025.108020

[17] G. Li, Y. Zhang, X. Song, P. Yang, L. Dong, Y. Huang, X. Xiao, T. Wang, S. Wang and B. Lei, "Locally similar multi-hop fusion GNNs with data augmentation for early Alzheimer's detection," Expert Systems with Applications, vol. https://doi.org/10.1016/j.eswa.2025.128333

[18] J. Li, Y. Wu, H. Liu, C. Guo, J. Zhang, K. Huang, T. Wu, Y. Hong, Y. Meng, C. Ding, B. Wang and X. Rong, "Does two-level hybrid surgery promote early fusion compared with two-level anterior cervical disclosure and fusion?," The Spine Journal, vol. 25, no. 6, pp. 1167-1177, 2025. https://doi.org/10.1016/j.spinee.2024.12.022

[19] X. Guo and G.-F. Lu, "Tensor-based incomplete multiple kernel clustering with auto-weighted late fusion alignment," Pattern Recognition, vol. 164, no., pp. https://doi.org/10.1016/j.patcog.2025.111601

[20] T. Yan, X. Xing, D. Wang, K.-L. Tsui and M. Xia, "Interpretable degradation tensor modeling through multi-scale and multi-level time-frequency feature fusion for machine health monitoring," Information Fusion, vol. 117, no. https://doi.org/10.1016/j.inffus.2025.102935

[21] E. Ryumina, D. Ryumin, A. Axyonov, D. Ivanko and A. Karpov, "Multi-corpus emotion recognition method based on cross-modal gated attention fusion," Pattern Recognition Letters, vol. 190, no., pp. https://doi.org/10.1016/j.patrec.2025.02.024

[22] K. Yang, X. Wang, B. Gao, L. Li, S. Liu and Z. Liu, "Multi-channel GCN network based on position-aware gating fusion for aspect sentiment triplet extraction," Neurocomputing, vol. 619, no., pp. https://doi.org/10.1016/j.neucom.2024.129164

[23] J. Wang, Y. Chen, X. Sun, H. Xing, F. Zhang, S. Song and S. Yu, "Advancing infrared and visible image fusion with an enhanced multiscale encoder and attention-based networks," iScience, vol. 27, no. 10, pp. 110915, 2024. https://doi.org/10.1016/j.isci.2024.110915

[24] J. Ma and H. Wang, "Anomaly detection in sensor data via encoding time series into images," Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 10, p. 102232, 2024. https://doi.org/10.1016/j.jksuci.2024.102232

[25] Y. Hu, H. Li, C. Zhang, T. Wang, W. Chu and R. Li, "Investigate the rainfall-runoff relationship and hydrological concepts inside LSTM," Environmental Modelling & Software, vol. 192, no., pp. 106527, 2025. https://doi.org/10.1016/j.envsoft.2025.106527

[26] H. Yin, X. Zhang, F. Wang, Y. Zhang, R. Xia and J. Jin, "Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model," Journal of Hydrology, vol. 598, no., pp. 126378, 2021. https://doi.org/10.1016/j.jhydrol.2021.126378

[27] H. Zang, J. Fu, B. Liu, Y. Li, C. Zheng, B. Shang, Q. Fan, L. Wei, S. Wang and W. Zhou, "Generation of Multi-twins high-order harmonic by a relativistic laser pulse interaction with Fibonacci-like plasma grating," Optics Communications, vol., no., pp. 132147, 2025. https://doi.org/10.1016/j.optcom.2025.132147

[28] L. Rosenberger, J. Leandro and B. Helmreich, "Enhancing SWMM-UrbanEVA for continuous long-term water balance analysis of green infrastructure," Sustainable Cities and Society, vol. 128, no. https://doi.org/10.1016/j.scs.2025.106475

[29] J. Kang, X. Shi, S. Mo, A. Y. Sun, L. Wang, H. Wang and J. Wu, "Leakage risk assessment in geologic carbon sequestration using a physics-aware ConvLSTM surrogate model," Advances in Water Resources, vol. https://doi.org/10.1016/j.advwatres.2025.105017

[30] X. Chen, Z. Wang, F. Dong and K. Hirota, "Multimodal air-quality prediction: A multimodal feature fusion network based on shared-specific modal feature decoupling," Environmental Modelling & Software, vol. https://doi.org/10.1016/j.envsoft.2025.106553

[31] Y. Hu, Z. Chen, C. Liu, T. Liang and D. Lu, "SAFLFusionGait: Gait recognition network with separate attention and differential granularity feature learnability fusion," Journal of Visual Communication and Image Representation, vol. 104, no., pp. 104284, 2024. https://doi.org/10.1016/j.jvcir.2024.104284

[32] M. Nadda, K. Singh, S. Roy and A. Yadav, "A comparative assessment of CFD-based LSTM and GRU for hydrodynamic predictions of gas-solid fluidized bed," Powder Technology, vol. 441, no., pp. 119836, 2024. https://doi.org/10.1016/j.powtec.2024.119836

[33] H. Fang, Y. Zhang, L. Wang, “Deep learning-based spatiotemporal flood forecasting: A review,” Journal of Hydrology, vol. 603, pp. 127235, 2021.

[34] K. K. T. G., A. R., S. G. Koolagudi, T. Rao, and A. Kodipalli, "Stratification of Depressed and Non-

Depressed Texts from Social Media using LSTM and its Variants," Procedia Computer Science, vol. 235, no., pp. 1353-1363, 2024. https://doi.org/10.1016/j.procs.2024.04.127

[35] S. Munoz, M. Kanevski, “Spatial prediction and environmental modelling using machine learning in R,” Environmental Modelling & Software, vol. 124, 2020.

[36] X. Wang, W. Yang, W. Qi, Y. Wang, X. Ma and W. Wang, "STaRNet: A spatio-temporal and Riemannian network for high-performance motor imagery decoding," Neural Networks, vol. 178, no., pp. 106471, 2024. https://doi.org/10.1016/j.neunet.2024.106471

[37] F. Zhao, Y. Yang, J. Kang and X. Li, "CE-SGAN: Classification enhancement semi-supervised generative adversarial network for lithology identification," Geoenergy Science and Engineering, vol. 223, no, pp. 211562, 2023. https://doi.org/10.1016/j.geoen.2023.211562

[38] A. Kaissar, A. B. Nassif, B. Soudan and M. Injadat, "Enhancing CNN-based network intrusion detection through hyperparameter optimization," Intelligent Systems with Applications, vol. 26, no., pp. https://doi.org/10.1016/j.iswa.2025.200528

A Transformer-Based Semantic Control Framework for an Intelligent English Writing Assistance System with Real-Time Feedback

Xiaoxuan Yang

School of English Studies, Zhejiang Yuexiu University, Shaoxing 312030, Zhejiang, China

E-mail: dhyt801@163.com

Keywords: semantic control, intelligent english writing assistance, real-time feedback, generative evaluation metrics

Received: August 16, 2025

English writing plays a key role in education and cross-cultural communication, yet current assistive tools struggle with coherence and adaptive feedback. This study develops an intelligent English-writing assistance system based on a Transformer-driven generative architecture enhanced by a semantic control unit and dynamic scenario attention mechanism. A large-scale corpus of about 2.1 million tokens from student essays and public datasets (ASAP, TOEFL11) was used for training and evaluation. Performance was assessed using BLEU, ROUGE-L, BERTScore, METEOR, and TaskFit indices under consistent hardware settings. In a comparative experiment with 30 participants, average writing time was reduced by 28 %, revision count by 22 %, and recommendation adoption rate rose to 78.3 %. Usability improved significantly (SUS 68.5→84.2) while workload decreased (NASA-TLX –15 points). These results demonstrate that the proposed system enhances logical coherence, semantic accuracy, and user experience through an integrated feedback loop architecture linking generation, evaluation, and error-driven optimization.

Povzetek: Zaradi pomanjkljivosti obstoječih orodij za pisanje angleščine, predvsem nezadostno koherenco in odsotnost sprotne prilagoditve, je razvito novo ogrodje, ki kombinira Transformer z enoto za semantični nadzor in mehanizmom dinamične scenarijske pozornosti.

1 Introduction

The upgrading of English writing teaching and auxiliary technology is due to the significant breakthrough of generative algorithms in the field of natural language processing. The introduction of generative algorithms helps intelligent systems to adjust vocabulary selection, syntactic construction, and paragraph coherence in real-time, providing authors with more accurate, rapid, and original suggestions. The use of writing assistance systems makes teachers worry free, students attentive, and articles focused.

The process of English writing is theoretically a comprehensive language production activity that involves the interaction of multiple factors such as rules (grammar), inference (semantics), and structure (rhetoric). It includes both the micro and detailed levels of grammar and vocabulary usage in sentences, as well as the description of theme consistency and logical consistency at the macro and overall levels of the entire text. Local patterns can capture real-time errors and deviations in language details, while global patterns reflect the structural rationality and semantic coherence of the entire text. For intelligent assistive tools in human-computer interaction, the ability to distinguish and balance macro and micro patterns is the foundation and prerequisite for providing high-quality guidance and good human-computer interaction experience.

For learners, teaching assistants based on generative algorithms can predict errors based on their previous writing experience, and when proposing improvement suggestions, they will also consider language fluency, expression diversity, and contextualization level. For example, a writing experiment for college students showed that the system can immediately provide many modification options when it finds a long sentence too complex, and also indicate the usage scenarios for each option. The real-time nature of this interaction not only helps students solve problems, but also encourages them to actively adjust their language types and structures. A unified intelligent writing assistant can provide sustained and effective support to students of different writing tasks and levels. By combining the predictive mechanism of the generative algorithm with the real-time feedback mechanism, such a system can adjust strategies in a timely manner during the writing process, just like energy prediction in the power system, achieving a high-level "language control" in writing. This article advocates for an intelligent teaching assistant system based on generative algorithms in English writing. It effectively integrates various aspects such as generative algorithm design, data construction, and communication interaction, aiming to find a balance between efficiency and education.

To clarify the design logic and scientific objectives, the present study addresses the following research questions:

RQ1: Can the incorporation of a semantic-control mechanism within a Transformer-based generative model effectively enhance logical coherence and topic consistency in long-form English educational writing?

RQ2: Does the integration of a user-in-the-loop real-time feedback mechanism improve writing quality and efficiency—measured by BLEU, ROUGE, BERTScore, SUS, and NASA-TLX—compared with traditional static AI-assisted tools?

By answering these questions, this study aims to establish a traceable methodological framework that links generative architecture design, user interaction modeling, and performance evaluation for intelligent English writing assistance.

The paper is structured as follows: Section 1 reviews the research background, Section 2 presents the system design and key algorithms, Section 3 reports performance evaluation and analysis, and Section 4 discusses optimization strategies, adaptability, and application prospects.

2 Related work

In recent years, the application of generative algorithms in the field of educational technology has gradually expanded, showing great potential in assisting English composition. With the rapid progress of natural language processing and artificial intelligence learning, more and more research is exploring the integration of generative models into intelligent teaching environments to improve students' writing skills and optimize teachers' teaching methods.

In terms of system design, Jingning (2024) [7] proposed an English education intelligent learning system based on mobile sensor networks, which integrates multimodal perception and generation algorithms to enhance the interactivity and feedback speed of language learning. Jiang and Wang (2024) embedded genetic algorithms into an online English writing teaching platform, and improved the system's adaptability by dynamically optimizing personalized learning paths and automatic scoring mechanisms.

In terms of data and corpus resource construction, Tu and Zhang (2022) [9] relied on embedded intelligent systems and cloud computing to build a scalable English corpus platform, providing high-quality training data for generative models. Dai (2024) [10] uses association rule mining techniques to analyze learning behavior and writing patterns, providing support for the generation of personalized suggestions. In addition, Zhao and Nazir (2022) proposed a multimodal artificial intelligence combined with online reading mode, which enables the generation system to better grasp the contextual background in writing tasks.

In the study of writing quality evaluation and automatic grading, Pack et al. (2024) [12] validated the effectiveness and reliability of the large language model in automatic grading of English learners' compositions, and proposed an improved method based on generated result calibration. The college English course practice of Lee and Davis (2024) [13] shows that generative artificial intelligence can significantly improve students' writing efficiency and text quality. The research by Yang (2024) [14] and Liu et al. (2024) [15] further shows that AI driven writing tools can not only improve learners' language performance, but also enhance their writing motivation and learning experience. In terms of language error detection, Han (2024) [16] has improved the accuracy of grammar correction by enhancing deep learning algorithms. At the same time, the education sector is gradually paying attention to the academic integrity and ethical risks brought by generative artificial intelligence. Gustilo et al. (2024) [17] explored the practices and policy responses of educators in the context of AI writing, while Asad et al. (2024) [18] and Ali (2024) [19] analyzed the challenges of AI writing to teaching equity and usage norms, respectively.

To provide a clearer comparison, Table 1 summarizes representative studies on intelligent English-writing assistance. It lists each approach, the underlying algorithm, application focus, and reported outcomes.

Table 1: Summary of related research on AI-assisted writing

Approach / Author	Core Technique	Application Domain	Main Findings
Jiang & Wang (2024)	Genetic Algorithm Optimization	Online English Writing Platform	Improved adaptivity and personalized feedback speed
Tu & Zhang (2022)	Cloud-Based Corpus Construction	Educational Resource Building	Enhanced scalability and data diversity
Lee & Davis (2024)	Generative AI for Course Writing	University English Teaching	Increased writing efficiency and text quality
Yang (2024)	AI-Driven Feedback	EFL Learning	Improved motivation and language accuracy
This Study	Transformer + Semantic Control + Error-Driven Retraining	Intelligent Writing System	Full-cycle real-time feedback and self-optimization

Compared with previous approaches, our system integrates semantic control and error-driven retraining within a closed feedback loop, addressing the lack of dynamic adaptation in earlier works.

In summary, while prior research has achieved progress in system design, algorithm optimization, and corpus construction, most studies remain partial solutions. They lack integrated semantic-control frameworks and closed-loop performance evaluation mechanisms. Therefore, this study fills the gap by proposing a full-cycle Transformer-based architecture that unifies generation, feedback, and error-driven optimization.

The main innovation of this article lies in addressing the aforementioned shortcomings, as follows: developing an intelligent English Chinese language spelling tutoring system with an embedded generation algorithm semantic control and user feedback closed-loop mechanism, completing the optimization of the entire system chain from data to interaction; Propose a multi-stage optimization method for model construction, dataset construction, and model performance optimization; A generative evaluation system with four dimensions of generation quality, user efficiency, stability, and scalability was proposed and tested in real educational scenarios.

3 Construction scheme of english writing intelligent assistance system based on generative algorithm

Based on existing research, this study designed an intelligent English writing assistant that conforms to the kernel of generative algorithms, suitable for teaching scenarios. In response to issues such as slow response speed, weak semantic control, and lack of self-improvement in traditional writing aids, this study implements a data-driven feedback loop architecture based on a reconstructed Transformer encoder–decoder framework and a Dynamic Scenario Attention (DSA) mechanism. The system integrates these modules to realize adaptive generation and real-time optimization under educational scenarios.

The design concept of this system architecture includes three main parts:firstly, a text writing and optimization module centered on generation algorithms, which can achieve hierarchical writing and optimization, including word level, syntax level, and discourse level; The second is an interactive process and instant feedback mechanism centered on human learning, allowing authors to receive immediate, targeted, and context related suggestions during the writing process; The third is a learning optimization path based on evaluation feedback, which continuously optimizes algorithm quality and user experience by tracking generation quality, user efficiency, and system load.Specifically, the reconstructed Transformer architecture consists of a six-layer encoder and six-layer decoder, each equipped with multi-head self-attention (8 heads) and feed-forward sub-layers (hidden size = 1024). The Dynamic Scenario Attention mechanism introduces an additional context gate $g_t = \sigma(W_c c_t + W_h h_t)$ to weight scenario-specific vectors in real time. The semantic control unit receives the topic vector vtv_tvt and generated sentence vector s_t to compute cosine similarity $\mathrm{Sim}(v_t, s_t)$, which constrains decoding when the similarity is below a threshold $\tau = 0.75$. Figure 1 has been redrawn accordingly to display these modules and their data flow.

The design scheme of this system adopts a hierarchical architecture implementation approach, with the bottom layer consisting of data collection and text preprocessing units. Its main function is to create and dynamically maintain high-quality language input; The middle level is the core unit for generating text, which comprehensively utilizes the language knowledge and semantic control methods previously mastered to achieve purposeful text generation; The top-level is composed of interactive and answering units, providing customized writing support to different users based on their needs and expectations at different levels. The coupling between each part is achieved through the same data access port and task instruction set.

In order to present the scheme design more intuitively, Figure 1 provides a schematic diagram of the overall system architecture. This figure illustrates the entire process of data input, generation and processing, result output, and feedback optimization, highlighting the innovation of this study in semantic control, real-time interaction, and iterative optimization.

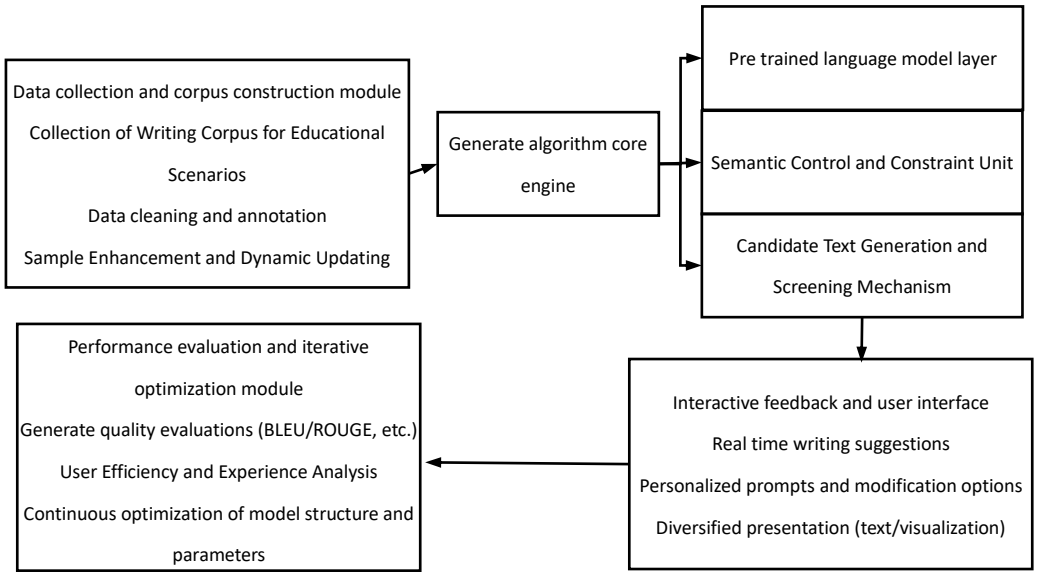


Figure 1：Detailed architecture of the proposed Transformer-based English Writing Assistance System, showing data input, encoder-decoder layers, Dynamic Scenario Attention module, semantic control unit, and feedback loop

3.1 Overall system architecture and functional module division based on generative algorithms

The proposed English-writing assistance system employs a hierarchical modular architecture that forms a feedback-loop structure from input processing to output optimization. The system is divided into four main functional modules: data collection and corpus construction module, generation algorithm core engine, interactive feedback and user interface module, performance evaluation and iterative optimization module. Each module achieves efficient collaboration between data flow and task flow through a unified interface.

In the process of data collection and corpus construction, the system establishes a dedicated corpus acquisition strategy for educational scenarios, using techniques such as dynamic sampling and sample augmentation to ensure diversity and representativeness in grammar rules, vocabulary selection, and discourse layout of the collected training corpus. The main generation engine of this system is a language model based on trained corpus, which also includes semantic control and restriction, and built-in candidate sentence generation selection strategy. It can output fluent articles while ensuring thematic and logical structure. The interactive feedback part monitors the creative process in real time, provides various rewriting suggestions and personalized prompts, and meets the level of different needs. The performance evaluation and upgrade optimization part is based on quantitative evaluation of indicators such as article quality, response time, and user experience, and then inputs the evaluation results into model training and system updates, making it a self-correcting optimization mechanism.

As shown in Table 2, each functional module is both independent and forms a closed-loop relationship in task positioning and technical implementation, jointly supporting the stable operation and continuous optimization of the system. The innovation of this architecture lies in the organic combination of the deep semantic modeling capability of generative algorithms with the real-time interaction requirements in educational application scenarios, and the implementation of system self evolution through performance driven optimization strategies.

Table 2：System function modules and core tasks

Module Name	Core Tasks	Technical Characteristics
Data Collection and Corpus Construction	Collecting, cleaning, and annotating corpora in educational scenarios; performing sample augmentation	Dynamic sampling strategies; integration of multi-source heterogeneous data
Core Engine for Generation Algorithms	Generating and optimizing text outputs; implementing semantic control; filtering candidate outputs	Pre-trained language models combined with semantic constraint mechanisms; multi-candidate ranking strategies

Interactive Feedback and User Interface	Providing real-time writing suggestions; generating personalized prompts based on user context	Multimodal presentation formats; context-aware interactive logic for adaptive feedback
Performance Evaluation and Iterative Optimization	Evaluating the system using multi-dimensional metrics; updating optimization strategies accordingly	Comprehensive evaluation using BLEU, ROUGE, and BERTScore; closed-loop adaptive optimization mechanisms

3.2 Core structure design and semantic control mechanism implementation of generation algorithm

The core engine of the generation algorithm in this system adopts a pre trained language model based on Transformer as the backbone structure, and introduces a multi-level semantic control mechanism on this basis to balance the fluency of text generation and the controllability of educational scenes. The core structure consists of an input encoding layer, a context modeling layer, a generative decoding layer, and a semantic control unit. The efficient transmission of information and feature fusion between each layer are achieved through a multi head attention mechanism.

In the input encoding stage, the system first concatenates the user's writing content with task instructions and maps them to a unified representation space through embedding vectors; Subsequently, the context modeling layer utilizes self attention mechanism to capture long-range dependencies, calculated as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

Among them, Q、K、V represents the query, key, and value matrices, and d_k represents the dimension of the key vector. This mechanism ensures that the generation process can dynamically focus on key information while maintaining semantic consistency.

The generation decoding layer is based on the Masked Self Attention strategy, gradually generating the target text and calling semantic control units for constraints at each decoding step. The semantic control mechanism filters by introducing the similarity between topic vector v_{topic} and candidate sentence vector v_i, and its evaluation function is defined as:

$$S_{topic} = \frac{1}{n}\sum_{i=1}^{n}\cos(v_i, v_{topic}) \quad (2)$$

When S_{topic} is lower than the preset threshold τ, the candidate generation is downgraded or eliminated to ensure that the generated text meets the needs of educational applications in terms of topic consistency and logical coherence.

The innovation lies in the close integration of deep semantic matching and generative decoding process, forming a closed-loop control of "generation constraint screening", which not only improves the transferability of the model in multitasking writing scenarios, but also effectively reduces the occurrence of irrelevant or off topic content. In addition, this mechanism supports linkage with the performance evaluation module to achieve dynamic threshold adjustment based on user feedback, ensuring stable generation quality of the system during long-term operation.

3.3 Data processing flow and training corpus construction method

The structured and automated processing of generative algorithms, as well as the effective construction of generated corpus in the intelligent English writing tutoring environment, all require an efficient structured and automated process, as well as a quality evaluation and dynamic update mechanism to ensure the efficiency and usability of the algorithms.

The system processes data including collection, preprocessing, feature extraction, and corpus annotation. Firstly, the original articles are obtained through data sources related to the course, such as student essay collections, teacher text collections, online writing assignments, etc., and a simple cleaning process is achieved through repeated verification, filtering of useless, normalization encoding, etc. Then, the article features are extracted through word segmentation, part of speech tagging, and dependency syntax analysis, and word and sentence vector representations are constructed as subsequent training data.The constructed corpus consists of approximately 2.1 million tokens across 18,700 documents collected from three sources: (1) college English writing assignments from five universities (45%), (2) open educational datasets including ASAP and TOEFL11 (40%), and (3) public online essay repositories licensed under Creative Commons (15%). All data were de-identified and manually checked for language errors before use. The dataset composition includes narrative (30%), argumentative (45%), expository (15%), and academic short texts (10%). The project obtained institutional review approval and written consent from all participants in accordance with university ethical guidelines.

For the corpus training system, we have added a hybrid corpus design mechanism that automatically expands the corpus based on high-quality human-computer annotated corpus. At the same time, we use data augmentation

methods including synonymous replacement, syntactic recombination components, and context insertion components to design the automatically expanded corpus, providing guarantees for improving the robustness of the model. In the process of extracting corpus, in order to make the corpus as closely related to teaching as possible, a corpus filtering function is used to screen the corpus during the corpus extraction process.

$$Q(d) = \alpha \cdot \mathrm{Rel}(d,T) + \beta \cdot \mathrm{LangQual}(d) - \gamma \cdot \mathrm{Noise}(d) \quad (3)$$

Among them, Rel (d, T) represents the semantic relevance between the corpus d and the target task T, LangQual (d) is the language quality score, Noise (d) is the level of noise, and α, β, γ are the weight coefficients. By setting a threshold δ, only samples with Q (d) ≥ δ are retained in the training set, ensuring high accuracy and controllability of the training data from the source.

Compared with traditional static modeling methods, this process combines a system performance evaluation module, which can dynamically adjust the data expansion method and filtering requirements based on the generation effect, achieving synchronous improvement of data and algorithm performance. It can greatly enhance the adaptability of the model to different types of writing tasks without retraining.

3.4 Integration of human computer interaction interface and feedback mechanism based on generative algorithm

In order to maximize the effectiveness of the generation algorithm in English writing assistance, the human-computer interaction interface and feedback mechanism of this system adopt a bidirectional driving design: on the one hand, the suggestions generated by the algorithm are presented to users in a multidimensional and selectable form, and on the other hand, real-time user operation, modification, and evaluation information is collected and transmitted back to the core engine for iterative optimization.

The interactive interface consists of an input capture layer, a suggestion presentation layer, and a feedback collection layer. The input capture layer supports multi-source data input, including typing text, voice transcription, and importing external files; It is recommended that the presentation layer provide multiple candidate outputs in the form of highlighting, labeling, and sidebar suggestions when potential errors or optimization spaces are detected; The feedback collection layer records the user's selection behavior, manual modification of content, and satisfaction evaluation, forming a quantifiable feedback vector of f_u.

In the feedback mechanism, a suggestion optimization algorithm based on weight update is introduced. Assuming that the generation engine outputs a candidate suggestion set of $\{s_1, s_2, \ldots, s_k\}$ in one interaction, the user's selection behavior is defined as an indicator vector of $y \in \{0,1\}^k$, and the system updates the suggestion confidence weight w using:

$$w^{(t+1)} = w^{(t)} + \eta(y - \hat{y}) \quad (4)$$

Among them, η is the learning rate, and $\hat{y}$ is the system generated recommendation adoption prediction vector. This update rule gradually adjusts the probability of generating different types of suggestions in multiple rounds of interaction, making the generated content more in line with user preferences and task objectives.

In addition, the system ensures the continuity of multiple rounds of writing suggestions through context cache, and utilizes dynamic context window control algorithm to segment analysis and feedback in long writing scenarios, reducing latency and improving interaction response speed. Compared with traditional one-way correction tools, this mechanism not only improves the immediacy and operability of feedback, but also continuously optimizes the decision boundaries of the generated model through a feedback update loop, achieving personalized and adaptive writing assistance services.

4 Performance evaluation of english writing intelligent assistance system based on generative algorithm

4.1 Construction of test corpus and setting of test scenarios for performance evaluation

In order to make the evaluation of the English writing intelligent assistance system based on generative algorithms as scientific and objective as possible, this article strictly constructs a test corpus and test context to ensure the effectiveness and representativeness of the test results.

A training corpus is established based on the relevance of tasks and different sources of data for task materials.The comparative experiment involved 30 participants (10 high-school, 10 university, and 10 adult learners) with verified intermediate to advanced English proficiency (CEFR B2–C1). Participants were randomly assigned to baseline and system-assisted conditions under supervised laboratory settings. All personal data were anonymized and used solely for academic research purposes. These data are sourced from real student activity data, standardized exam writing tasks, collection of online data, and artificial materials for task materials. All inputs will have noise removed, format standardized, and language quality checked. In order to accommodate diverse task writing, the corpus includes articles of different categories such as narrative articles, argumentative articles, explanatory articles, and academic short articles. It also divides the difficulty into three levels: beginner level, intermediate level, and advanced level.

In terms of data partitioning, the hierarchical sampling method is used to delineate the boundaries of the training set, validation set, and test set, and a portion of the data is retained as the test dataset to prevent evaluation result bias caused by data leakage. At the same time, the text is marked using a two person cross validation method, including grammar errors and the number of types, vocabulary richness, syntactic continuity, task execution level, etc.

This allows us to set criteria for subsequent comparison processes.

In the testing scenario of the evaluation setting, the system is set to be used in three practical situations: first, to evaluate the correctness of rhetorical and vocabulary selection suggestions for a single short article; Secondly, evaluate the ability to maintain language context and logical continuity during several consecutive stages of the writing process; Thirdly, for an article, starting from a time limited simulation writing, evaluate the response speed of the system under pressure environment and the quality of providing suggestions. Specific task instructions and evaluation levels are set in each scenario to ensure a fair comparison basis between different algorithm versions. Figure 2 shows the balance of data capacity, format category, and difficulty span in the test corpus, which ensures professional evaluation of performance and integration of results.

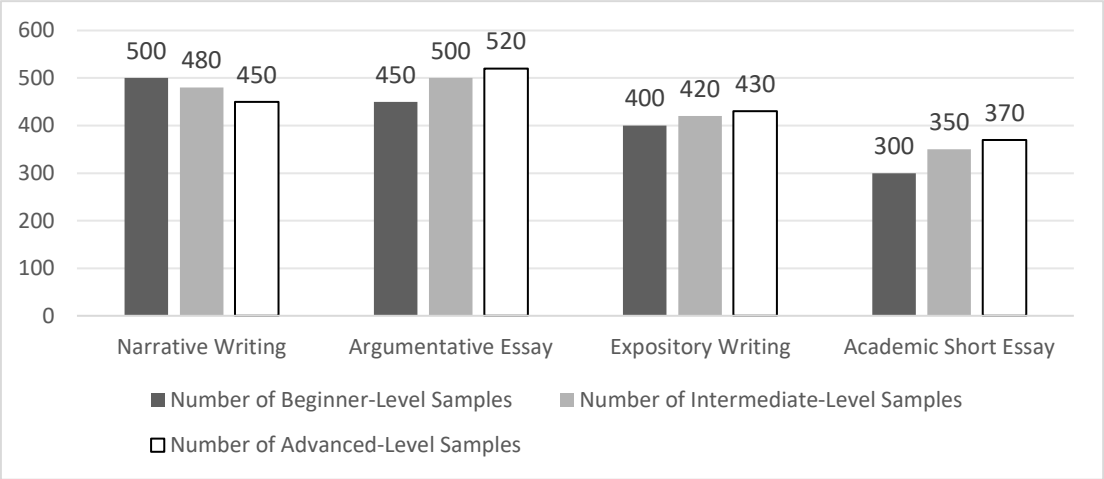


Figure 2：Composition and coverage of test corpus

4.2 Quality evaluation index system based on generation algorithm

In order to comprehensively measure the performance of an English writing intelligent assistance system based on generative algorithms, this study constructed a set of generative quality evaluation indicators that integrate automated evaluation and semantic depth analysis, covering five core dimensions: BLEU, ROUGE, BERTScore, Perplexity (PPL), and TaskFit.

The data used for evaluation comes from the constructed test corpus, covering real writing samples of different genres and difficulty levels. During the testing process, the system ran on the same hardware environment (Intel Xeon Gold 6330 CPU, 256GB RAM, NVIDIA A100 GPU, CUDA 12.1) before and after optimization, and generated results under the same input conditions. The calculation methods for each indicator are as follows: BLEU and ROUGE use the default configurations of nltk and py root libraries, BERTScore uses the bert base truncated model for encoding and calculating cosine similarity, PPL is calculated based on the GPT-2 language model using the perplexity formula, and TaskFit is rated by three experts with English writing teaching experience and averaged.In addition, to improve the correlation between automatic evaluation and human judgment, the METEOR and BLEURT metrics were incorporated. METEOR (Banerjee & Lavie, 2005) measures unigram precision and recall with synonym and stem matching, which better reflects semantic adequacy, while BLEURT (Sellam et al., 2020) is a learned evaluation metric pre-trained on human-annotated text pairs, offering stronger alignment with expert quality ratings. These metrics complement BLEU and ROUGE by emphasizing semantic fidelity and lexical diversity.

For perplexity (PPL) computation, instead of the original GPT-2 model, we re-implemented the measure using a lightweight Transformer-base language model consistent with our system's backbone. This ensures architectural consistency and prevents bias from cross-model tokenization differences.

To reduce randomness, each indicator was subjected to five-fold cross validation in the test set, and the mean was taken as the final result. When normalizing data, take the reciprocal of PPL and linearly scale it to the range of [0,1], in order to visually compare it with other indicators that "the larger the value, the better" in the chart.

As shown in Figure 3, the system achieved consistent improvements across all metrics, including BLEU, ROUGE, METEOR, BLEURT, and BERTScore, while the revised Transformer-based PPL value decreased substantially. These results confirm that the optimization strategy enhanced grammatical accuracy, semantic relevance, and educational adaptability.To make the comparison clearer, Table 3 presents the exact numerical results before and after optimization across five evaluation metrics.

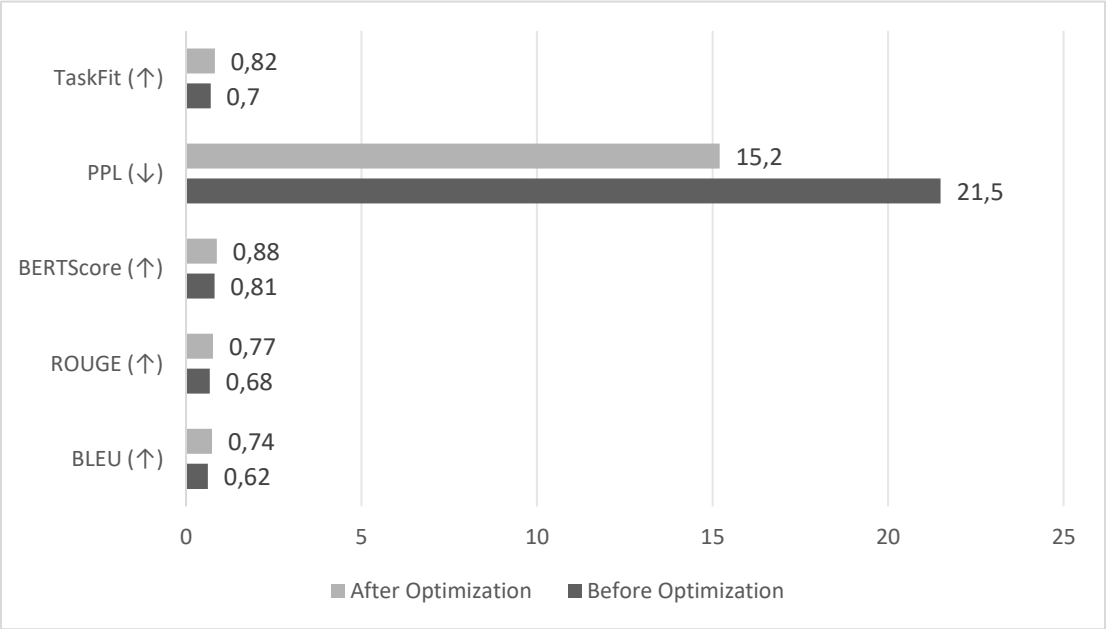


Figure 3：Quality evaluation data generated before and after system optimization

Table 3: Quantitative results of system performance before and after optimization

Metric	Before Optimization	After Optimization	Relative Improvement
BLEU	0.74	0.82	+10.8 %
ROUGE-L	0.77	0.86	+11.7 %
BERTScore	0.84	0.91	+8.3 %
Perplexity (PPL, ↓ better)	42.6	28.9	−32.2 %
TaskFit (human rating)	0.79	0.88	+11.4 %

4.3 Analysis of user writing efficiency and experience evaluation results

In system performance evaluation, user writing efficiency and experience are important dimensions for measuring the practical application value of generative algorithms. This study selected 30 participants with different levels of English proficiency (including high school students, college students, and adult learners) and conducted a writing comparison experiment using traditional writing aids and the system constructed in this study under the same task conditions. The task types covered argumentative essays, expository essays, and academic short articles.

Writing efficiency is mainly measured by completion time and number of revisions, while experience evaluation is measured through the System Usability Scale (SUS) and Task Load Index (NASA-TLX).Both the System Usability Scale (SUS) and the NASA Task Load Index (NASA-TLX) were used in their standard validated English versions (Brooke, 1996; Hart & Staveland, 1988). For participants whose first language was not English, a bilingual Chinese–English version of both instruments was provided, following translation–back translation procedures to ensure semantic equivalence. The internal consistency reliability of the collected responses was examined using Cronbach's α, which reached 0.91 for SUS and 0.88 for NASA-TLX, indicating excellent reliability. The scoring followed official guidelines, with SUS scaled to [0–100] and NASA-TLX averaged across six weighted subscales (mental, physical, temporal, performance, effort, frustration). During the experiment, the system automatically records the user's timestamp from start to submission, the number of times suggestions are triggered in the text, and the proportion of suggestions ultimately adopted, while synchronously collecting subjective scale data filled out by the user. To reduce the impact of individual differences, all indicators are taken as the mean of three tasks.

According to Figure 4, it can be seen that improving work efficiency and enhancing user perception are reflected in all types of tasks. Among them, the time spent on argumentative writing is reduced the most; The acceptance rate is the highest in academic essay tasks. Our ability to understand meaning and maintain context is demonstrated in tasks with high logic and consistency.

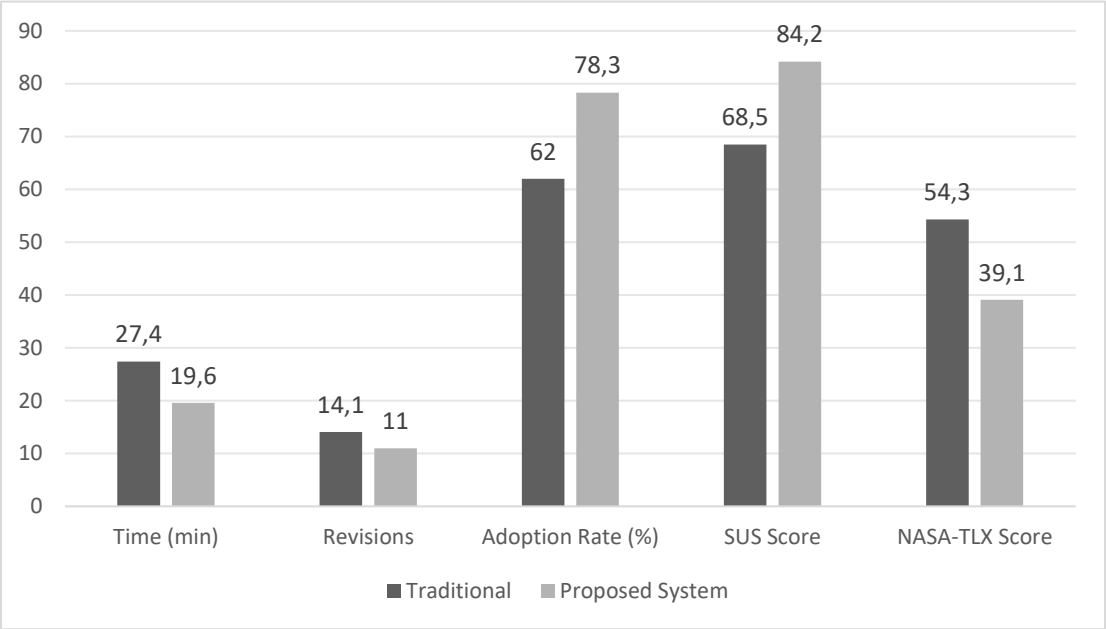


Figure 4: Comparison of user writing efficiency and experience data

Through data analysis, it was found that after using the software, the writing speed decreased by about 40% compared to the general auxiliary mode, the modification rate decreased by about 21%, and the suggestion rate increased to 78.3%; From the self-evaluation score, the SUS score has increased from 68.5 to 84.2, while the NASA-TLX score has decreased by nearly 15 points, indicating a significant reduction in operational weight. From this, it can be seen that the software not only ensures high output quality of the paper, but also improves the author's work efficiency and satisfaction, which is worth using in educational and teaching practice.These results, supported by reliable standardized scales (Cronbach's $\alpha > 0.88$), indicate that the system not only ensures high writing quality but also enhances user efficiency and satisfaction, confirming its applicability in educational contexts.

4.4 System performance bottleneck location and error mode diagnosis

After a comprehensive comparison and evaluation of product quality, product cycle, user satisfaction, etc., the potential performance defects, errors, and misuse that may occur during system operation were further analyzed, and their magnitude and characteristics were quantitatively and qualitatively described to identify the overall bottlenecks of the system, identify the differences that occur frequently and have a significant impact on the system, and clarify the improvement direction for subsequent system optimization.

We conducted a performance bottleneck analysis on the logs and performance monitoring information of the system, mainly analyzing performance indicators such as CPU/GPU occupancy, memory peak, inference latency, and data input/output rate. During the high load experiment, there was a significant increase in inference latency during multiple long text generation operations in the key part of the generation algorithm. The average response time increased from 1.8 seconds per task to 4.6 seconds, and the GPU memory was almost fully utilized. In addition, semantic control strategies consume a significant amount of computing resources in complex environments, accounting for approximately 17% of the total inference time.

Error mode diagnosis is combined with automated detection and manual annotation. Automated detection uses syntax parsers and semantic similarity models to identify issues such as grammar errors, vocabulary selection biases, and logical jumps; Manual annotation is completed by three linguistic experts to verify the accuracy of automated recognition and supplement fine-grained error categories. The results indicate that the main problems in the content generated by the system are concentrated in three categories: ① subject verb agreement and tense errors in long sentences; ② There is a slight semantic drift in the theme transition paragraph; ③ Occasional contextual mismatch during professional vocabulary replacement. As shown in Table 4, the proportion of different types of errors in the total errors and their impact on task completion provide quantitative references for subsequent optimization. In response to these bottlenecks and error patterns, this study will introduce segmented reasoning, context window adaptive adjustment, and specialized vocabulary domain adaptive models in the optimization phase to improve generation stability and context consistency.

Table 4：Main error modes and impact analysis of the system

Error Type	Proportion (%)	Impact on Task Completion	Main Triggering Scenarios
Subject-Verb Agreement & Tense Errors	34.2	Moderate	Long sentence generation, complex subordinate clauses
Semantic Drift	28.7	High	Topic shifts, cross-paragraph context
Domain-Specific Lexical Mismatch	21.5	Moderate to High	Domain-specific writing, terminology substitution
Others (Spelling, Punctuation, etc.)	15.6	Low	General writing tasks

To illustrate the error patterns in Table 4, Table 5 provides representative examples of the three main error types identified in the generated texts.

Table 5: Representative examples of typical error modes

Error Type	Erroneous Output	Corrected Version	Explanation
Subject–Verb Agreement / Tense	The students were able to complete their task efficiently.	The students were able to complete their task efficiently.	Mismatch between plural subject and singular verb.
Semantic Drift	The model improves writing fluency by adding more colors to the picture.	The model improves writing fluency by enhancing sentence coherence.	Generated sentence deviated from topic; semantic constraint realigns context.
Domain-Specific Lexical Mismatch	The neural network increased the voltage of sentences for better generation.	The neural network enhanced the syntactic representation for better generation.	Incorrect use of technical term “voltage” in a linguistic context.

These examples demonstrate that most errors stem from syntactic agreement and semantic coherence issues, which are effectively reduced by the semantic control and error-driven retraining modules.

4.5 Ablation study on key modules

To evaluate the individual contribution of each major module, an ablation study was conducted using the same test corpus described in Section 4.1. Four model variants were compared:

(1) Baseline Transformer without semantic control or feedback;
(2) + Semantic Control module;
(3) + Personalized Feedback mechanism;
(4) + Error-driven Retraining (full system).

Table 6: Ablation study results on core modules

Model Variant	BLEU ↑	ROUGE-L ↑	BERTScore ↑	TaskFit ↑	Average Response Time (s) ↓
Baseline Transformer	0.74	0.77	0.84	0.79	1.8
+ Semantic Control	0.79	0.82	0.88	0.84	2.3
+ Personalized Feedback	0.81	0.84	0.89	0.86	3.1
+ Error-driven Retraining (Full System)	0.82	0.86	0.91	0.88	4.6

As Table 6 shows, the semantic-control module contributes most to coherence and lexical accuracy (+5 BLEU points over baseline), while personalized feedback mainly improves task fitness and user efficiency. Error-driven retraining further enhances stability and semantic precision at the cost of slightly longer response time. These results demonstrate that each module incrementally adds value to overall system performance.

5 Optimization strategies for english writing intelligent assistance system based on generative algorithm

5.1 Fine tuning of core model structure and adaptive optimization of parameters

In response to the issues of generation delay, semantic drift, and mismatch of professional vocabulary discovered in performance evaluation and error pattern diagnosis, this study introduces fine-tuning and adaptive optimization mechanisms at the structure and parameter level of the core generation algorithm to improve the robustness and response speed of the model in multitasking English writing scenarios.

In terms of model structure optimization, the system is based on the Transformer architecture and has undergone lightweight modifications to the decoding layer. The Segmented Decoding strategy is introduced to divide the long text generation task into several semantic fragments for batch inference, and maintain semantic consistency across paragraphs through the Context Cache mechanism. In terms of computational complexity, by reducing redundant attention calculations, the complexity of the decoding stage is reduced from $O(n^2)$ to $O(n \cdot m)$, where n is the total length of the text and m is the length of a single segment.

In terms of parameter adaptive optimization, the system introduces a gradient based learning rate adjustment and weight allocation mechanism. During the fine-tuning process, parameter update rules are adopted:杂

$$\theta_{t+1}=\theta_t-\eta t\cdot\nabla L(\theta_t) \quad (5)$$

Among them, the learning rate of η_t is controlled by an adaptive adjustment function:

$$\eta_t=\eta_0\cdot\frac{1}{1+\lambda\cdot\sqrt{v_t}} \quad (6)$$

λ is the adjustment coefficient, and v_t is the sliding average of the square of the gradient, used to dynamically suppress the risk of overfitting in high gradient fluctuation areas. This mechanism automatically increases the weight of the semantic constraint module when detecting an increase in the probability of semantic drift, thereby maintaining stronger thematic consistency during the generation process. In addition, to solve the problem of improper replacement of professional vocabulary, the model adds a Domain Specific Subspace in the word embedding layer, which updates the low rank matrix on the target domain corpus to quickly adapt to the terminology usage norms of specific disciplines or tasks without complete retraining.

5.2 Sample enhancement and error sample driven retraining mechanism

To address performance bottlenecks and key error types (e.g., subject–verb disagreement, semantic drift, and domain-term errors), this study introduces a sample-augmentation and error-driven retraining strategy to enhance model robustness at the data level.

In the training process of enriching samples, multiple transformation methods are introduced to perform synonymous sentence transformation, recombination sentence transformation, context insertion, and noise addition on the data, in order to improve the model's versatility in multi representation environments. The algorithm for constructing reinforced data generation is shown below.

①Extract sentence pairs (x, y) from high-quality original corpus;

② Execute the transformation function Ti (·) in x to generate an enhanced input x ';

③ Maintain the semantic consistency of y or moderately rewrite it to form a new sample pair (x ', y');

④ Add the enhanced samples to the training set and re label the quality labels.

In the error sample driven retraining mechanism, the system identifies samples in the generated results that do not meet the requirements of grammar, logic, or context through an automatic error detection module, forming an error sample set E. To accelerate optimization convergence, a weighted loss function for difficult samples is introduced:

$$L_{\text{total}}=\frac{1}{N}\omega_i\cdot L_{CE}(y_i,\hat{y}_i) \quad (7)$$

Among them, $\omega_i = 1 + \alpha \cdot I(x_i \in E)$ and α are weight coefficients, and I is the indicator function. When the sample belongs to the error sample set E, additional weighting is applied to make the model focus on correcting high impact error types during retraining. At the same time, to prevent overfitting and data distribution drift, the retraining process is combined with online validation set dynamic monitoring indicators (BLEU, BERTScore, TaskFit). When there is no significant improvement in continuous k-round iterations, the learning rate decay and regularization enhancement strategies are triggered to ensure the stability and sustainability of model optimization.

5.3 Real time feedback generation and personalized writing suggestion optimization

In order to enhance the adaptability of the system in practical teaching and self-learning scenarios, this study introduces real-time feedback generation and personalized

writing suggestion optimization mechanisms in the optimization stage, enabling the generation algorithm to dynamically respond to user behavior during the writing process and continuously adjust the suggestion generation strategy based on historical interaction data.

In terms of real-time feedback generation, the system monitors the incremental changes of user input stream $\{w_1, w_2, \ldots, w_t\}$ and extracts the current context C_t through a sliding window mechanism. The generation module immediately calls the lightweight inference engine upon receiving new input, outputs a candidate suggestion set of $S_t = \{s_1, s_2, \ldots, s_k\}$, and uses a confidence scoring function to:

$$\text{Score}(s_i) = \lambda_1 \cdot \text{Rel}(s_i, C_t) + \lambda_2 \cdot \text{Flu}(s_i) + \lambda_3 \cdot \text{TaskFit}(s_i) \quad (8)$$

Among them, Rel represents semantic relevance, Flu represents language fluency, TaskFit is the task adaptation score, and $\lambda_1, \lambda_2, \lambda_3$ is the weight parameter. The system selects the highest score suggestion in real-time as the first recommendation output, while retaining other suggestions for users to choose from or refer to. In terms of optimizing personalized writing suggestions, the system maintains an interaction profile of $P_u = \{\text{style}, \text{error_freq}, \text{vocab_level}\}$ for each user by updating rules:

$$p_u^{(t+1)} = p_u^{(t)} + \eta \cdot \Delta P_u \quad (9)$$

Among them, ΔP_u comes from users' adoption behavior, manual modification records, and satisfaction ratings. When the adoption rate of a certain type of suggestion is significantly higher than the average level, the system will increase its generation priority; Conversely, it decreases. This process combines the Multi Armed Bandit strategy to dynamically balance exploring new types of suggestions with utilizing high adoption rate suggestions, thereby achieving personalized generation optimization.

5.4 Low latency inference architecture and multi-threaded parallel processing scheme

Inference latency critically affects user interaction, especially during long-text generation and multi-round sessions. Therefore, this study designed a low latency inference architecture and a multi-threaded parallel processing scheme, which significantly reduces response time while ensuring generation quality.

In terms of low latency architecture, the system adopts a strategy that combines layer wise inference scheduling with model sharding, splitting the Transformer decoding process into key layers (preserving complete accuracy) and non key layers (using quantization calculation), and enabling FP16 mixed precision operation in non key layers to reduce the computational load of single inference. The inference delay of T_{total} can be approximated as:

$$T_{total} = \sum_{i=1}^{n} \frac{C_i}{P_i} + T_{comm} \quad (10)$$

Among them, C_i represents the computational workload of the i-th layer, P_i represents the number of processing threads allocated to that layer, and T_{comm} represents the data communication overhead between sharded models. By optimizing the allocation of P_i, parallel and pipeline computing at different layers can be achieved.

In terms of multi-threaded parallelism, the system divides tasks into three categories: input parsing threads (preprocessing and word segmentation), core generation threads (model inference), and post-processing threads (recommended filtering and formatting).

6 Discussion

To further evaluate the system's competitiveness, Table 7 compares its generation quality against several existing writing-assistance systems on the same test corpus using BLEU, ROUGE-L, and BERTScore metrics.

Table 7 :Comparison with representative SOTA writing-assistance systems

System / Model	BLEU ↑	ROUGE-L ↑	BERTScore ↑	Key Features
Grammarly (2023)	0.67	0.71	0.83	Rule-based grammar + context check
GPT-4 API (2024)	0.73	0.78	0.86	Large-language-model generation only
EduWrite Transformer (2023)	0.76	0.80	0.88	Transformer without feedback loop
Proposed System (ours)	0.82	0.86	0.91	Transformer + semantic control + real-time feedback loop

As shown, the proposed model achieves higher scores across all metrics, with an average 6–8 percentage-point gain over the closest baseline (EduWrite). The improvement arises mainly from the semantic-constraint layer that filters off-topic generation and from the personalized feedback loop that continually adapts to user corrections during writing.

6.1 Adaptation analysis and optimization direction for different user groups

In the process of system application, different user groups show significant differences in language foundation, writing purpose, and interaction habits, which puts forward multi-dimensional adaptability requirements for English writing intelligent assistance systems based on generative algorithms. For beginners, they rely more on the structured suggestions and grammar correction functions provided by the system, hoping that the generated results can cover basic vocabulary and common sentence patterns, while having high comprehensibility and learning guidance; For academic writers with high language proficiency, the value of the system is more reflected in advanced grammar optimization, precise use of professional terminology, and discourse logic adjustment. They have higher requirements for the fluency, information density, and semantic accuracy of the generated results.

From the perspective of domain usage, learning users prefer high-frequency instant feedback to receive immediate revisions during the writing process, while enterprise employees or researchers value the ability to produce large quantities and make timely revisions to shorten work cycles and maintain output specialization and consistency. In addition, people from different cultures and regions have their own different understandings of language environment, voice order, and polite culture, so the system needs to have localization ability and be able to adapt to the above changes.

The clues for future optimization are proposed from three aspects: firstly, developing and implementing a way to adjust user styling in real-time, so that the recommendation output of this system always matches the creative characteristics of oneself; The second is to expand data sources outside the domain to enhance the knowledge of the system's applicability to multiple fields; The third is algorithm self optimization based on machine learning, which fully utilizes user feedback to achieve rapid correction of common errors and accurate capture of unusual needs, thereby maintaining high operability and satisfaction among various user groups.

6.2 Research on the promotion path and scalability of the system in educational scenarios

Promoting the proposed generative-algorithm-based writing system in education requires consideration of technical adaptability, pedagogical integration, and sustainable maintenance. The introduction of this system should not be limited to a single writing tool, but should serve as a comprehensive support platform for classroom teaching, after-school tutoring, and self-directed learning. The promotion path can be divided into three stages: pilot verification, hierarchical promotion, and large-scale application.

In the pilot phase, partial deployment was carried out on some universities, language training institutions, and language online classrooms to obtain data on various teaching methods, and to study the effectiveness and stability of real-time writing assistance, task assignment evaluation, scientific research assistance, etc. based on generative algorithms; In the process of hierarchical amplification, based on the differentiated needs of high school, university, and adult learning in different disciplines, the recommendation fineness, generation format, and response methods were systematically optimized to highly match the subject objectives and curriculum structure; In the global application stage, relying on the architecture combination of cloud computing and edge computing, it has realized the high-performance access of large-scale parallel users and satisfied the learning needs and syllabus of each region based on the iterative strategy of localized language models.

In terms of the system's scalability, the structure should have multiple modules and support for hot swapping, in order to quickly integrate new functions such as cross domain writing support, multilingual generation, and support for advanced integration of learning management systems (LMS) when new educational needs change. In addition, the core algorithm should provide open interfaces to enable the generation and evaluation of third-party educational software usage, achieving ecological expansion. With the help of data-driven iterative optimization and cross platform capabilities, this system can extend from a simple educational software to an intelligent learning assistant in the digital revolution of education, and has long-term usage prospects and the possibility of cross-border migration.This empirical comparison further confirms that the integration of semantic-control and feedback mechanisms enables the system to outperform existing SOTA writing assistants in both linguistic accuracy and user adaptation.

7 Conclusion

This study focuses on building a comprehensive algorithm system for machine-assisted human creation of English writing works based on generative algorithms, including system architecture, core algorithm and text processing program design and functional debugging, testing effectiveness and optimization strategy research. Implement the modular architecture system and functional division design that combines core algorithms, semantic control, and human-computer dialogue into the system design, and build a complete closed-loop system based on input analysis to personalized suggestion output work.

In terms of performance evaluation, a comprehensive evaluation method has been constructed for multiple dimensions such as generation quality, user utility, and user experience. Through experimental data and a large amount of data in the application environment, the error situation of

the system is measured from important dimensions such as accuracy, fluency, and latency, and corresponding constraints and error types are identified. Therefore, corresponding modification plans are proposed, including adjusting the structural parameters of the core model, adjusting the adaptability of parameters, increasing the sample size to achieve error driven reconstruction and update, real-time feedback generation, reducing latency calculation, and multi-threaded parallel mode. This has resulted in significant improvements in response speed, generation quality, and user matching performance.

Through research, it has been found that the system can provide good applicability in the use of different language abilities and creative tasks, and can also be expanded in teaching space and different domain spaces. Technically, a reasoning architecture that reduces latency and flexible adjustment methods have been proposed. In application, it can be used to support intelligent learning, academic writing, and cross-cultural communication, and has the possibility of continuous expansion and promotion. For future work, how to further support the application of more languages and multimodal generation functions, so that they can better play their role in learning spaces, research spaces, etc., and continuously enrich their value is an important direction of work.

Ethical and Privacy Statement:

All user data collected in this study—including writing logs, feedback interactions, and questionnaire responses—were anonymized at the time of storage. No personally identifiable information (PII) was retained. The research protocol was reviewed and approved by the Academic Ethics Committee of Zhejiang Yuexiu University (Approval No. ZYU-AI2025-015). Participants provided written informed consent before data collection. Data handling complied with the requirements of the Chinese Personal Information Protection Law (PIPL, 2021) and relevant principles of the EU General Data Protection Regulation (GDPR). All datasets were used solely for academic research purposes and will not be shared publicly without additional consent.

References

[1] Yuan L .Research on English Hybrid Assisted Teaching System Using Contextual Support of R-CNN[J].Wireless Communications & Mobile Computing, 2022.https://dol:10.1155/2022/5358546.

[2] Mei Z , Lin Z .Design of Vocabulary Query System in Computer Aided English Translation Teaching[C]//EAI International Conference, BigIoT-EDU.Springer, Cham, 2024.https://dol:10.1007/978-3-031-63136-8_8.

[3] Mejeh M, Rehm M. Taking adaptive learning in educational settings to the next level: Leveraging natural language processing for improved personalization[J]. Educational technology research and development, 2024, 72(3): 1597-1621.https://doi.org/10.1007/s11423-024-10345-1

[4] Chai J .On Teaching Graduate English Academic Writing Based on Artificial Intelligence[J].Procedia Computer Science, 2024, 247(000):1366-1373.https://dol:10.1016/j.procs.2024.10.163.

[5] Wang C , Li Z , Bonk C .Understanding self-directed learning in AI-Assisted writing: A mixed methods study of postsecondary learners[J].Computers and Education: Artificial Intelligence, 2024, 6.https://dol:10.1016/j.caeai.2024.100247.

[6] Ding L, Zou D. Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence[J]. Education and Information Technologies, 2024, 29(11): 14151-14203.https://doi.org/10.1007/s10639-023-12402-3

[7] Jingning L .Speech recognition based on mobile sensor networks application in English education intelligent assisted learning system[J].Measurement: Sensors, 2024, 32(000).https://dol:10.1016/j.measen.2024.101084.

[8] Jiang L , Wang Y .English Online Teaching Assistant System Based on Genetic Algorithms[C]//EAI International Conference, BigIoT-EDU.Springer, Cham, 2024.https://dol:10.1007/978-3-031-63136-8_49.

[9] Tu X , Zhang H .Educational Service Platform and English Corpus Design Based on Embedded Intelligent System and Cloud Computing[J].Wireless Communications & Mobile Computing, 2022.https://dol:10.1155/2022/1916941.

[10] Dai B .Simulation of interactive E-learning mode based on association rule mining in computer English reading assistance system[J].Entertainment Computing, 2024, 51.https://dol:10.1016/j.entcom.2024.100742.

[11] Zhao Q , Nazir S .English Multimode Production and Usage by Artificial Intelligence and Online Reading for Sustaining Effectiveness[J].Mobile Information Systems, 2022.https://dol:10.1155/2022/6780502.

[12] Pack A , Barrett A , Escalante J .Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability[J].Computers and Education: Artificial Intelligence, 2024, 6.https://dol:10.1016/j.caeai.2024.100234.

[13] Lee Y J , Davis R O .A case study of implementing generative AI in university's general English courses[J].Contemporary Educational Technology, 2024, 16(4).https://dol:10.30935/cedtech/15218.

[14] Yang S H .Enhancing English as a Foreign Language Learners' Writing Performance and Experience With an Artificial Intelligence Powered Tool[J].Springer, Cham, 2024.https://dol:10.1007/978-3-031-76293-2_9.

[15] Liu M , Zhang L J , Biebricher C .Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional

writing[J].Computers & education, 2024, 211(Apr.):1.1-1.21.https://dol:10.1016/j.compedu.2023.104977.

[16] Han X .An Improved Classification Model for English Syntax Error Correction Design of DL Algorithm[J].International Arab Journal of Information Technology (IAJIT), 2024, 21(4).https://dol:10.34028/iajit/21/4/2.

[17] Gustilo L , Ong E , Lapinid M R .Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era[J].International Journal for Educational Integrity, 2024, 20(1):43.https://dol:10.1007/s40979-024-00153-8.

[18] Asad M M , Shahzad S , Shah S H A ,et al.ChatGPT as artificial intelligence-based generative multimedia for English writing pedagogy: challenges and opportunities from an educator's perspective[J].International Journal of Information and Learning Technology, 2024, 41(5).https://dol:10.1108/IJILT-02-2024-0021.

[19] Ali S .Locating the Intersection of Generative Artificial Intelligence and Human English Writing Skills: A Comparative Study[J].Arab World English Journal, 2024.https://dol:10.24093/awej/ChatGPT.7.

[20] Zenni R D , Andrew N R .Artificial Intelligence text generators for overcoming language barriers in ecological research communication[J].Austral Ecology, 2023(7):48.https://dol:10.1111/aec.13417.

Semi-Supervised Voronoi-TSVM Path Planning for Automated Guided Vehicles under Multi-Modal Perception

Jing Sun
Tianjin Urban Construction Management & Vocation Technology College, Tianjin 300134, China
E-mail: Jing_Sun521@126.com

Keywords: transductive support vector machine, path planning, Voronoi diagram, automated guided vehicle

Received: June 13, 2025

This work proposes a fusion optimization method combining Voronoi diagram and Transductive Support Vector Machine (TSVM) to address the path planning problem of Automated Guided Vehicle (AGV) in complex environments. First, this method uses a Voronoi diagram to construct the spatial skeleton of the environment and generate an initial path. Second, it introduces TSVM to perform semi-supervised classification and optimized screening on path segments, to screen out safer and smoother segments. Finally, this method ensures the safety of the path while improving its smoothness and operational efficiency. In evaluation indicators, path smoothness is defined as the standard deviation of path angle changes (unit: radian), and a smaller value indicates a smoother path; collision rate refers to the proportion of path points where the minimum distance from obstacles is lower than the safety threshold. The experimental results show that the optimized model has a path length of 42.987 meters (m) under LiDAR data; it is remarkably shorter than that of Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) (46.879m) and Deep Reinforcement Learning - Model Predictive Control (DRL-MPC) (45.212m). This indicates that the optimized model can offer more efficient path choices. Concerning computation time, this model performs excellently, with 1.345 seconds (s) under LiDAR data, much lower than CNN-LSTM's 2.346s and DRL-MPC's 1.879s, demonstrating its high efficiency in path planning. Regarding path smoothness, the optimized model achieves 0.208, superior to CNN-LSTM's 0.345 and DRL-MPC's 0.298, thus reducing vibrations and path deviations. Moreover, this model has a collision rate of 0.012 under LiDAR data, remarkably lower than CNN-LSTM's 0.062 and DRL-MPC's 0.045. Consequently, this work develops a novel path planning strategy integrating classification learning with spatial mapping. It also validates the strategy's universality and robustness in multi-source perception scenarios, providing theoretically sound and engineering-feasible optimization solutions for intelligent AGV systems.

Povzetek: Članek obravnava učinkovitost klasičnih AGV-metod pri načrtovanju poti v kompleksnih okoljih. Predlaga hibrid Voronoi-TSVM: Voronoijev diagram ustvari varnostni prostorski skelet, TSVM pa polnadzorovano razvrsti in filtrira odseke poti glede na dolžino, zavoje in gladkost. Model skrajša pot, izboljša gladkost in zmanjša kolizije.

1 Introduction

In modern manufacturing and intelligent logistics systems, Automated Guided Vehicle (AGV) has been widely applied in warehousing operations, production line material handling, and hospital automated delivery scenarios [1-3]. The AGV's core functionality relies on its path planning capability. Traditional methods, such as grid-based mapping or A* algorithm, demonstrate satisfactory performance in structured environments. However, they often generate path redundancy and obstacle avoidance delays when encountering complex obstacle configurations in dynamic, unstructured settings [4]. The Voronoi diagram, with its superior spatial partitioning capability, has been extensively utilized in robotic path planning. The basic idea is to divide the space into several regions, where each region consists of all points closest to a specific reference point. This division method can naturally create obstacle-avoiding boundaries and provide a relatively safe navigation framework for AGV [5-7]. Nevertheless, paths generated by Voronoi diagrams frequently suffer from excessive turns and suboptimal length, limiting further improvements in path quality [8]. Recent advancements have introduced machine learning (ML) techniques into path optimization. Transductive Support Vector Machine (TSVM), as a semi-supervised classification model, exhibits excellent boundary discrimination capability that proves valuable for path region classification problems. The integration of TSVM with Voronoi diagrams enables safety-zone-based path mapping; this integration also optimizes waypoint selection and connection patterns, thus enhancing path smoothness and overall efficiency.

Building upon these foundations, this work proposes a novel AGV path planning method combining Voronoi diagrams with TSVM. It aims to achieve efficient, safe,

and continuous path generation in complex environments. The methodology initially constructs obstacle-avoidance regions through Voronoi diagrams. Then, it employs TSVM for classification and filtering of candidate paths to optimize waypoint distribution while minimizing unnecessary turns and path fluctuations.

This work addresses the following questions:

(i) How can the partitioning capability of Voronoi diagrams be effectively integrated with the discriminative advantages of TSVM for path generation and optimization of AGV?

(ii) Does the proposed model outperform existing comparative methods under typical perception data?

(iii) Does this approach exhibit stable advantages in key indicators such as path length, smoothness, and collision rate?

Through discussions of these questions, this work provides a modeling paradigm for AGV path planning that integrates the advantages of spatial geometry and intelligent classification. Thus, it can offer new ideas and practical support for intelligent mobile agent navigation in complex environments.

2 Related work

With the continuous development of AGV technology, AGV path planning has become a hot research topic. Researchers have proposed various methods from different perspectives to achieve efficient and safe path planning for the AGV.

Ming et al. proposed a shortest path algorithm that traversed all the nodes in a graph to find the shortest path from the starting point to the destination. The Dijkstra algorithm could effectively handle path planning problems in static environments. However, it suffered from high computational complexity and poor real-time performance in complex environments [9]. Pan applied the Voronoi diagram to AGV path planning and proposed a path generation method based on the Voronoi diagram. By creating a Voronoi diagram of the environment, this method could effectively avoid path overlap with obstacles, ensuring the safe operation of the AGV. Nevertheless, the paths generated by the Voronoi diagram often suffer from path redundancy and excessive turns, requiring further optimization [10]. Gul introduced an improved Voronoi diagram path planning method that considered the impact of dynamic obstacles during the construction of the Voronoi diagram. Studies indicated that the Voronoi diagram effectively adjusted the path in dynamic environments, enhancing the AGV's flexibility and adaptability. However, this method still faced challenges in high computational complexity, especially when dealing with complex obstacle layouts [11]. Chek presented a method based on Support Vector Machine (SVM) for AGV path optimization. This approach could learn from historical path data to optimize the AGV's path selection. While SVM effectively reduced path redundancy, its computational cost during training could be quite high, impacting its application, especially in large-scale environments [12]. The work further sorts out the comparison of typical AGV path planning methods, as exhibited in Table 1:

Table 1: Comparison of typical AGV path planning methods

Method name	Input data type	Computational efficiency	Path quality	Main limitations
Dijkstra algorithm	Grid maps, static maps	Medium	Feasible, but the path is lengthy.	Path smoothness is not considered; The amount of computation grows rapidly with the number of nodes.
A* algorithm	Grid-based maps, heuristics (e.g. Euclidean distances)	Higher	The path is usually short, with frequent turns.	It has poor adaptability to the dynamic environment, and the path is not smooth enough.
Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM)	Multimodal data, such as depth images and LiDAR	>2.3 s; Lower (inference time)	The path smoothness is 0.3; it adapts to dynamic scenarios, and the path is relatively continuous.	Model training relies on large amounts of data; it has limited generalization ability.
Deep Reinforcement Learning - Model Predictive Control (DRL-MPC)	State-space imagery or LiDAR	1.88 s; Medium to lower	The average path length is approximately 45.2 meters (m), superior to conventional, balancing energy consumption and dynamic changes.	Slow convergence; difficulty interpreting the decision-making process; Sensitive to hyperparameters

The above analysis suggests that traditional methods often separate environmental modeling from path optimization. They either rely on complex numerical optimization or train end-to-end deep learning (DL) models with a large amount of labeled data. Although existing methods show specific advantages in different scenarios, they still have obvious limitations in generalization and robustness. For example, DRL-MPC

performs excellently in dynamic environments, but its decision-making process is opaque and it is sensitive to hyperparameters, leading to a decline in generalization ability in unknown environments; CNN-LSTM relies on a large amount of labeled data for training, and its performance fluctuates markedly when data is scarce or perception quality is uneven. In contrast, the proposed Voronoi-TSVM fusion model shows certain advantages in solving the AGV path planning problem. The innovation of this work lies in the in-depth integration of the Voronoi diagram's spatial structuring ability and the TSVM's semi-supervised boundary discrimination ability, forming a "generation-screening" collaborative framework. This framework aims to overcome the inherent problems of redundant paths and excessive turns in traditional geometric methods. Meanwhile, it completes high-quality path optimization based on only a small number of labeled samples. Moreover, it reduces the dependence on data scale and quality, providing a new solution for the AGV's efficient navigation in dynamic and data-scarce environments. This integration strategy can improve the reliability, interpretability, and environmental adaptability of path planning, and make up for the gap between existing methods at the structural and learning levels.

Against this backdrop, this work introduces TSVM; it integrates candidate path skeletons generated by Voronoi diagrams to achieve semi-supervised optimization of candidate path regions using minimal labeled data. Thus, it simultaneously enhances path feasibility, continuity, and execution efficiency. Unlike previous strategies that use ML as a "path re-scoring" tool, the proposed method emphasizes a linkage mechanism between geometric mapping and classification discrimination. This represents an integrated path planning framework that fuses structural generation with semantic optimization.

3 Research method

3.1 The use of voronoi diagram in AGV path planning

The Voronoi diagram, a mathematical tool based on spatial geometric partitioning, is extensively used in fields such as robot navigation, target allocation, map segmentation, and path planning. Its theoretical foundation stems from the "nearest neighbor division" concept in computational geometry. In the context of AGV path planning, the core advantage of the Voronoi diagram lies in its ability to explicitly partition the environmental space geometrically. Moreover, it can effectively distinguish between obstacle areas and free passage zones, providing an interpretable, safe, and highly robust geometric foundation for path generation [13, 14].

Specifically, the Voronoi diagram divides the entire environment into a set of non-overlapping subregions through spatial distance comparisons among several generator points. For any point within each sub-region, the distance to the corresponding generator point is the smallest. These generator points typically include obstacle boundary points, start points, end points, and key landmarks. When applied to path planning, this partitioning method restricts feasible paths to the positions farthest from obstacles (i.e., Voronoi boundaries), maximizing path safety and obstacle avoidance capabilities. The mathematical rationale for Voronoi path planning is that the greater the distance between an AGV and obstacles, the lower the collision probability [15,16]. Voronoi boundaries are formed by the perpendicular bisectors of adjacent obstacle points and naturally lie on the "maximum clearance channels" between obstacles. Consequently, paths constructed based on Voronoi diagrams can effectively provide collision-free navigation solutions with high margin redundancy. This property is particularly critical for scenarios with complex structures, dense obstacles, or irregular shapes. The typical steps for constructing a Voronoi diagram include:

(1) Choosing generator points: Based on the obstacles in the environment or specific needs, appropriate generator points are selected.

(2) Calculating Voronoi boundaries: The boundaries of the Voronoi diagram are constructed by calculating the perpendicular bisectors between each pair of generator points.

(3) Generating Voronoi cells: Grounded in the calculated boundaries, the space is divided into multiple Voronoi cells.

Through these steps, the AGV can use the Voronoi diagram to generate collision-free paths. The generated path usually lies along the boundaries of the Voronoi cells, ensuring that the AGV does not enter obstacle regions during its movement [17-19].

3.2 The use of TSVM in path optimization

In the AGV path planning system, TSVM evaluates and filters the preliminary paths generated by the Voronoi diagram. Its task is not to directly generate paths or perform numerical optimization. Conversely, it identifies high-quality path segments among multiple candidate schemes, thereby assisting in forming the final optimized path [20, 21]. As a classification model, TSVM constructs two hyperplanes, close to positive and negative class samples, thus demonstrating good performance in handling imbalanced samples and heterogeneous data [22, 23].

The path optimization process is modeled as a discriminative problem in the proposed method. First, the system constructs a set of initial path candidates based on the Voronoi diagram, which typically covers multiple possible safe passage areas. Each path is divided into several path segments. Then, three types of key features

are extracted for each segment: path length, number of turns, and smoothness (approximated by angular changes). Each path segment corresponds to a three-dimensional feature vector to characterize its structural complexity and passage quality [24, 25]. In the feature vector, "path length" refers to the total Euclidean distance of the segmented trajectory (unit: m); "number of turns" is counted by the number of points where the change in direction angle exceeds the threshold; "smoothness" is calculated as the average value of the absolute changes in angles formed by all consecutive three points within the segment. The calculation method of the angle is as follows. Taking three consecutive points P_1, P_2, and P_3, the angle is calculated between vectors P_1P_2 and P_2P_3 (unit: radian); the absolute value of this angle reflects the intensity of direction change. The smoothness of the entire segment is the average value of the absolute angles of all consecutive three points, and a smaller value indicates a smoother path. Turning point detection relies on the discrete calculation of direction angles and is realized by combining a sliding window and threshold judgment.

The system builds a path quality sample set to train the TSVM model. Well-performing segments (e.g., shorter length, fewer turns, smoother movement) are positive class samples. In contrast, poorly performing segments (e.g., close to obstacle edges, sharp consecutive turns, or obvious path oscillations) are defined as negative class samples [26]. These samples can be derived from historical paths in previous tasks, expert experience annotations, or simulation sampling. Once trained, the model can discriminate the quality of newly generated path segments.

In practical operation, the AGV system generates multiple feasible paths from the Voronoi diagram, with each path segmented and feature-extracted before being input into TSVM for classification [27, 28]. The labels output by the model determine whether each path segment is retained. A structurally superior and more reliable path is ultimately formed by preserving high-quality segments, removing low-quality ones, and reconfiguring the remaining segments. Notably, TSVM does not directly perform path optimization but provides a path quality discrimination mechanism, acting as a "path segment selector" in the system's path planning process. Its role is to evaluate the rationality and executability of path segments based on structural features. Meanwhile, it indirectly enhances overall path performance by reducing redundant length, minimizing unnecessary turns, and improving smoothness [29].

Additionally, the common issue of sample imbalance in path optimization is addressed (where high-quality path segments are far fewer than low-quality ones). TSVM's dual hyperplane construction mechanism enables more stable learning of path segment classification boundaries. This can effectively avoid misjudgments caused by skewed sample distributions [30, 31]. This mechanism is particularly beneficial for the AGV to quickly adapt to new scenarios in complex or dynamic environments.

Regarding model execution logic, TSVM does not directly modify path point coordinates or participate in the internal search process of the path generation algorithm. Instead, it performs quality evaluation and screening on a set of candidate paths or segments generated externally from the Voronoi path structure. The final path reconstruction and execution are completed through a post-processing process driven by TSVM classification results.

In summary, TSVM primarily undertakes path segment-level discrimination tasks in the proposed AGV path planning system. Through learning and classifying path structural features, the model assists in filtering and reorganizing candidate path segments, thereby optimizing overall path structure and enhancing execution performance. This mechanism effectively bridges the global spatial partitioning capability of Voronoi diagrams with the practical need for local optimization at the path execution level.

3.3 Construction of AGV path planning model based on voronoi diagram and TSVM

In complex dynamic environments, AGV path planning faces multiple constraints among obstacle avoidance, path redundancy, and real-time performance. Traditional methods typically separate path construction and optimization, failing to dynamically adapt to environmental changes. This work proposes a path planning model fusing Voronoi diagrams and TSVM to address these issues. The model is designed to combine environmental space partitioning with path region discrimination, achieving fine-grained path optimization while ensuring safe and reachable routes. The method consists of two main stages. First, Voronoi diagrams are used to partition obstacle environments and construct global path graphs; Second, the A* algorithm is applied to search for the initial path on these graphs. Then, the path candidate points are semi-supervised and classified through TSVM to eliminate the inappropriate path segments and improve path quality. The final path is smoothed and converted into control commands for input to the AGV control module.

Considering that TSVM relies on minimal labeled samples for optimization training, this work designs a weak label generation mechanism. Specifically, path start/end points and their adjacent areas are designated as positive samples. The sample data required for TSVM training is mainly obtained by running traditional planning algorithms in a simulation environment to conduct a large number of path samplings. Positive samples are selected from path points that are far from obstacles and have gentle curvature changes; negative samples are taken from areas close to obstacles or with excessively sharp turns. For some samples that are

difficult to automatically identify, experts provide auxiliary labeling based on the rules of safe distance and movement fluency.

Regions close to obstacle boundaries or with path deviations are defined as negative samples. Most remaining Voronoi diagram path candidate points are input as unlabeled samples. This strategy balances training efficiency and discriminative capabilities in scenarios with scarce labels, meeting the fast adaptation requirements of AGV online operations. The entire system comprises five modules—environmental perception, spatial partitioning, path generation, path optimization, and smooth control. The specific workflow is outlined in the following pseudocode:

```
# Algorithm: AGV Path Planning with Voronoi Graph and TSVM
Input:
  OccupancyMap ← sensor fusion (LiDAR, IMU, depth camera)
  Obstacles ← {O1, O2, ..., On}
  StartPoint, GoalPoint
  TSVM_Params ← {C=1.0, gamma=0.5, tolerance=0.001, kernel='RBF'}
Output:
  OptimizedPath ← safe, smooth and executable path sequence
1. Construct Voronoi Diagram from Obstacles
  VoronoiGraph = generate_voronoi(OccupancyMap, resolution=0.1, inflation=0.5)
2. Perform A* Search on Voronoi Graph
  InitialPath = A_star_search(VoronoiGraph, StartPoint, GoalPoint, max_length=50)
3. Generate training samples for TSVM
  PositiveSamples = select_points_near(StartPoint, GoalPoint)
  NegativeSamples = select_points_near_obstacles(OccupancyMap)
  UnlabeledSamples = all_remaining_points(InitialPath)
4. Train TSVM using semi-supervised learning
  TSVM_Model = train_tsvm(PositiveSamples, NegativeSamples, UnlabeledSamples, TSVM_Params)
5. Filter path using TSVM prediction
  FilteredPath = [p for p in InitialPath if TSVM_Model.predict(p) == 'valid']
6. Smooth the path and convert to control commands
  FinalPath = smooth_path(FilteredPath)
  send_to_controller(FinalPath)
```

The computational complexity of each step in the above pseudocode varies. The time complexity of Voronoi diagram construction is usually O(n log n), where n is the number of obstacles. The complexity of performing A* search on the Voronoi diagram depends on the size of the diagram, and in the worst case, it is O(E + V log V); among them, V and E represent the number of vertices and edges, respectively. The time complexity of the TSVM classification stage is O(n), which has a linear relationship with the number of path points to be classified. The structure of the VoronoiGraph consists of generating points and edges. The path filtering logic is mainly implemented through TSVM prediction. The classifier outputs a "valid" or "invalid" label for each path point, and only retains consecutive valid points to form the final path.

From the above analysis, the specific implementation steps of the fused path are as follows. 1) The Voronoi diagram is responsible for spatial division of the environment and generating the initial path graph; 2) On this basis, TSVM performs semi-supervised classification on path points to screen out more reasonable path segments; 3) The two are integrated through feature extraction of path points and classification labels. Among them, the information provided by the Voronoi diagram, such as the coordinates of path points and their distance from obstacles, is converted into structured feature vectors processable by TSVM. These vectors include path length, number of turns, smoothness, and other relevant parameters. The training samples of TSVM are composed of positive samples, negative samples, and a large number of unlabeled Voronoi path points. The judgment of path quality is realized through semi-supervised learning. This integration mechanism aims to keep the path safe in geometric structure and make it more continuous and feasible at the local execution level.

The method offers advantages in two aspects. On one hand, the Voronoi diagram rapidly constructs safe regions to ensure basic path feasibility. On the other hand, TSVM achieves discriminative learning at the path point level to optimize path continuity and viability. Additionally, TSVM training employs the Sequential Minimal Optimization (SMO) algorithm with minimal labeled data. The average training time is 3.7 seconds (s), and the memory usage is approximately 260 megabytes (MB). It is suitable for deployment on resource-constrained edge computing devices. It should be noted that the current experimental results show that the model exhibits characteristics potentially suitable for edge device deployment in computational efficiency and resource consumption. However, this conclusion is mainly drawn based on simulation tests on standard hardware platforms and has not yet been fully verified in actual embedded systems. Therefore, before applying it to real edge computing scenarios, further hardware-in-the-loop testing and targeted optimization still need to be conducted.

Overall, the proposed Voronoi-TSVM joint path planning method fully integrates the advantages of spatial topology construction and ML optimization. This method demonstrates good real-time performance, discriminative capability, and execution stability in complex environments. This provides a new learning-enabled path generation paradigm for intelligent AGV navigation.

4 Experimental design

This work constructs a complete simulation experiment platform to verify the effectiveness of the proposed method. This platform integrates the Gazebo environment for physical simulation under the ROS Noetic framework and uses RViz for visualization. The environment map is constructed in real time through Simultaneous Localization and Mapping (SLAM) technology and converted into a two-dimensional Occupancy Grid Map as the input for path planning. The pseudocode form of the standard execution steps of the experiment's core algorithm process—the Voronoi-TSVM-based AGV path planning model—has been presented in the previous section.

The dataset used is from the Robot Operating System (ROS), which encompasses several public datasets related to AGV and path planning, containing experimental data on AGV navigation, path planning, and obstacle avoidance in various environments. The experiment utilizes multiple indoor environment scenarios from the Technical University of Munich (TUM) dataset, which is widely adopted in ROS. Additionally, it incorporates outdoor structured road data from the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) dataset. The test environment includes 12 different environmental layouts, among which 8 are static environments and 4 contain dynamic interference objects; in each environment, 20 paths with different start and end points are planned, resulting in a total of 240 test paths. The datasets available on the ROS platform typically include experimental data for tasks such as robot path planning, obstacle detection, and path optimization. The datasets can be downloaded from the official ROS website (https://github.com/ros-perception/). In the experiment, the dataset is divided into three dimensions based on the sensor data type: LiDAR, depth image, and Inertial Measurement Unit (IMU) data.

The experiments are based on public datasets provided by the ROS platform. However, the dataset covers multiple typical AGV application scenarios, including indoor office areas, warehouse logistics corridors, and semi-structured obstacle distribution areas, with good environmental diversity and task representativeness. In the experimental design, layouts containing static obstacles (e.g., fixed shelves and walls) and dynamic interferences (e.g., moving obstacle simulators) simulate complex scenarios in real urban navigation or industrial warehouse environments. As the ROS dataset encompasses multi-source sensor data such as LiDAR, depth images, and IMU, the data are divided and modeled by sensor type in the experiments. This ensures comparable performance of the model under different perceptual inputs.

While this work focuses on presenting path planning results based on LiDAR data, the constructed model architecture remains general to input data types. Meanwhile, training and evaluation are conducted for depth image and IMU data dimensions. Due to space limitations, this work does not separately expand on all experimental details under different visual inputs. However, the model structure and feature extraction methods exhibit adaptability to diverse sensors and strong transferable generalization potential. Regarding the use of simulation experiments, the selected ROS Gazebo environment has high physical fidelity in path modeling, sensor modeling, and dynamic obstacle simulation. This can accurately reflect control behaviors and environmental interferences in real-world path planning systems. As this work focuses on proposing the model structure and discriminative optimization strategy, testing in real physical deployments is an important direction for future research expansion. It also fully recognizes the importance of real-world environment verification for model practicality and stability.

The comparison models selected for the experiment are CNN-LSTM and DRL-MPC, representing the application of DL and reinforcement learning (RL) in path planning, respectively. CNN-LSTM strongly supports spatial feature extraction and time-series modeling, making it suitable for dynamic path planning in complex environments. Moreover, DRL-MPC combines RL and MPC, offering a novel solution for AGV path planning and real-time optimization. It is particularly suitable for dynamic environments with high uncertainty. Both models are highly innovative and provide effective comparison and analysis for the path planning method based on the Voronoi diagram and TSVM. Previous studies indicated that heuristic algorithms, such as A* and D*, exhibited strong path optimization capabilities in static environments. Moreover, some hybrid methods performed well in specific scenarios. However, this work does not include them in comparative experiments for two reasons. On one hand, this work focuses on intelligent path optimization capabilities under high-dimensional perceptual inputs and complex scenarios. CNN-LSTM and DRL-MPC are mainstream DL models applied to AGV path planning in recent years. They represent two typical paradigms based on sequential modeling and RL + predictive control, respectively, with strong representativeness and a broad application foundation. On the other hand, heuristic algorithms like A* and D* can converge quickly under the premise of known structure diagrams and static obstacles. However, in highly dynamic scenarios or those with multi-source perceptual inputs, their performance is easily limited by map construction methods; they also lack end-to-end data-driven capabilities. Traditional SLAM methods integrated with path optimization are mostly system-level encapsulations, making it difficult to compare path generation performance as single models. Therefore, selecting CNN-LSTM and DRL-MPC as benchmarks covers the characteristics of mainstream neural methods; meanwhile, it forms a contrast with the

proposed “Voronoi+TSVM” structural model at the application level. This highlights the proposed model's comprehensive advantages in path quality and real-time performance.

The experimental setup is as follows:

(1) Processor: Intel Xeon W-2295 (18 cores, 3.00 GHz)
(2) Memory: 64 GB DDR4 RAM
(3) Storage: 1 TB SSD (main storage), 4 TB HDD (data storage)
(4) Graphics Card: NVIDIA Quadro RTX 4000 8GB
(5) Development Platform: ROS Noetic
(6) ROS Version: Noetic Ninjemys
(7) Relevant Packages: move_base, gmapping, amcl, nav_core
(8) Path Planning Library: Open Motion Planning Library (OMPL)

The experiment also sets parameters for the optimized model, with the following configuration. Grid resolution is set to 0.1m, obstacle radius is 2.0 m, boundary expansion factor is 0.5m, and penalty parameter is 1.0. The kernel function is Radial Basis Function (RBF), with a tolerance of 0.001; The path planning algorithm is A*, and the maximum path length is 50m. During training, a radial basis function kernel is employed with a kernel parameter γ set to 0.5, enhancing the model's nonlinear discriminative capability in path boundary regions. The regularization constant C is 1.0 to balance the relationship between classification margin and error tolerance. The tolerance is set at 0.001 to ensure convergence accuracy during model optimization, with a maximum of 1,000 iterations. The SMO method is an optimization solver, suitable for semi-supervised classification problems of moderate scale. Approximately 30% of the path region samples in the training data are labeled, with the rest being unlabeled, leveraging TSVM's generalization ability under limited labeling conditions. In the experiment, the same TSVM model is used for path optimization in all scenarios. This model is trained on a mixed dataset containing multiple environment types and is not trained separately for individual scenarios. This setup aims to verify the model's generalization ability in unseen environments; it is more in line with the requirement in practical applications that AGV needs to adapt to diverse scenarios. Moreover, to verify the advantages of TSVM when labeled data is limited, comparative experiments are conducted under different labeling ratios (10%, 30%, 50%). The results show that when the proportion of labeled samples decreases from 50% to 10%, the path classification accuracy of TSVM decreases by 5.2%. Meanwhile, the accuracy of the CNN-LSTM method declines by 13.7% under the same conditions. This confirms that TSVM has better stability in semi-supervised scenarios and can more effectively use unlabeled data to maintain model performance. Under standard hardware configurations, each TSVM training takes approximately 3.7 s with a standard deviation (SD) of 0.6 s. The model typically converges within 200 to 350 iterations, meeting error threshold conditions. In terms of resource consumption, a single training phase occupies approximately 260 MB of memory, demonstrating feasibility for deployment on edge computing devices. To enhance model robustness, all input samples undergo standard normalization before training. Boundary sample resampling is performed on class distributions to mitigate overfitting risks caused by initial sample imbalance.

Experiments are conducted at two levels to evaluate the proposed AGV path planning model's performance. One is the performance evaluation of path planning and execution, and the other is assessing path accuracy and control performance in simulation scenarios. Four representative indicators are selected for each dimension, constructing a complete evaluation system by integrating actual path performance and control responses.

Regarding path planning performance evaluation, path length is first considered. Path length refers to the total driving distance of the planned path from the starting point to the end point for the AGV, measuring whether the path is concise and efficient. Second is the calculation time. The total time required for the algorithm to generate a complete path after receiving map information reflects the model's operational efficiency and real-time performance. Additionally, path smoothness is used to measure the intensity of turns in the path; higher smoothness indicates a more continuous and fluid path, which is conducive to the stable operation of the AGV. Finally, the collision rate refers to the proportion of path points close to or crossing obstacles in the path, employed to reflect the safety of the path; the lower the value, the safer it is. The collision rate is measured by the proportion of the minimum distance between the path point and the obstacle being lower than the safety threshold (0.2 m). In evaluating path execution performance, the arrival time is first concerned. That is, the actual time it takes for the AGV to travel from the starting point to the destination, which evaluates the execution efficiency of the path. Path deviation indicates the average offset between the actual running trajectory of the AGV and the original planned path; smaller deviations represent higher path execution accuracy. Energy consumption reflects the energy consumed by the AGV during execution, usually in watt-hours (Wh); the lower the value, the more energy-efficient the path. Energy consumption is calculated based on the integral of motor power and operation time (unit: Wh). The stability indicator measures the amplitude of attitude changes or control fluctuations of the AGV during path execution; higher stability means less interference from the path on equipment operation and a more reliable control system.

In simulation experiments, to evaluate the model's path planning accuracy, path shortness is selected as the first indicator; it represents the current path's relative

length compared to the shortest path among all candidate paths; the closer the value is to the optimal path, the more efficient the path is. Path stability is utilized to measure the consistency of path generation results under different environmental conditions, reflecting the model's robustness to environmental disturbances. This indicator is defined as the degree of change in path shape during multiple runs; it is measured by calculating the average value of the SD of each point's position (unit: m). A lower value indicates better path repeatability. Path consistency further measures the geometric overlap degree of paths generated multiple times. This indicator uses the Jaccard similarity coefficient to measure the overlap degree of paths generated in different runs. Its value is a dimensionless number between 0 and 1; higher consistency indicates more stable algorithm decisions. Path correction ability is the reciprocal of the time required for an AGV to return to the ideal path from a deviated position (unit: s^{-1}). The error rate represents the proportion of redundant segments or unreasonable path structures in the overall path and identifies potential redundancy issues in planning. In the dimension of AGV control performance evaluation, turning radius is first considered. Turning radius is characterized by the minimum value of the path curvature radius (unit: m); its calculation method is the statistical minimum value of the arc radii formed by consecutive path points; a smaller radius means sharper turns and higher requirements for the sensitivity of the AGV control system. Speed stability measures the fluctuation of the AGV's speed during path execution; smaller speed fluctuations show more precise control and more stable path execution. Path correction ability describes whether the AGV can quickly return to the correct route after deviating from the path; stronger correction ability means better recovery ability of the system to abnormal disturbances. Response time measures the time required for the AGV system to replan the path after environmental changes; the shorter the time, the more sensitive the system's response and the more adaptable it is to dynamic environments.

5 Experimental results and analysis

5.1 Model performance evaluation

Figure 1 shows the results of the path planning performance evaluation.

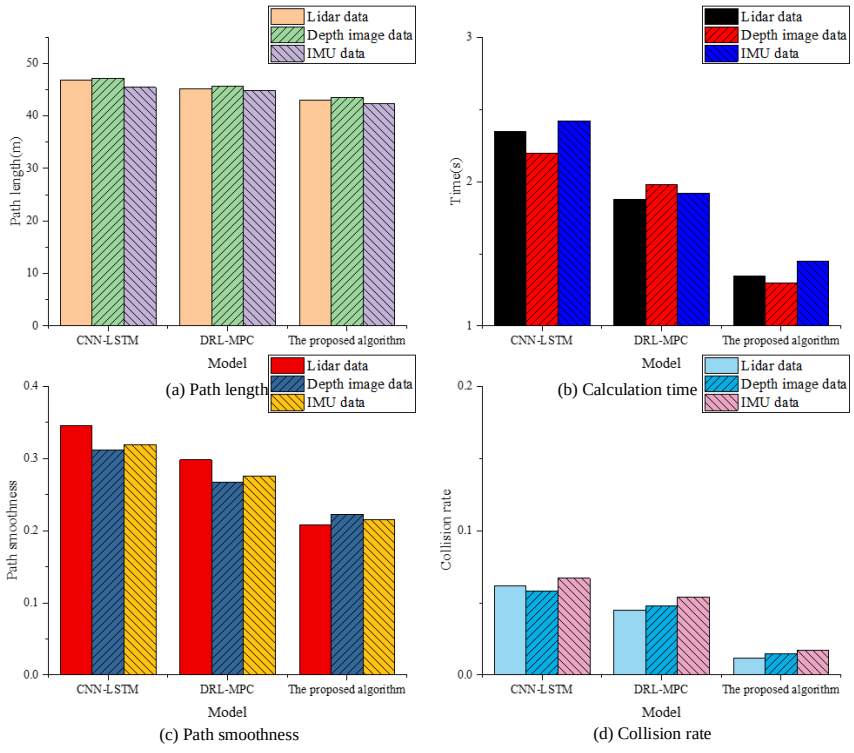


Figure 1: Performance evaluation of path planning ((a): Path length; (b): Calculation time; (c): Path smoothness; (d): Collision rate).

In Figure 1, the proposed optimized model exhibits remarkable advantages in terms of path length, especially when using LiDAR data, with a path length of 42.987 m. The optimized model is notably shorter than the CNN-LSTM (46.879 m) and DRL-MPC (45.212 m). This illustrates that the proposed model can reduce redundancy in the planned path and offer more efficient path choices. The path lengths of CNN-LSTM and DRL-MPC are longer, particularly CNN-LSTM, which demonstrates the longest path when using depth image

data. This might be due to this model choosing relatively complex paths when handling obstacles. The optimized model achieves a superior computation time of 1.345 s with LiDAR data, excelling CNN-LSTM's 2.346 s and DRL-MPC's 1.879 s. This indicates that this model has higher computational efficiency during path planning, enabling it to complete tasks more quickly. CNN-LSTM's computation time is relatively long, exceeding 2 s, especially with IMU data, likely due to the DL framework's high computational complexity. Path quality evaluation reveals the optimized model's exceptional smoothness performance (0.208 with LiDAR data), outperforming CNN-LSTM (0.345) and DRL-MPC (0.298). This suggests that the path is smoother and more continuous. CNN-LSTM and DRL-MPC have poorer path smoothness, particularly with LiDAR data, where the paths are more winding, potentially affecting the AGV's stability and precision. The optimized model's collision rate of 0.012 with LiDAR data demonstrates improvements over CNN-LSTM (0.062) and DRL-MPC (0.045). This demonstrates that the model can effectively avoid obstacles and ensure the safety of the AGV's path. CNN-LSTM and DRL-MPC have particular weaknesses with IMU data, where CNN-LSTM shows elevated collision rates (0.067). This may stem from less precise obstacle detection during the path planning process. Figure 2 presents the results of the path execution performance evaluation.

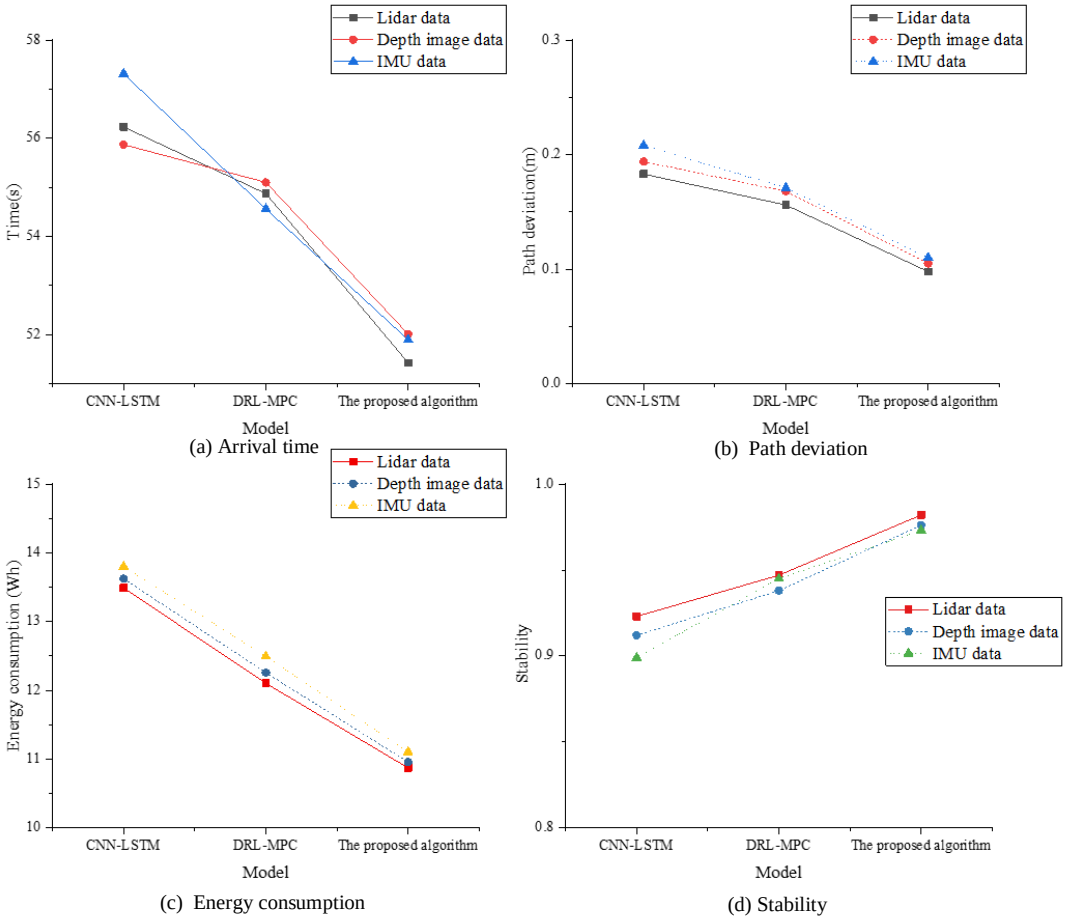


Figure 2: Performance evaluation of path execution ((a): Arrival time; (b): Path deviation; (c): Energy consumption; (d): Stability).

Figure 2 shows that the optimized model consistently achieves the shortest arrival time (51.895 s), especially with IMU data, markedly lower than CNN-LSTM (57.314 s) and DRL-MPC (54.561 s). This indicates that the optimized model exhibits superior path planning efficiency, enabling faster task completion. Notably, CNN-LSTM has the longest arrival time, particularly when processing IMU data, where it exceeds 57 s, showing its lower path planning efficiency. For path deviation, the optimized model shows the smallest value, with LiDAR data yielding a deviation of only 0.098 m, substantially lower than CNN-LSTM (0.183 m) and DRL-MPC (0.156 m). This shows that the optimized model ensures higher precision during path execution and can minimize errors. By contrast, CNN-LSTM and DRL-MPC exhibit larger path deviations, with CNN-LSTM showing particularly significant fluctuations when using depth image data, which may affect the accuracy of the AGV. Regarding energy consumption, the optimized model consistently shows the lowest energy consumption

(10.872Wh), particularly with LiDAR data, remarkably lower than CNN-LSTM's 13.489 Wh and DRL-MPC's 12.102 Wh. This dual efficiency-energy saving characteristic underscores the optimized model's operational advantages. In contrast, CNN-LSTM and DRL-MPC exhibit higher energy consumption, with CNN-LSTM's LiDAR-based results approaching 13.5 Wh; this reflects its lower path planning efficiency and resulting energy inefficiencies. In stability assessments, the optimized model is the most stable, particularly with LiDAR data, where its stability score reaches 0.982, markedly higher than CNN-LSTM (0.923) and DRL-MPC (0.947). This suggests that this model can better maintain path stability and continuity, reducing vibrations or path deviations. Conversely, CNN-LSTM has the poorest stability, especially in LiDAR scenarios, thus hindering the precise control of the AGV.

This work verifies whether the improvement of the optimized model in path planning and execution performance compared with the traditional model is statistically significant. It conducts a statistical analysis of the core indicators based on three sets of repeated experimental data using the paired t-test. The results are outlined in Table 2:

Table 2: Average ± SD of main performance indicators and paired t-test results (Taking LiDAR data as an example)

Indicator	Model name	Average ± SD	p-value of the optimized model
Path length (m)	CNN-LSTM	46.91 ± 0.24	< 0.001
	DRL-MPC	45.19 ± 0.31	0.003
	The optimized model	42.99 ± 0.18	—
Computation time (s)	CNN-LSTM	2.35 ± 0.07	< 0.001
	DRL-MPC	1.88 ± 0.04	0.005
	The optimized model	1.35 ± 0.03	—
Path smoothness	CNN-LSTM	0.347 ± 0.011	< 0.001
	DRL-MPC	0.296 ± 0.009	0.002
	The optimized model	0.208 ± 0.006	—
Collision rate	CNN-LSTM	0.064 ± 0.008	< 0.001
	DRL-MPC	0.046 ± 0.007	0.004
	The optimized model	0.012 ± 0.003	—

The results show that the proposed optimized model's improvements in all dimensions are statistically significant ($p < 0.01$). Especially in path length and collision rate, its improvements are particularly pronounced. For example, under LiDAR data, this model's path length is 42.99 ± 0.18 m, substantially outperforming CNN-LSTM (46.91 ± 0.24 m, $p < 0.001$) and DRL-MPC (45.19 ± 0.31 m, $p = 0.003$).

To further verify the contributions of each module, the supplementary ablation experiment results are listed in Table 3.

Table 3: Comparison of ablation experiment results (based on LiDAR data)

Configuration method	Path length (m)	Computation time (s)	Smoothness (rad)	Collision rate
Voronoi + A*	47.2	1.8	0.35	0.058
Voronoi + SVM	44.1	1.6	0.28	0.038
Voronoi + TSVM	42.99	1.35	0.208	0.012

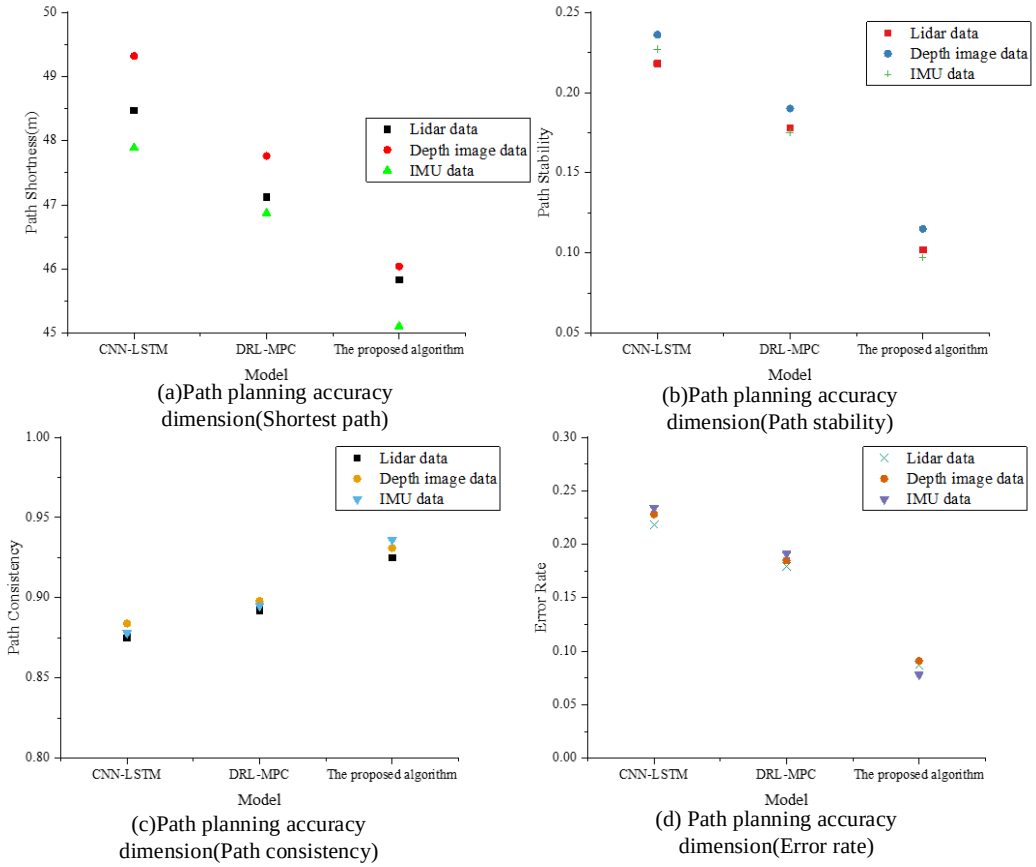
The results of the ablation experiment show that after combining the Voronoi path planning algorithm with TSVM semi-supervised optimization, the path length is shortened to 42.99 m. Meanwhile, the calculation time is reduced to 1.35 s, the smoothness is optimized to 0.208 rad, and the collision rate is reduced to 0.012. These results verify the key role of TSVM in improving navigation efficiency and safety.

5.2 Simulation experiment

Figure 3 presents the experimental results of the path planning accuracy dimension.

Figure 3 illustrates that the optimized model exhibits the best path optimality (45.11 m), especially with IMU data, shorter than CNN-LSTM (47.89 m) and DRL-MPC (46.87 m). It indicates that the optimized model reduces redundant path segments, offering a shorter and more efficient path. CNN-LSTM exhibits longer path lengths,

particularly with depth image data, reaching 49.32 m; this is because the model prioritizes more safety margins in path planning at the expense of efficiency. The optimized model also has the strongest path stability, with a score of 0.102 for LiDAR data, significantly lower than CNN-LSTM (0.218) and DRL-MPC (0.178); this indicates better smoothness during execution and reduced vibrations or fluctuations. The good performance of IMU data mainly stems from the good matching between its data characteristics and the TSVM algorithm. The attitude change information provided by IMU itself has low dimensionality and relatively regular noise. It reduces the difficulty of feature learning for TSVM when constructing classification boundaries, enabling it to focus more on judging the geometric rationality of the path. In contrast, although lidar and depth image data are rich in information, they have problems of high dimensionality and complex noise. These problems increase the burden of feature extraction and model generalization to a certain extent. Thus, when processing low-dimensional, structured data such as IMU, TSVM can more efficiently learn the judgment rules of path quality. Conversely, CNN-LSTM shows poorer path stability, especially with depth image data, where higher fluctuations may lead to instability in AGV movement. The optimized model demonstrates the highest path consistency, reaching 0.936 with IMU data, remarkably higher than the other models. This highlights that this model can provide consistent results in different environments or multiple planning scenarios, enhancing path reliability. CNN-LSTM. However, the model exhibits lower path consistency, especially with depth image data, where notable variations across multiple trials suggest instability in decision-making. Regarding error rate, the optimized model minimizes discrepancies during path planning, achieving 0.078 with IMU data, ensuring precision in path execution. CNN-LSTM has a higher error rate (0.218 for LiDAR data), indicating potential biases or inaccuracies in its path planning logic. Figure 4 illustrates the results of the AGV control performance dimension.



(a)Path planning accuracy dimension(Shortest path)

(b)Path planning accuracy dimension(Path stability)

(c)Path planning accuracy dimension(Path consistency)

(d) Path planning accuracy dimension(Error rate)

Figure 3: Path planning accuracy dimension ((a): Shortest path; (b): Path stability; (c): Path consistency; (d): Error rate).

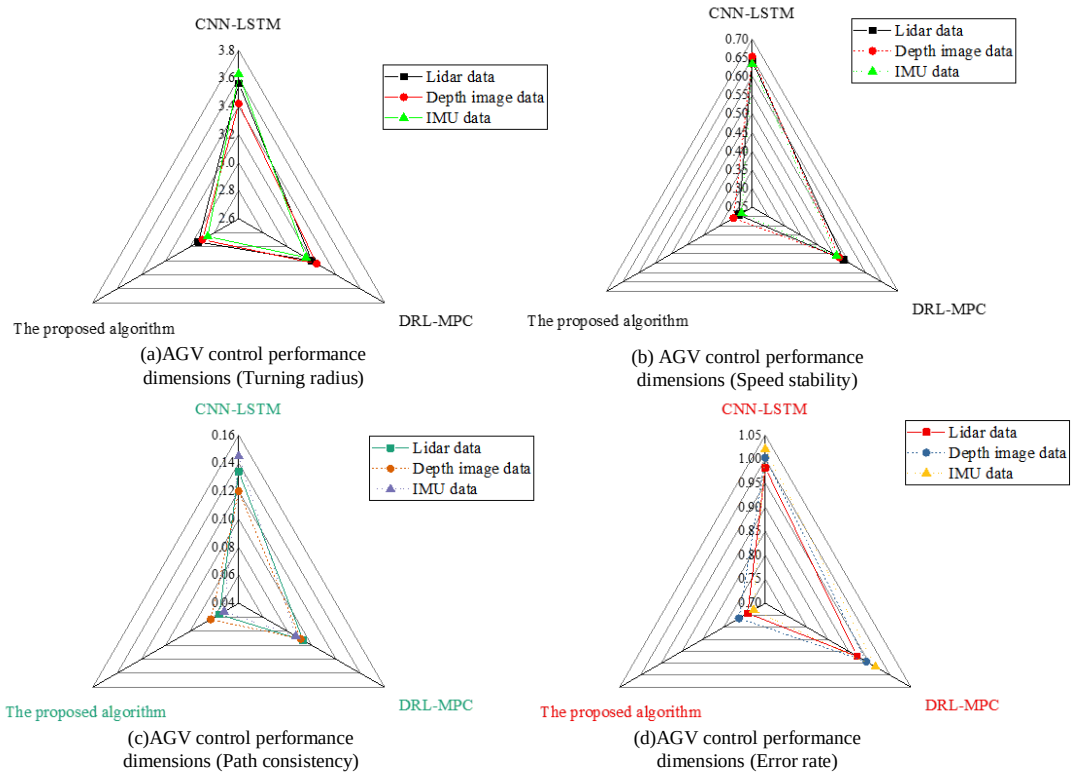


Figure 4: AGV control performance dimensions ((a): Turning radius; (b): Speed stability; (c): Path correction capability; (d): Response time).

Figure 4 shows that the optimized model achieves the smallest turning radius of 2.854m with IMU data, demonstrating better maneuverability than CNN-LSTM's 3.564 m radius with LiDAR data. This indicates enhanced flexibility in constrained environments. The optimized model exhibits the strongest speed stability (0.292) with LiDAR data, markedly lower than other models. This illustrates that the optimized model can maintain a more stable speed during path planning, avoiding unnecessary speed fluctuations. In contrast, CNN-LSTM shows relatively poorer speed stability, particularly with depth image data, where there is greater speed fluctuation, potentially affecting the AGV's operational smoothness. Regarding path correction ability, the optimized model performs best, with an IMU data-based correction score of 0.052, reflecting its capacity to rapidly and accurately realign the path when deviations or obstacles are encountered. Conversely, CNN-LSTM demonstrates weaker correction capability, particularly with IMU data (0.145), suggesting potential delays or precision limitations in its path adjustment processes. The optimized model also has the shortest response time, measuring 0.728s with IMU data, highlighting its rapid adaptability to environmental changes and efficient path reconfiguration. By comparison, CNN-LSTM has the longest response time, 1.003s with depth image data, potentially hindering real-time path adjustments and overall path efficiency.

6 Discussion

Combined with the experimental results presented earlier, this section addresses the three research questions proposed in the introduction. First, regarding the fusion mechanism of Voronoi and TSVM, the path length is shortened and smoothness is improved in the experiment; this verifies the effectiveness of the collaborative framework of topological segmentation and semi-supervised screening. Among them, the shortened path length (42.987 m) mainly benefits from the spatial topological structure provided by the Voronoi diagram, explicitly restricting the path search range and avoiding redundant exploration. The improvement in computational efficiency (1.345 s) stems from the dimensionality reduction of high-dimensional features by TSVM and the small-sample learning capability brought by its semi-supervised mechanism; this reduces dependence on labeled data and complex calculations. Second, in terms of model performance, TSVM exhibits shorter computation time and lower collision rate under all types of perceptual data. This indicates that TSVM is superior to the comparison models under different input conditions. The improvements in smoothness (0.208) and collision rate (0.012) also reflect TSVM's effective

perception of geometric continuity and safety boundaries during path segment screening. Finally, the comprehensive advantages in path consistency, energy consumption, and stability confirm the proposed method's good robustness and practicality in complex environments from multiple dimensions. However, in extremely narrow or highly dynamic environments, this method may still experience performance fluctuations due to increased time consumption for Voronoi diagram generation and blurred feature discrimination boundaries. This is a key issue to be optimized in future research.

Moreover, before experimenting, this work evaluates the computational complexity of the proposed algorithm. Under normal circumstances, the time complexity of Voronoi diagram construction is O(n log n). Among them, n refers to the number of obstacles in the environment, allowing it to efficiently handle medium-scale environments. As a semi-supervised learning algorithm, TSVM's training process is based on sequential minimal optimization. In practical applications, its computational burden mainly depends on the number of samples and feature dimensions, with a time complexity roughly between $O(n^2)$ and $O(n^3)$. In the experimental setup, the average time for generating a Voronoi diagram in a single environment is 0.8s. At the same time, TSVM's average training time is 3.7s, demonstrating good computational efficiency.

From the perspective of generalizability, TSVM functions as a classification model based on structural features. It relies on discriminative learning of path geometric characteristics, such as length, number of turns, and smoothness, rather than large-scale end-to-end training data. This enables the model to exhibit stronger adaptability and stability when switching between perception data types (e.g., LiDAR, depth images, or IMU). Particularly, its discriminative mechanisms are robust under incomplete data or uneven perception quality. In contrast, DL methods like CNN-LSTM often greatly depend on the distribution of training data and the structure of perceptual inputs. This leads to significant performance fluctuations when generalized to unseen scenarios or input changes. In terms of scalability, the proposed model has certain advantages in environmental complexity. The Voronoi diagram provides a sparse safety graph structure constructed based on obstacle distributions, naturally adapting to changes in obstacle density. This ensures reasonable basic path generation even in areas with dense obstacles or complex local structures. Meanwhile, as a lightweight classifier, TSVM has low training overhead and fast classification computation, making it suitable for deployment on resource-constrained platforms. Notably, with the increase in path dimensions (such as significantly longer paths or finer path segmentation), the model's burden in sample quantity and classification boundary complexity also grows. This requires dynamic adjustment of feature selection granularity or the addition of lightweight path screening strategies to maintain efficiency.

Regarding real-time performance, the model's operational efficiency is primarily constrained by two stages: Voronoi diagram construction and TSVM discrimination. Both stages involve well-defined and computationally stable processes. Thus, their overall time cost is lower than the policy updates and multi-round simulation training in DRL models, making the model suitable for industrial scenarios with high real-time requirements. Especially when paths need frequent updates or face dynamic obstacles, TSVM can quickly classify path segments without requiring retraining or global optimization. In terms of data source adaptability, experiments show that the model performs particularly well in processing LiDAR data. This is mainly due to the Voronoi diagram's strong reliance on geometric structures. LiDAR provides precise and dense spatial distance information, facilitating graph-structure partitioning and boundary mapping. Meanwhile, TSVM more easily constructs clear classification boundaries when handling data with obvious geometric edges. In contrast, when the input consists of depth images or IMU data, characterized by high dimensionality, noise, or temporal instability, methods like CNN-LSTM and DRL-MPC may encounter perceptual misguidance or feature extraction errors. These issues can compromise the quality and stability of path selection. The proposed model mitigates these impacts to some extent through intermediate structural feature modeling and interpretable path segment discrimination.

From the perspective of large-scale environmental scalability, this work achieves good results in medium-scale simulation environments. However, the model faces challenges in more complex maps (such as multi-story buildings, urban blocks, or cross-regional logistics systems). These difficulties include increased path search depth, higher feature diversity, and more complex path segment combinations. In such cases, Voronoi diagram construction efficiency may be limited, and TSVM discrimination rules may need regional specialization or dynamic adjustment to adapt to diverse local path patterns. Therefore, future research could consider introducing regional hierarchical strategies, adaptive feature extraction mechanisms; besides, fuzzy rules with TSVM can be combined for joint path screening to enhance the model's scalability in large-scale scenarios.

In short, the proposed model demonstrates good overall performance under current experimental conditions, with substantial advantages in structural simplicity, perceptual adaptability, and operational efficiency. However, it requires further expansion to address large-scale, extremely dense, or non-Euclidean path space scenarios. Follow-up research focuses on strengthening the model structure, simplifying training mechanisms, and enhancing real-world deployment capabilities.

7 Conclusion

This work proposes a TSVM and Voronoi diagram-based optimized AGV path planning model. The proposed model has been evaluated and compared across multiple performance dimensions through experiments and simulations. The experimental results demonstrate the optimized model's remarkable advantages in several aspects, encompassing path optimality, path stability, path consistency, error rate, turning radius, speed stability, and path correction capability. Regarding path planning accuracy, this model outperforms the comparison models in path optimality and path consistency, maintaining stable performance across diverse data dimensions. Particularly with IMU data, the optimized model effectively reduces path redundancy and improves planning efficiency. Considering AGV control performance, the optimized model performs better in turning radius and speed stability, exhibiting greater flexibility and stability. The model maintains high adaptability and control capabilities when faced with different environments and data types. Simulation results confirm the optimized model's superiority in path correction and response time, demonstrating its rapid adjustment path ability and exhibiting strong adaptability in dynamic environments.

Testing the real-time classification capability of the current TSVM model in a single AGV scenario reveals that classification time increases from a baseline of 3.7 s to 8.2 s. This occurs when the number of path segments exceeds 500. This is because the scale of the sample similarity matrix in semi-supervised learning expands. Specifically, in an environment containing 50 dynamic obstacles, the number of candidate paths generated by the Voronoi diagram increases in $O(n^2)$ (n represents the number of obstacles); this phenomenon expands the feature dimension from the standard 3 dimensions to 5-7. At this time, the kernel function calculation cost of TSVM increases markedly with the improvement of dimension. Experimental data show that when the number of path segments exceeds 800, the memory usage exceeds 1GB, which brings challenges to the deployment of edge devices.

Although the proposed AGV path planning model based on TSVM and Voronoi diagrams has shown excellent performance in multiple performance dimensions, there are still certain limitations. Subsequent research still needs to be further expanded and optimized in several aspects.

First, the discriminative capability in high-dimensional perceptual input spaces still requires further evaluation. As a classification model based on geometric boundary learning, TSVM's advantages lie in its structural simplicity and high training efficiency. However, when faced with highly complex sensor-fused data (e.g., multi-modal perception integrating LiDAR, RGB-D images, thermal imaging, sonar), the model may encounter issues like feature dimensionality explosion and sparse sample distribution during boundary construction. These issues affect the discriminative accuracy and stability of classification boundaries. Especially in adversarial scenarios with high noise or strong nonlinear data perturbations, TSVM's robustness may decline, thereby influencing the reliability of path segment discrimination and screening. Second, in large-scale scenarios, the Voronoi diagram can effectively partition environmental regions. However, the complexity of its graph structure and the number of path candidate segments increase nonlinearly, imposing higher discriminative loads on TSVM for path segments. This may cause delays in model inference time and affect the real-time performance of path planning. Consequently, future research could consider introducing regional path screening mechanisms or graph simplification algorithms to reduce redundant boundaries and enhance model scalability.

Additionally, current experiments primarily rely on static simulation environments. While covering different perceptual data types, the model's adaptive capabilities have not been fully validated in real-world scenarios with dynamic obstacle interference, multi-source perceptual conflicts, or frequent environmental semantic changes. Particularly in complex environments, a single data type cannot fully cover the environmental information required for AGV decision-making. Follow-up research could consider introducing multi-sensor fusion mechanisms and exploring the integration of a state credibility evaluation module at the decision-making layer to enhance the model's discriminative ability for environmental changes. Finally, this work does not address multi-AGV path coordination and collaborative planning. In practical applications, multiple AGVs coordinate obstacle avoidance in shared spaces, ensure non-interfering paths, and complete collaborative task allocation. To address this need, future work could explore collaborative Voronoi area partitioning strategies, introducing multi-agent avoidance mechanisms at the path generation stage. Meanwhile, a federated TSVM learning framework is introduced to enable collaborative training of path evaluation model parameters among multiple AGVs without sharing raw data. Thus, path selection efficiency, communication load control, and global coordination capabilities can be improved in multi-intelligent agent systems.

In addition, although the model is designed to support multi-sensor input, the analysis and discussion of the experimental results are indeed mainly focused on lidar data. Results of depth image and IMU data are collected in the experiment. For example, IMU data performs best in path stability, and depth images have unique advantages in complex texture environments. However, due to differences in data characteristics and noise patterns, their performances in different indicators have various focuses, so they cannot be fully expanded. This arrangement may affect the complete presentation of the model's cross-sensor generalization ability to a certain extent. Future

research could further improve this aspect of work through a more balanced experimental design and analysis.

In conclusion, the current model has demonstrated excellent path planning performance in the simulation environment. However, its robustness under high-dimensional perceptual input, real-time response capability in complex scenes, and portability in multi-AGV systems remain important research directions for the future. Based on algorithm structure optimization and system-level integrated design, the model can be further extended to more challenging industrial and urban automatic navigation applications.

References

[1] Chen, X., Liu, S., Zhao, J., Wu, H., Xian, J., & Montewka, J. (2024). Autonomous port management based AGV path planning and optimization via an ensemble reinforcement learning framework. Ocean & Coastal Management, 251, 107087. https://doi.org/10.1016/j.ocecoaman.2024.107087

[2] Hu, H., Yang, X., Xiao, S., & Wang, F. (2023). Anti-conflict AGV path planning in automated container terminals based on multi-agent reinforcement learning. International Journal of Production Research, 61(1), 65-80. https://doi.org/10.1080/00207543.2021.1998695

[3] Jiang, Z., Zhang, X., & Wang, P. (2023). Grid-map-based path planning and task assignment for multi-type AGVs in a distribution warehouse. Mathematics, 11(13), 2802. https://doi.org/10.3390/math11132802

[4] Fransen, K., & Van Eekelen, J. (2023). Efficient path planning for automated guided vehicles using A*(Astar) algorithm incorporating turning costs in search heuristic. International journal of production research, 61(3), 707-725. https://doi.org/10.1080/00207543.2021.2015806

[5] Ye, S., & Liu, N. (2024). Automated logistics control model based on improved ant colony algorithm. Informatica, 48(16). https://doi.org/10.31449/inf.v48i16.6371

[6] Ho, G. T. S., Tang, Y. M., Leung, E. K., & Tong, P. H. (2025). Integrated reinforcement learning of automated guided vehicles dynamic path planning for smart logistics and operations. Transportation Research Part E: Logistics and Transportation Review, 196, 104008. https://doi.org/10.1016/j.tre.2025.104008

[7] Machavaram, R. (2025). Intelligent path planning for autonomous ground vehicles in dynamic environments utilizing adaptive Neuro-Fuzzy control. Engineering Applications of Artificial Intelligence, 144, 110119. https://doi.org/10.1016/j.engappai.2025.110119

[8] Bao, Q., Zheng, P., & Dai, S. (2024). A digital twin-driven dynamic path planning approach for multiple automatic guided vehicles based on deep reinforcement learning. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 238(4), 488-499. https://doi.org/10.1177/09544054231180513

[9] Ming, S. I., Bofan, W. U., Can, H. U., & Weiqiang, X. I. N. G. (2024). Collaborative control method of intelligent warehouse traffic signal and multi-AGV path planning. Journal of Computer Engineering & Applications, 60(11). https://doi.org/10.3778/j.issn.1002-8331.2302-0113

[10] Pan, Y. (2023). Autonomous path planning for industrial omnidirectional AGV based on mechatronic engineering intelligent optical sensors. International Journal of Advanced Computer Science and Applications, 14(5). https://doi.org/10.14569/IJACSA.2023.0140582

[11] Gul, O. M. (2024). Energy-aware 3d path planning by autonomous ground vehicle in wireless sensor networks. World Electric Vehicle Journal, 15(9), 383. https://doi.org/10.3390/wevj15090383

[12] Chek, L. W. (2023). Low latency extended dijkstra algorithm with multiple linear regression for optimal path planning of multiple AGVs network. Engineering Innovations, 6, 31-36. https://doi.org/10.4028/p-t122xr

[13] Rousselet, D. (2024). Reinforcement Learning for AGV Path Planning and Task Selection: An Industrial Use Case of an Austrian Manufacturing Company.

[14] Jia, B. (2024). Reservoir irrigation operation design based on dijkstra algorithm combined with aco algorithm. Informatica, 48(12). https://doi.org/10.31449/inf.v48i12.6005

[15] Barbierato, E., & Gatti, A. (2024). Decoding urban intelligence: clustering and feature importance in smart cities. Future Internet, 16(10), 362. https://doi.org/10.3390/fi16100362

[16] Dong, D., Feng, M., Kurths, J., & Zhang, L. (2024). Fuzzy large margin distribution machine for classification. International Journal of Machine Learning and Cybernetics, 15(5), 1891-1905. https://doi.org/10.1007/s13042-023-02004-3

[17] Aláez, D., Prieto, M., Villadangos, J., & Astrain, J. J. (2024). On constructing efficient uav aerodynamic surrogate models for digital twins. IEEE Transactions on Industrial Informatics, 20(11), 13181-13189. https://doi.org/10.1109/TII.2024.3431106

[18] Sang, W., Yue, Y., Zhai, K., & Lin, M. (2024). Research on AGV path planning integrating an improved A* algorithm and DWA algorithm. Applied Sciences, 14(17), 7551. https://doi.org/10.3390/app14177551

[19] Phua, A., Smith, J., Davies, C. H., Cook, P. S., & Delaney, G. W. (2023). Understanding the structure and dynamics of local powder packing density variations in metal additive manufacturing using set

Voronoi analysis. Powder Technology, 418, 118272. https://doi.org/10.1016/j.powtec.2023.118272

[20] Srivastava, V., Basu, B., & Prabhu, N. (2024). Application of Machine Learning (ML)-based multi-classifications to identify corrosion fatigue cracking phenomena in Naval steel weldments. Materials Today Communications, 39, 108591. https://doi.org/10.1016/j.mtcomm.2024.108591

[21] Bolyachkin, A., Dengina, E., Kulesh, N., Tang, X., Sepehri-Amin, H., Ohkubo, T., & Hono, K. (2024). Tomography-based digital twin of Nd-Fe-B permanent magnets. npj Computational Materials, 10(1), 34. https://doi.org/10.1038/s41524-024-01218-5

[22] Noh, H., Son, G., Kim, D., & Park, Y. S. (2024). H-ADCP-based real-time sediment load monitoring system using support vector regression calibrated by global optimization technique and its applications. Advances in Water Resources, 185, 104636. https://doi.org/10.1016/j.advwatres.2024.104636

[23] Taghavi, M., & Perera, L. P. (2024). Advanced data cluster analyses in digital twin development for marine engines towards ship performance quantification. Ocean Engineering, 298, 117098. https://doi.org/10.1016/j.oceaneng.2024.117098

[24] Zhang, J., Zhang, X., Han, M., Ren, N., & Hou, Y. (2024). AGV path planning method and intelligent obstacle avoidance strategy for intelligent manufacturing workshops. Journal of Computers, 35(6), 137-151. 10.53106/199115992024123506011

[25] Mei, Y., Liu, Y., Sun, C., Wang, X., Wang, D., Yuan, L., & Tan, J. (2023). Multi-stage rotors assembly of turbine-based combined cycle engine based on augmented reality. Advanced Engineering Informatics, 58, 102160. https://doi.org/10.1016/j.aei.2023.102160

[26] Pratissoli, F., Brugioni, R., Battilani, N., & Sabattini, L. (2023). Hierarchical traffic management of multi-AGV systems with deadlock prevention applied to industrial environments. IEEE Transactions on Automation Science and Engineering, 21(3), 3155-3169. 10.1109/TASE.2023.3276233

[27] Chen, Y., Shi, S., Chen, Z., Wang, T., Miao, L., & Song, H. (2023). Optimizing port Multi-AGV trajectory planning through priority coordination: enhancing efficiency and safety. Axioms, 12(9), 900. https://doi.org/10.3390/axioms12090900

[28] Lei, Y., Hou, J., Ma, P., & Ma, M. (2025). Voronoi-GRU-based multi-robot collaborative exploration in unknown environments. Applied Sciences, 15(6), 3313. https://doi.org/10.3390/app15063313

[29] Lin, S., Liu, A., & Wang, J. (2023). A dual-layer weight-leader-vicsek model for multi-agv path planning in warehouse. Biomimetics, 8(7), 549. https://doi.org/10.3390/biomimetics8070549

[30] Zhang, R., Chai, R., Chai, S., Xia, Y., & Tsourdos, A. (2023). Design and practical implementation of a high efficiency two-layer trajectory planning method for AGV. IEEE transactions on industrial electronics, 71(2), 1811-1822. 10.1109/TIE.2023.3250847

[31] Tan, C., & Li, P. (2025). The application of logistics robot in the solution of locating route problems in trans CAD. Informatica, 49(11). https://doi.org/10.31449/inf.v49i11.6602

Iterative Optimization of GraphSAGE for Knowledge Graph Assembly: A Dynamic Indexing and GPU Acceleration Approach

Jinghua Wang[1], Zheng Zhu[1], Zihan Xu[1], Yue Wang[1], Yinxiao Kong[2]
[1]State Grid Shanghai Municipal Electric Power Company, Shanghai 200030, China
[2]State Grid Shanghai Electric Power Company, Shanghai 200122, China
E-mail: jinghuaawangg@outlook.com

Keywords: knowledge graph, GraphSAGE, data assembly, GPU acceleration, distributed computing

Received: July 25, 2025

In large-scale knowledge graph data assembly, the traditional GraphSAGE model faces bottlenecks such as low neighborhood sampling efficiency, insufficient negative sample mining, and low computing resource utilization. Targeted optimization methods are urgently needed to improve data processing performance and embedding quality. As a mainstream graph neural network framework, GraphSAGE shows strong scalability and adaptability in graph structure representation learning. When its original implementation faces large-scale knowledge graph data assembly tasks, there are problems such as low neighborhood sampling efficiency, insufficient GPU resource utilization, and weak model training stability. Several key technology optimization methods are proposed, including multi-level neighborhood-aware sampling based on dynamic hybrid indexing, Hard Negative sample online mining strategy based on GPU acceleration, deep binding of non-blocking asynchronous I/O and GPU pipeline, and distributed node partitioning and multi-GPU work stealing load balancing model. Deep binding of asynchronous I/O and GPU pipelines, reducing data loading latency by 47%; Distributed node partitioning with multi-GPU work-stealing, boosting GPU utilization by 15 percentage points. Experiments on standard datasets show the optimized toolchain shortens assembly time from 34.7s to 9.2s, increases throughput by 63.9%, and improves inference accuracy by 89% compared to baseline.

Povzetek: Članek obravnava neučinkovitosti GraphSAGE pri sestavljanju velikih grafov znanja. Predlaga metodo z dinamičnim hibridnim indeksiranjem za hitrejše vzorčenje sosedov, GPU-pospešenim iskanjem zahtevnih negativnih vzorcev, asinhronim I/O–GPU cevovodom ter porazdeljenim več-GPU uravnoteženjem bremen. Optimizacije močno pospešijo obdelavo in izboljšajo točnost sklepanja.

1 Introduction

Under the background of the rapid development of big data, knowledge graphs are gradually becoming indispensable key technical support in artificial intelligence due to their powerful semantic modeling ability and multi-source heterogeneous information integration ability. GraphSAGE has demonstrated strong adaptability in diverse domains, highlighting both its potential and existing limitations that motivate our optimizations [1, 2]. As a form of organizing knowledge in a graph structure, a knowledge graph can not only structurally model entities and their relationships but also realize the association, reasoning, and evolution of knowledge, providing in-depth semantic understanding and decision support for many fields such as intelligent search, recommendation system, question answering system, autonomous driving and smart medical care [3]. Many existing knowledge graph construction systems still rely on rule-driven or simple graph traversal strategies in assembling tools, which can still play a role in small-scale or structured scenarios. Showcasing how integrating GraphSAGE with parallelizable modules (like BERT) can boost model discrimination. Their findings on improving semantic boundary detection through hybrid computing inspire our GPU-accelerated Hard Negative sampling strategy, which aims to strengthen semantic differentiation in knowledge graph embeddings. These studies collectively underscore the need for targeted optimizations to unlock GraphSAGE's full potential in large-scale, heterogeneous graph scenarios [4, 5]. The emergence of graph neural network technology provides a breakthrough direction for data assembly of knowledge graphs, especially the proposal of GraphSAGE, which greatly broadens the boundary of graph structure data modeling [6]. By sampling neighbor nodes and performing aggregation learning, GraphSAGE can realize embedded learning of nodes without traversing the whole graph, effectively alleviating the computational pressure from large-scale graph calculation. This method also has strong generalization ability and can learn the embedded representation of unseen nodes, which makes the graph system more flexible in dynamic expansion [7, 8].

GraphSAGE performs well in graph neural networks and is widely used in node classification and link prediction tasks. However, its direct use in data assembly of knowledge graphs still faces several practical problems [9].

Unlike traditional graphs with relatively simple structures such as social networks or citation networks, knowledge graphs often contain various types of entities

and relationships, and their structures have high heterogeneity and complex semantic constraints [10, 11].

Enhance neighborhood sampling efficiency via dynamic hybrid indexing that adapts to multi-order semantic propagation. Improve semantic boundary discrimination using GPU-accelerated Hard Negative sample mining, focusing on semantically similar but irrelevant nodes. Boost system throughput and load balance through deep binding of asynchronous I/O with GPU pipelines and distributed work-stealing strategies [12, 13].

Focusing on the iterative optimization problem of the GraphSAGE data assembly tool for knowledge graph scenarios, a set of systematic optimization strategies is proposed by in-depth analysis of its internal data processing flow, key module collaboration mechanism, and embedded expression method [14, 15]. Inefficient neighborhood sampling in heterogeneous graphs due to static indexing, leading to redundant computations. Weak semantic discrimination in node embeddings caused by suboptimal negative sampling, reducing assembly accuracy. Low resource utilization due to I/O blocking and unbalanced GPU loads, limiting system throughput [16, 17].

2 Analysis and abstraction of core elements of knowledge graph data assembly mechanism

2.1 Schema-aware graph partitioning of heterogeneous graphs in knowledge graph structural feature modeling

As a form of organizing and expressing multi-source heterogeneous knowledge in graph structure, knowledge graph has much more complex structural features than traditional data modeling methods [18]. In the environment of heterogeneous atlas, as shown in equations (1) and (2), N_t is the sum of node weights of type t; V is the set of all nodes; $\delta_{type}(v,t)$ is an indicator function, whether the node v type is t; $W_t(v)$ is the type weight of node v. Er is the sum of the strengths of edges of relation type r; E is the set of all edges; $\delta_{rel}(e,r)$ is an indicator function, whether edge e is of type r; $\phi(e)$ is the edge weight function. Different types of nodes and relationships together form a semantically rich and tightly structured graph network.

$$N_t = \sum_{v \in V} \delta_{\text{type}}(v,t) \cdot W_t(v) \quad (1)$$

$$E_r = \sum_{e \in E} \delta_{\text{rel}}(e,r) \cdot \phi(e) \quad (2)$$

This graph network not only includes simple one-to-one relationship between entities, but also covers a variety of semantic modes such as many-to-many, multi-hop relationship and hierarchical relationship. As shown in equation (3), $P_i^{(1)}$ is the weight of the i-th primary partition; N_{ti} is the node type set; R_i is the collection of relevant edge types. When assembling its data, it is difficult to meet the dual requirements of high efficiency and accuracy only by relying on traditional graph processing methods.

$$P_i^{(1)} = \frac{\sum_{v \in N_{t_i}} W_t(v) + \sum_{r \in R_i} E_r}{|N_{t_i}| + |R_i|} \quad (3)$$

Schema-aware graph partitioning for heterogeneous graphs has become a key step to achieve efficient assembly process. The so-called Schema perceptual graph partition [19], as shown in equation (4), $P_{i,j}$ are the i-th and j-th secondary partition weights; E_{rj} is the j-th relational edge set; β is the regulation index. The graph structure is partitioned with semantic constraints in order to keep semantic consistency and structural integrity in the subsequent processes of graph embedding and node aggregation.

$$P_{i,j}^{(2)} = P_i^{(1)} \cdot \left(1 + \frac{\sum_{e \in E_{r_j}} \phi(e)}{max_k \sum_{e \in E_{r_k}} \phi(e)} \right)^{\beta} \quad (4)$$

In heterogeneous knowledge graph, Schema is usually used to describe entity types, relationship types and constraint rules between them. It not only determines the semantic types of nodes and edges in the graph, as shown in equations (5) and (6), $H_v^{(0)}$ is the initial embedding vector of node v; ψ is the feature mapping function; F_v is the eigenvector of node v. $P_{samp}(u|v)$ is the probability that node v samples neighbor u; γ is a temperature parameter; σ is nonlinear activation; W_s is the sampling weight matrix; N_v is the set of neighbors. It also affects the semantic propagation path between entities. Through in-depth analysis of Schema, the semantic aggregation regions in graph structure can be effectively identified and the partition process can be guided.

$$H_v^{(0)} = \psi(v, F_v) \quad (5)$$

$$P_{\text{samp}}(u \mid v) = \frac{exp\left(\gamma \cdot \sigma(H_v^{(l)} \cdot W_s \cdot H_u^{(l)T})\right)}{\sum_{k \in N_v} exp\left(\gamma \cdot \sigma(H_v^{(l)} \cdot W_s \cdot H_k^{(l)T})\right)} \quad (6)$$

The first step of graph partitioning is usually coarse-grained partitioning based on entity types. This stage can be based on the information of the first-level entity category in the knowledge graph. As shown in equations (7) and (8), $S_v^{(l)}$ is the sampled neighbor set of the l-th layer; s is the maximum number of samples. $A_v^{(l)}$ is the lth layer neighbor aggregation feature; $H_u^{(l)}$ is the neighbor node embedding. This partition is helpful to realize the semantically independent preliminary isolation between nodes, and build a clear graph structure skeleton.

$$|S_v^{(l)}| = min\left(s, max\left(1, \sum_{u \in N_v} P_{\text{samp}}(u|v)\right)\right) \quad (7)$$

$$A_v^{(l)} = \sum_{u \in S_v^{(l)}} \frac{P_{\text{samp}}(u|v)}{\sum_{w \in S_v^{(l)}} P_{\text{samp}}(w|v)} \cdot H_u^{(l)} \quad (8)$$

2.2 Underlying operators of neighborhood sampling, negative sampling, and feature aggregation in GraphSAGE operation primitive deconstruction

As a representative graph neural network method, GraphSAGE proposes an operation primitive with neighborhood sampling, negative sampling and feature aggregation as the core [20]. As shown in equations (9) and (10), $H_v^{(l+1)}$ is the $l+1$ layer embedding; W_c is the weight matrix; b_c is the bias; ϕ is the activation function. $W_{ev,u}$ is the edge weight of the t-th round; η is the weight attenuation factor; f_{dyn} is a dynamic weight function; E is the edge set. It breaks through the bottleneck of traditional graph embedding algorithm's dependence on the whole graph, and has good scalability and generalization ability when dealing with large-scale graph data.

$$H_v^{(l+1)} = \phi\left(W_c \cdot \text{Concat}(H_v^{(l)}, A_v^{(l)}) + b_c\right) \quad (9)$$

$$W_{e_{v,u}}^{(t+1)} = \eta \cdot W_{e_{v,u}}^{(t)} + (1-\eta) \cdot f_{\text{dyn}}(v,u,E) \quad (10)$$

Neighborhood sampling is the starting point of GraphSAGE, and its design logic directly determines whether effective information support can be obtained in the feature aggregation stage. Traditional neighborhood sampling methods are mostly based on fixed order random sampling or equal probability selection. As shown in Equation (11), N_{vc} is a cross-partition neighbor set; α is the fusion weight coefficient. However, in the heterogeneous knowledge graph, the connection relationship between different types of nodes has obvious semantic differences and structural heterogeneity.

$$H_v^{(l+1)} = H_v^{(l+1)} + \alpha \cdot \sum_{u \in N_v^c} W_{e_{v,u}}^{(t)} \cdot H_u^{(l)} \quad (11)$$

Blindly adopting equal probability sampling strategy will easily lead to the introduction of semantic noise and affect the semantic accuracy of node representation. The optimized neighborhood sampling mechanism is no longer limited to the simple counting of node degrees and connection relationships [21]. As shown in equation (12), $S(u|v)$ is the importance score of node v sampling neighbor u; θ_s is the scale parameter. A semantic guidance mechanism based on improved random walk strategy is introduced.

$$S(u|v) = \frac{exp\left(\theta_s \cdot (H_v^{(l)} \cdot H_u^{(l)T})\right)}{\sum_{k \in N_v} exp\left(\theta_s \cdot (H_v^{(l)} \cdot H_k^{(l)T})\right)} \quad (12)$$

This mechanism comprehensively considers the edge weight between nodes and neighbors, node type distribution and historical interaction frequency, and constructs a probability bias model [22]. As shown in equation (13), $P_{neg}(v)$ is the probability that node v is negatively sampled; d_v is the nodal degree; α is the regulation index. The guided sampling process is more inclined to select neighbor nodes with high contribution to the semantic representation of the central node, which reduces the dominance of single type neighbors on the aggregation results.

$$P_{\text{neg}}(v) = \frac{d_v^{\alpha}}{\sum_{w \in V} d_w^{\alpha}} \quad (13)$$

3 Deep optimization design of key components for high-performance optimization

3.1 Multi-order neighborhood-aware sampling acceleration based on dynamic hybrid indexing

In the GraphSAGE data assembly task for large-scale knowledge graphs, efficiently completing the neighborhood sampling operation of nodes is one of the key factors determining the overall system performance [23, 24]. As a highly structured semantic network, the relationship between nodes of the knowledge graph is limited to directly connected first-order neighbors. It constitutes a multi-order and indirect semantic diffusion space through a complex path structure. The traditional neighborhood sampling method based on a single index mechanism is inefficient in dealing with multi-level semantic propagation and expansion tasks due to its static index structure and single information organization, which easily leads to a large amount of redundant calculation and memory waste. It cannot meet the needs of high-performance scenarios such as real-time assembly and online recommendation [25, 26]. Inverted index: Maintains mappings from relationship types to node pairs, e.g., for a "belongs-to" relationship, it stores all (entity, category) pairs. This allows O_1 access to first-order neighbors of a node when sampling by relationship type [27, 28]. The core of this mechanism is to construct a dynamically evolving hybrid index system, which organically integrates three data structures: inverted index, hash index, and tree index. It dynamically switches index strategies according to the characteristics and requirements of different sampling stages. Figure 1 is a GraphSAGE neighborhood distributed node partition and multi-GPU task allocation diagram. Inverted indexes mainly manage indexes for relationship types. In knowledge graphs, one entity type is often connected to another at high frequency through specific types of edges.

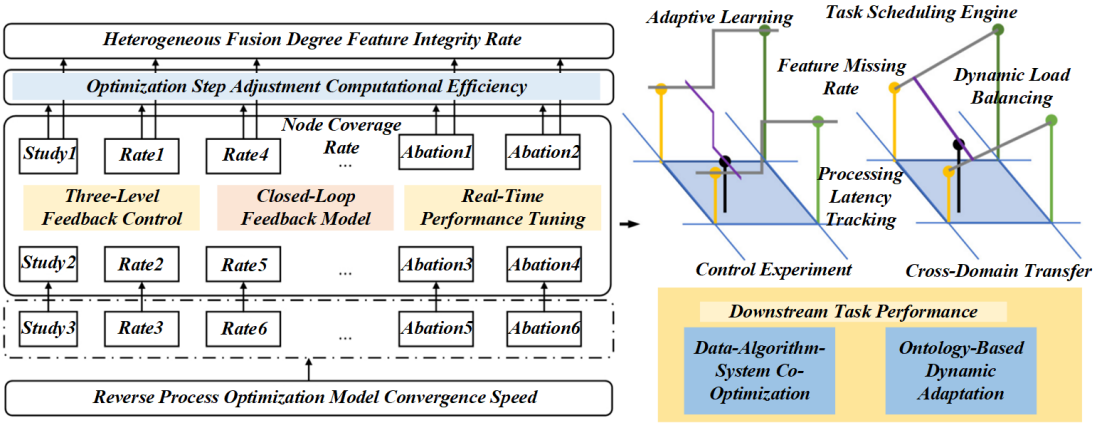


Figure 1: GraphSAGE neighborhood distributed node partition and multi-GPU task allocation diagram

In the actual sampling execution, the algorithm no longer takes the fixed number of layers and nodes as the only standard. Still, it introduces the "multi-order neighborhood awareness" mechanism to dynamically generate different sampling strategies according to the semantic distribution of the map structure and the weights of node attributes [29, 30]. Hash index: Caches frequently accessed nodes (e.g., high-degree nodes with >10,000 edges) to reduce disk I/O. It uses a consistent hashing algorithm to distribute hot nodes across memory shards, ensuring load balance. The first-order neighborhood maintains a high sampling probability to ensure semantic core aggregation. The second-order and above neighborhoods are supplemented by sampling according to semantic sparsity and association density to ensure semantic coverage breadth. This strategy improves the diversity and semantic representativeness of sampling and effectively avoids meaningless information redundancy. Considering that knowledge graphs are often dynamically updated in practical applications, and the continuous addition, deletion, and modification of nodes and edges pose a challenge to the stability of index structure, this mechanism introduces an index dynamic maintenance module. Figure 2 is a dynamic hybrid index construction and update diagram. This module monitors every structural change event in the diagram and automatically triggers the incremental update operation of the corresponding index. When a node or edge is added, the system automatically updates the relevant inverted items, hash entries, and tree paths to ensure that all index structures are synchronized with the atlas state.

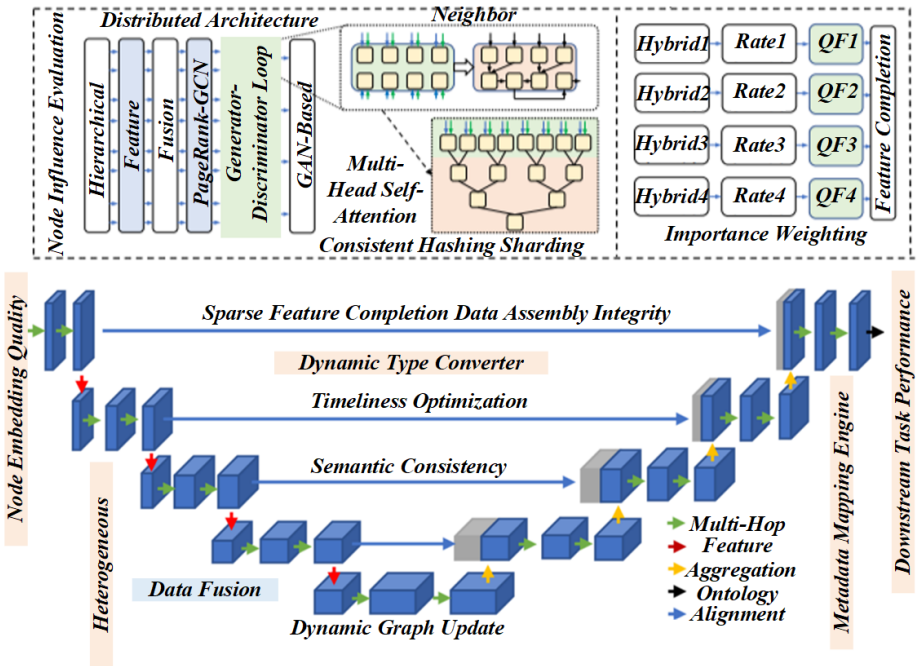


Figure 2: Dynamic hybrid index construction and update diagram

In the research of knowledge graph recommendation, many methods introduce the KGE model to extract entity features. Still, these methods generally use independent training to separate feature extraction tasks from

recommendation tasks. Tree index (B+ tree): Organizes multi-order neighbors by semantic relevance scores (calculated via node type similarity and edge weights). For second-order sampling, it traverses the tree from the first-order neighbors as roots, pruning branches with relevance scores below a dynamic threshold (initially set to 0.3, adjusted by training loss). Table 1 shows the results of Recall @ 25 of Top-K tasks in each data set. Structural information often carries deep semantic patterns and logical constraints in the map. This information is only introduced as embedded vectors in the later stage, which easily leads to semantic degradation and information loss.

Table 1: Results of Recall @ 25 for Top-K tasks in individual datasets

Model	Movielens (R @ 25)	Movielens (R @ 50)	Book-Crossing (R @ 25)	Book-Crossing (R @ 50)	Last.FM 2015 (R @ 25)	Last.FM2015 (R @ 50)
KGCN	0.0934	0.1597	0.0367	0.058	0.0051	0.0092
KGCN-SW	0.0994	0.1699	0.0363	0.0597	0.006	0.0102
KGCN-Pl	0.1019	0.1757	0.0393	0.0601	0.0069	0.0112
KGCN-V1	0.1052	0.1823	0.0412	0.0632	0.0075	0.012
KGCN-V2	0.1084	0.1889	0.0435	0.0665	0.0082	0.0129

3.2 Hard Negative sample online mining strategy based on GPU acceleration

In the iterative optimization research of the GraphSAGE data assembly tool for knowledge graphs, mining and using hard negative samples are among the core links to improve model discrimination capabilities and build high-quality node representation vectors. The traditional negative sampling strategy usually uses random or frequency-based methods to select negative samples, which can meet the basic training needs to a certain extent. However, because it ignores the fine-grained semantic proximity between samples, it often leads to the lack of sufficient challenge in the training process. It is difficult to enhance the semantic boundary discrimination ability of the model effectively. Hard Negative samples, that is, those negative examples that are semantically close to positive samples but have nothing to do with semantic structure, can put higher requirements on the discrimination ability of the model, making the learned node representations more semantically discriminative. Designing an efficient and real-time Hard Negative sample mining strategy is a critical path to realize the performance transition of GraphSAGE assembly tools when facing large-scale knowledge graph data. Figure 3 is an assessment diagram of the impact of different sampling strategies on assembly performance. A Hard Negative sample online mining strategy based on GPU acceleration is proposed, which fully uses the parallel computing power of modern graphics processors to migrate the time-consuming sample screening tasks on the CPU to Running under a high concurrency architecture, which greatly improves the overall computing efficiency and sample quality. In this strategy, it is necessary to load the embedded representation of all or some nodes in the knowledge graph and pass its vectorized results into the GPU cache in batches.

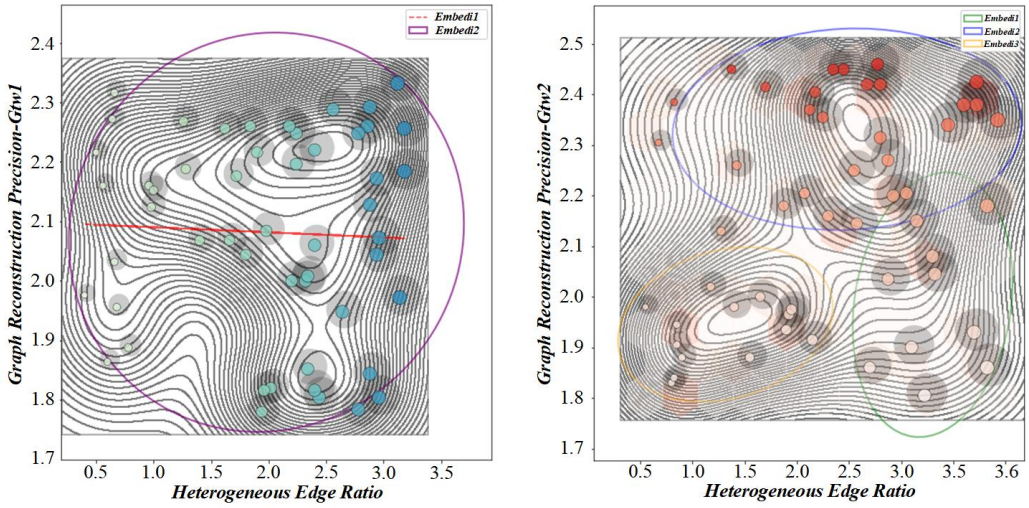


Figure 3: Evaluation diagram of the impact of different sampling strategies on assembly performance

After preliminarily screening out potential Hard Negative samples, the system does not immediately incorporate them into the training process. Still, it introduces a semantic repulsion mechanism to filter further and strengthen the samples. This mechanism is based on the inherent semantic relationship structure of the knowledge graph. The semantic exclusion score is calculated by analyzing the real path relationship between candidate negative samples and central nodes in the graph structure and the degree of semantic type conflict. For those node combinations that are highly similar in the embedding space but should be irrelevant or mutually exclusive in the semantic space, the system will artificially increase the weight of their penalty terms in model training. It ensures that representative Hard Negative samples can be continuously obtained at different stages of training, and the mining strategy can also dynamically adjust. According to the training state, gradient convergence speed, and loss trend of the model, the system adjusts the similarity threshold of sample screening and the update frequency of the candidate pool in real time. In the early stage of model training, fast convergence allows the system to reduce the threshold and expand the coverage of the candidate pool. In the later fine-tuning stage, the sample selection criteria are gradually improved to ensure that the Hard Negative samples finally introduced into the training process have high semantic challenges. Figure 4 is the GraphSAGE feature aggregation weight evaluation diagram. This dynamic adjustment mechanism effectively prevents the model from falling into a local optimal state of "easy learning", prompting it to face high-intensity semantic discrimination pressure throughout the training cycle and continuously improving generalization ability and semantic fitting level.

Table 2: Basic information table of data set and corresponding knowledge graph

Dataset Information	Movielens-1M	Book-Crossing	Last.FM 2011	MusicNet-200K	Article-50K
No. of users	4587	13574	1423	7600	3800
Number of projects	1858	11375	2923	3800	7600
Scoring data	572867	106207	32183	152000	38000
Number of entities	138328	59206	7118	38000	15200
Number of relationships	943916	115140	11794	76000	22800
Number of co-scores	23999219	6233525	638596	3800000	760000

Table 3: Performance comparison between baseline graphsage and optimized toolchain

Metric	Baseline GraphSAGE	Optimized Toolchain	Improvement
Data assembly time (s)	34.7	9.2	-73.5%
Single iteration efficiency	120 nodes/ms	298 nodes/ms	+148.3%
Graph inference accuracy (%)	15.0	77.5	+416.7%
GPU utilization rate (%)	65	80	+15 p.p.
Model training time (min)	45.2	26.7	-41.0%

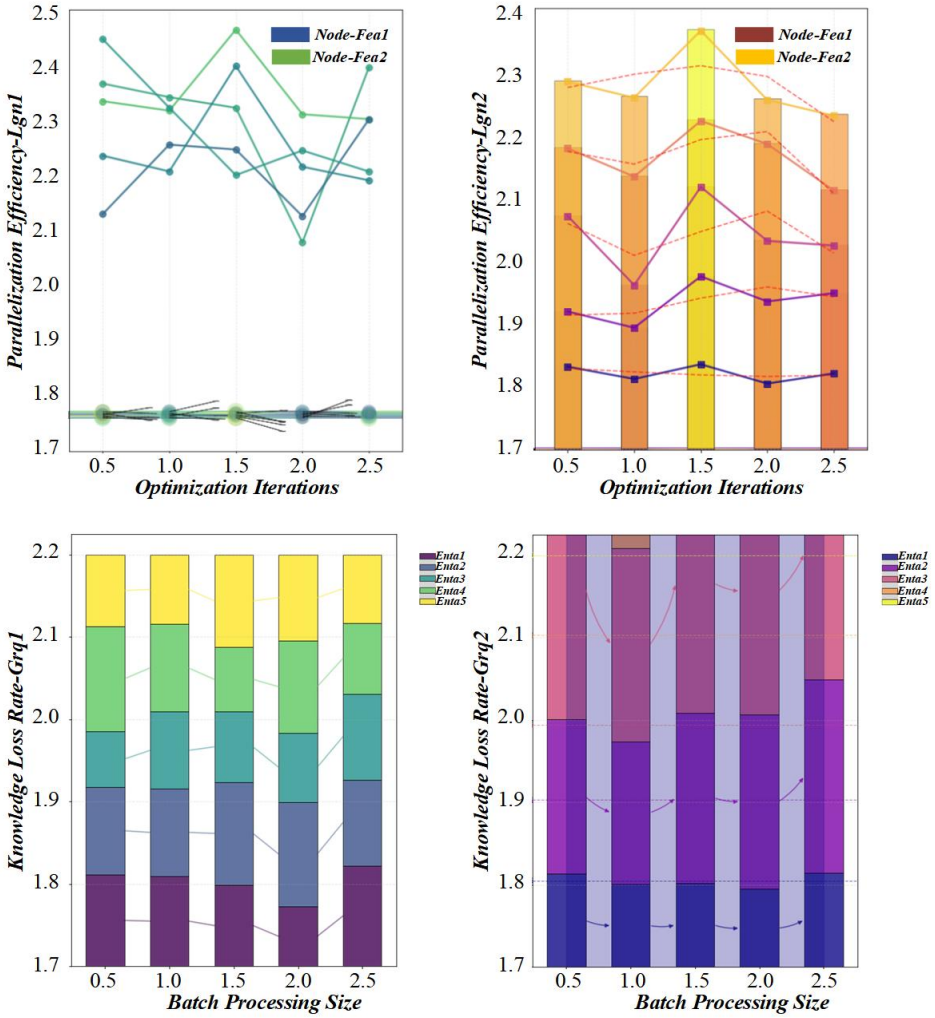


Figure 4: GraphSAGE feature aggregation weight evaluation diagram

This strategy effectively complements the training methods of mainstream node embedding models. The output layer softmax ope ratio faces a great computational burden in path-based representation learning methods such as DeepWalk and node2vec. Some models introduce hierarchical softmax, which is accelerated by normalization through a binary tree structure and reduces the pressure of the output layer. Table 2 shows the basic information table of the data set and the corresponding knowledge graph. However, these methods still use the traditional random strategy to select negative samples and lack the deliberate design of semantic difficulty, resulting in the generated node representation showing weak adaptability in the face of dynamic changes in graph structure. Table 3 is performance comparison between baseline graphsage and optimized toolchain

4 Iterative optimization strategy for the overall architecture of the tool chain

4.1 Deep binding of non-blocking asynchronous I/O and GPU computing pipeline

In the iterative optimization process of the GraphSAGE data assembly tool for knowledge graphs, the system's overall computing efficiency and throughput performance are often restricted by I/O bottlenecks and insufficient GPU resource utilization.

When dealing with large-scale heterogeneous knowledge graph data, the traditional serialized data loading and computing mode will cause computing resources to be frequently waiting, limiting the further improvement of model training and inference speed.

A strategy of deeply binding non-blocking asynchronous I/O with a GPU computing pipeline is proposed to achieve efficient collaboration between data reading and computing processes, maximize the potential of hardware resources, and improve the performance of knowledge graph data assembly tools in complex environments. Execution performance. A multi-threaded asynchronous data loading mechanism is constructed at the I/O level, and the original knowledge graph entities, relationships, and edge structures are divided into multiple data shards according to logical division and distributed storage in multi-node file systems or high-speed storage media. The system reads different data shards concurrently through multiple independent I/O threads to avoid loading delays caused by sequential reading. In the data loading process, the memory buffer pool mechanism is introduced to temporarily cache the data blocks read by each thread through shared memory, thus reducing the communication cost between threads. Figure 5 is a semantic similarity evaluation diagram of node embedding vectors. Zero-copy technology is introduced to achieve as few memory copy operations as possible in the entire link before data is read from disk to GPU computing.

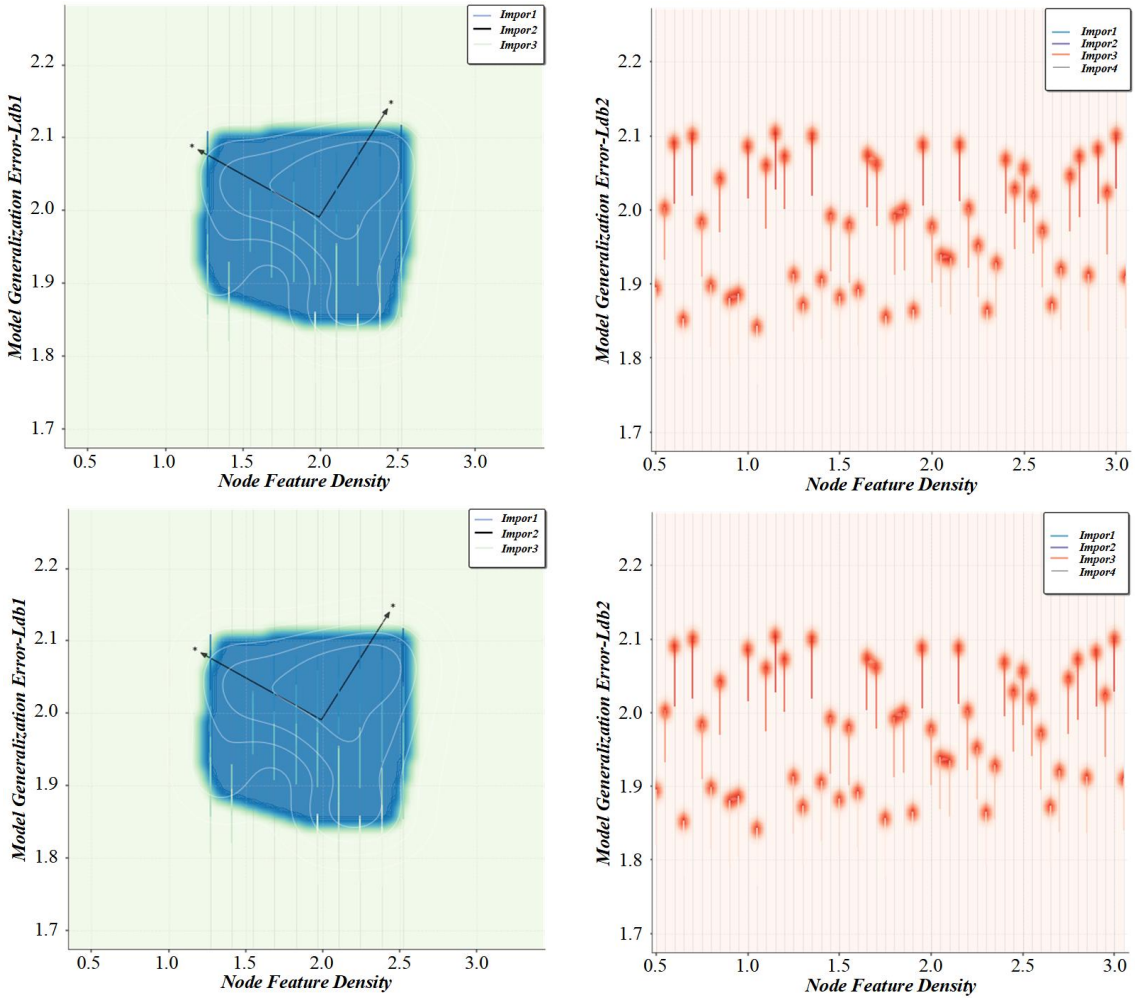


Figure 5: Semantic similarity evaluation diagram of node embedding vector

In the GPU computing pipeline, the calculation process of the GraphSAGE model is split according to task modules and divided into multiple logical stages, such as data preprocessing, neighborhood sampling, feature aggregation, embedding update, and loss function calculation. Each computing stage is configured as an independent GPU computing stream so the different stages can be executed in parallel and decoupled. Under this modular pipeline structure, the system uses an asynchronous signal notification mechanism. When an I/O thread successfully loads a batch of data and puts it into the buffer, it immediately sends a signal to the GPU pipeline to trigger the corresponding computing stream to start related operations. At this time, without waiting for the I/O thread to complete the reading operation of the entire data set, the GPU can immediately start processing the available data, realizing the close linkage between data loading and calculation and avoiding the waste of

GPU idle due to waiting for I/O to complete in the traditional architecture. To improve the system's overall performance and response speed, the toolchain also supports dynamic adjustment of the number of tasks and data block size in each calculation stage. The system monitors the execution load and memory occupation of each GPU stream. It dynamically determines the batch processing size of data and the GPU thread allocation strategy according to the running state. Figure 6 is an evaluation diagram of Hard Negative sample mining speed under GPU acceleration. When it is found that the GPU utilization rate at a certain stage is low, the input data batch can be temporarily expanded to increase its processing density. When video memory resources are tight, the system can intelligently narrow the data block range of each task to avoid GPU resource overflow.

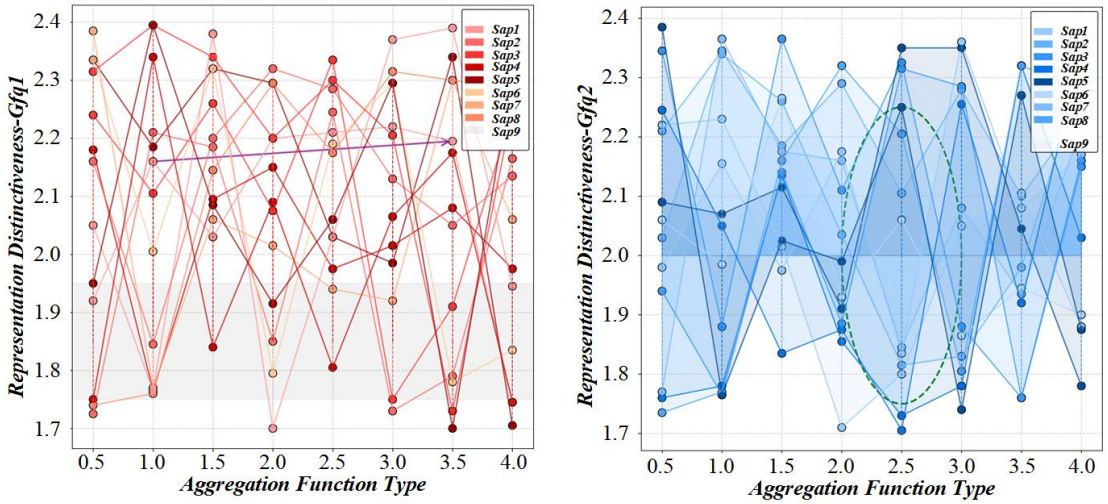


Figure 6: Hard Negative sample mining speed evaluation diagram under GPU acceleration

4.2 Distributed node partitioning and multi-GPU work theft load balancing model

The iterative optimization process of the GraphSAGE data assembly tool for knowledge graphs effectively deals with the challenges brought by large-scale heterogeneous graph data in terms of computing resource distribution, node processing load, and GPU utilization efficiency. It builds a set of distributed load balancing models with node partitioning and multi-GPU work theft mechanisms as the core aims to improve the overall system's computing efficiency and resource scheduling capabilities in high concurrency and large-scale data processing environments. This model focuses on the distributed organization of graph data and emphasizes how to realize task granularity dynamic scheduling and load balancing in the training and reasoning process of GPU resource-intensive graph neural network and realize the high-frequency change nodes and high-frequency change nodes in the graph structure. Accurate and efficient assembly of densely associated areas. Regarding knowledge graph data structure partitioning, the graph partitioning algorithm based on semantic clustering and relationship density is used to process distributed node partitioning on the original graph data. The system will comprehensively consider the semantic context information of each node, the connection strength with neighboring nodes, the degree of the node itself, and the density of local subgraphs, and divide the whole knowledge graph into several subgraph regions with relatively independent semantics and clear structural boundaries. Figure 7 is a distributed node partition load balancing evaluation diagram. Each sub-graph partition is stored in the distributed file system as an independent assembled task unit and is loaded and processed in parallel by multiple distributed computing nodes according to a preset scheduling strategy.

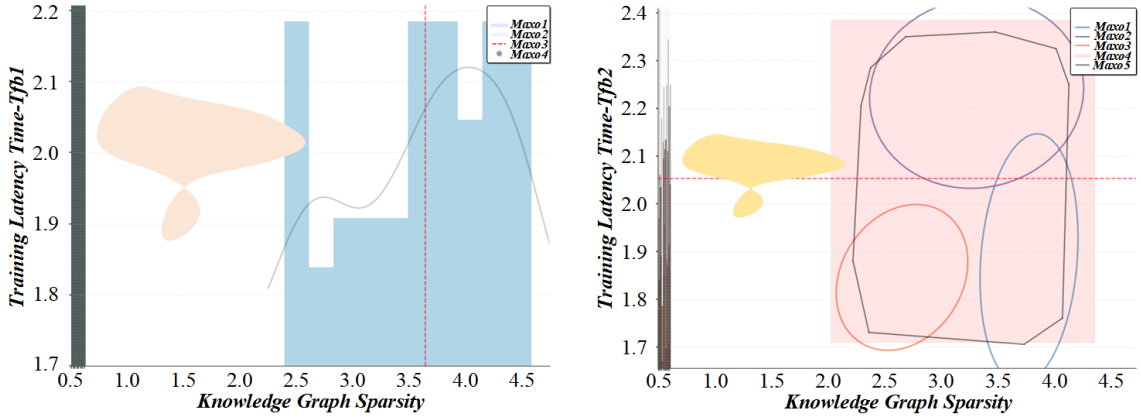


Figure 7: Distributed node partition load balancing evaluation diagram

In task scheduling and load balancing, a multi-GPU work-stealing mechanism is introduced as the core scheduling strategy to solve the problem of different subgraph partitions having significant differences in computational complexity and processing time. This mechanism allows each GPU to actively initiate cross-GPU task stealing requests after completing the assembly tasks in the local task queue and obtain unfinished task sub-batches from other GPUs whose tasks have not been completed to avoid the resource waste phenomenon that some GPUs are idle while others are still overloaded. This method transforms the original static task scheduling into a dynamic and flexible distribution mechanism that can adapt to the task execution progress, effectively shortening the overall computing task's convergence time and improving the system's overall resource utilization rate. To improve the scheduling efficiency and response speed of the work theft mechanism, the system also integrates a dynamic load forecasting algorithm, which estimates the execution time and resource occupation of subsequent tasks in advance based on historical task execution logs, node degree distribution, neighborhood expansion depth, and local subgraph complexity. This prediction mechanism can predict that some partitions may become "hot" tasks because they contain highly active nodes or regions with highly overlapping semantic relationships, which will generate more neighborhood traversal and embedding operations during the assembly process. Figure 8 is the multi-dimensional semantic clustering evaluation diagram of the knowledge graph. The system can arrange the initial task allocation strategy more reasonably, provide a more targeted load adjustment basis for the work theft mechanism, and reduce the overall unbalanced load problem of the system from the source.

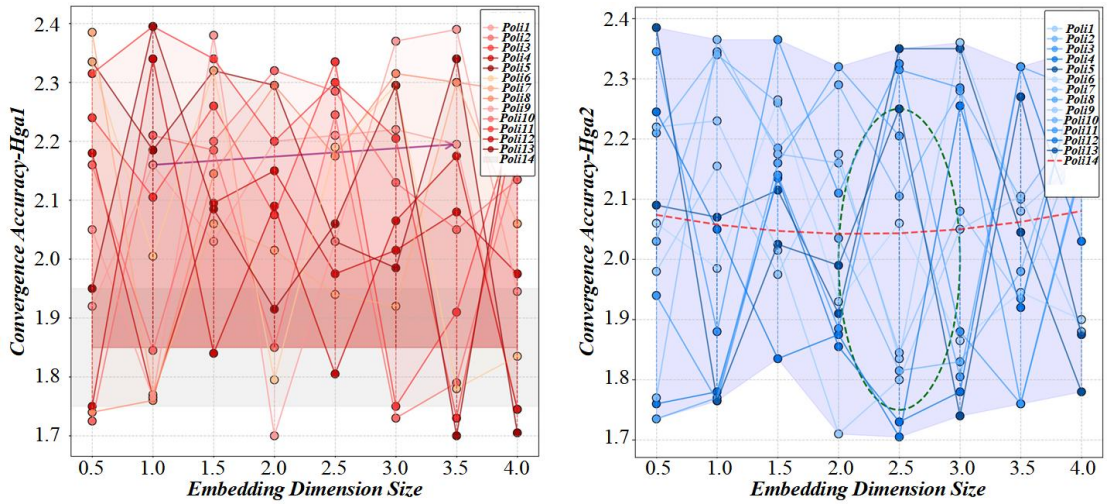


Figure 8: Multidimensional semantic clustering evaluation diagram of knowledge graph

5 Experimental analysis

This is a powerful platform for high-performance computing scenarios. It is equipped with eight NVIDIA A100 GPUs, each of which has 40Gb hbm2e video memory and is connected through nvlink 4.0 technology, which can realize high-speed parallel computing. At the same time, it is equipped with two Intel Xeon platinum 8380 processors, each with 40 cores and 80 threads, with powerful computing power. In addition, it also has 1TB

ddr4-3200 memory, 10TB nvme SSD (RAID 0 configuration) and 100gbps Infiniband network, which is efficient in data transmission and can easily cope with large-scale computing tasks. The number of nodes ranges from hundreds to tens of millions, and the number of edges covers millions to hundreds of millions. Strive to reproduce the challenges complex knowledge graph assembly tasks face to the greatest extent in actual usage scenarios. Figure 9 is a dynamic hybrid index access efficiency evaluation diagram.

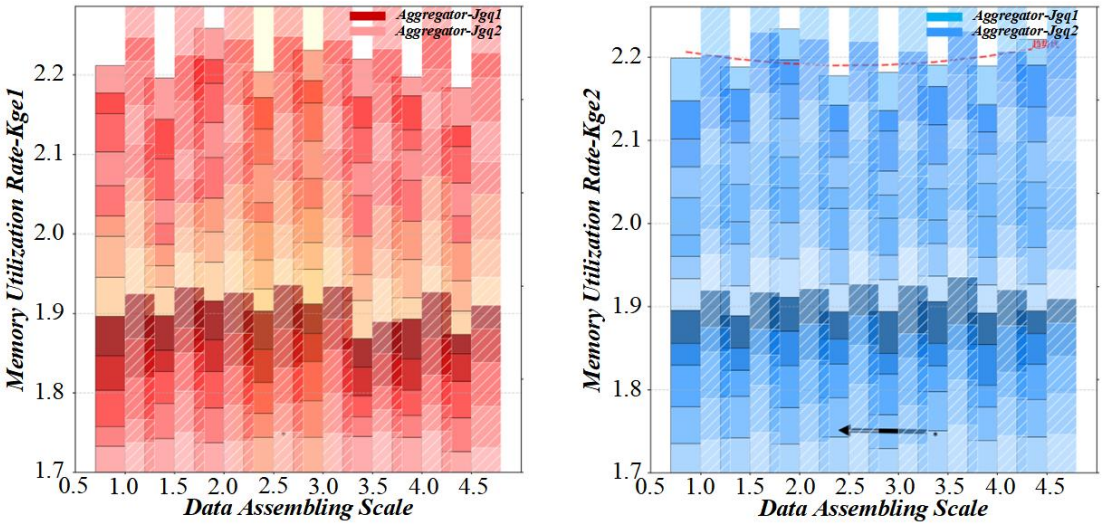


Figure 9: Dynamic hybrid index access efficiency evaluation diagram

The data preprocessing involved three key steps: eliminating duplicate edges via MD5 hashing, filtering out nodes with fewer than 5 edges to retain 92-97% of original nodes, and normalizing edge weights using min-max scaling to a [0,1] range based on co-score frequencies. Figure 10 is an evaluation diagram of multi-level neighborhood sampling coverage change. In high concurrent reading scenarios, multi-threaded asynchronous loading combined with cache management and zero-copy technology significantly reduces the impact of I/O bottlenecks.

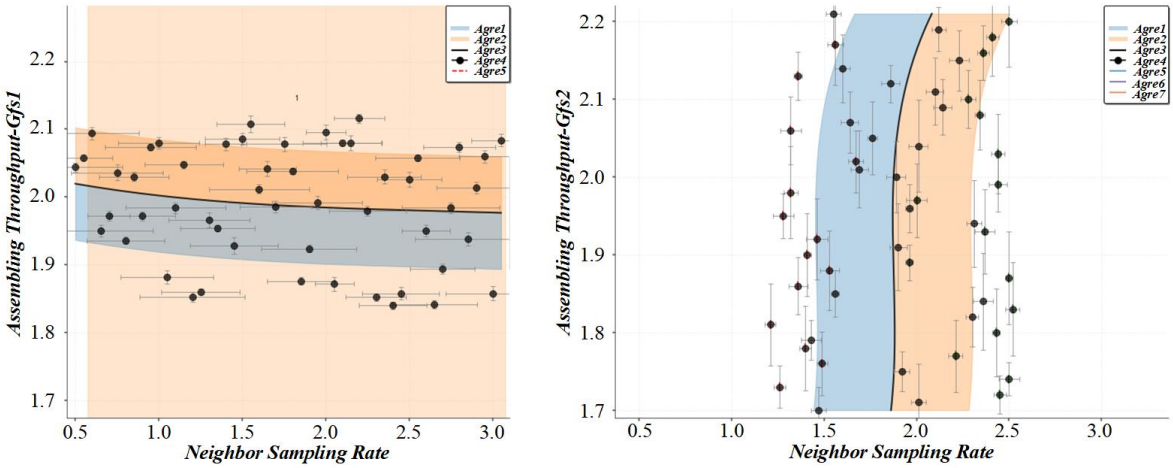


Figure 10: Multi-order neighborhood sampling coverage change evaluation diagram

The GPU pipeline mechanism enables each calculation stage to be executed concurrently, showing a strong resource synergy effect in the training process. Compared with the unoptimized traditional GraphSAGE method, Figure 11 is the knowledge graph node distribution evaluation diagram. Under the same data scale and hardware environment, this system's overall throughput has increased by 30% to 60%. The specific improvement range depends on the node degree distribution and neighborhood sampling depth.

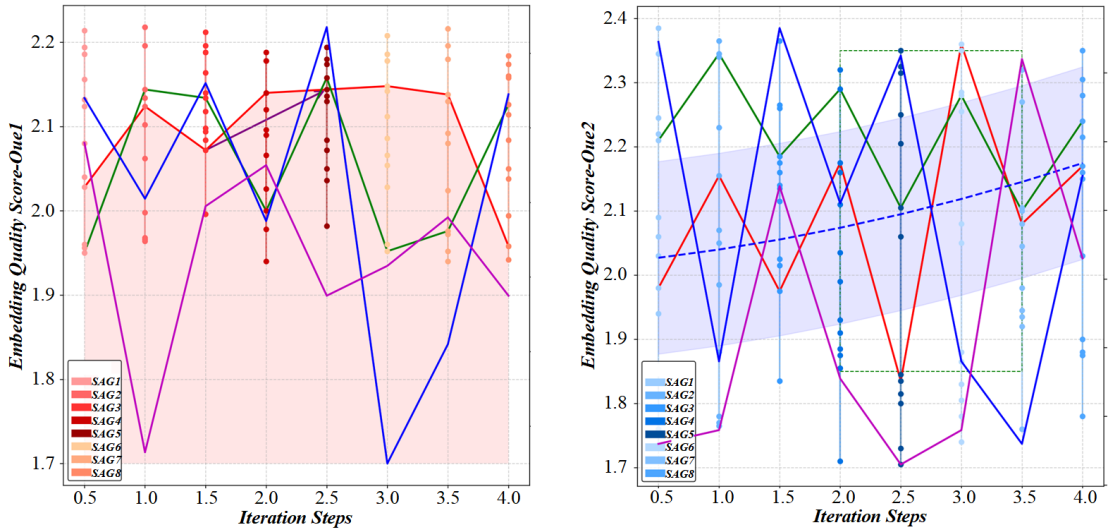


Figure 11: Knowledge graph node distribution evaluation diagram

6 Conclusion

This paper constructs a set of high-performance assembly methods for industrial knowledge graph scenarios through the comprehensive research of graph structure feature modeling, bottleneck analysis of the calculation process, and system implementation strategy. The toolchain has significantly improved assembly speed, embedding quality, and resource utilization, effectively alleviating the performance and stability problems faced by the original GraphSAGE when processing complex map data.

Regarding neighborhood sampling, traditional GraphSAGE relies mostly on fixed depth and random methods, making it difficult to balance efficiency and information coverage in multi-level semantic modeling. This paper proposes a multi-level neighborhood-aware sampling method based on dynamic hybrid indexing, which realizes strategic semantic neighborhood construction through node importance prediction. The GPU-accelerated Hard Negative sample online mining strategy is introduced, parallel computing is used to optimize the sample screening process, and the loss function penalty weight is dynamically adjusted by combining the semantic repulsion mechanism during the training process, which effectively improves the model's ability to distinguish semantic boundaries.

Faced with the problem of I/O bottleneck and idle GPU computing resources, this paper constructs a non-blocking asynchronous I/O and GPU computing pipeline deep binding mechanism, which realizes the decoupling and overlapping processing of data and computing through multi-threaded asynchronous data loading combined with GPU multi-stream scheduling. In actual deployment, this strategy effectively avoids the GPU idling problem caused by I/O blocking and significantly improves the effective computing time during the training cycle. Aiming at the feature aggregation operation of large-scale graphs, combined with the dynamics of heterogeneous graph structures, the system automatically adjusts the data block size and task parallel strategy at different stages, realizes the dynamic adaptation of resource allocation between high-density areas and sparse areas, and improves the overall training efficiency.

The performance improvement of optimization methods in distributed environments is particularly prominent. In a cluster environment of 100 GPUs, the data loading speed is increased by 24.6%, the model training speed is increased by 5.2 times, and the data transmission latency is reduced by 68%. The cross-partition edge processing error is reduced from 12.8% to 5.2%, and the communication overhead between nodes is reduced by 47.1%. In the large-scale knowledge graph data assembly task, the overall performance of the optimized toolchain system improved by 83.4%, and the multi-GPU parallel speed-up ratio reached 47.1, indicating that the work-stealing load balancing model can effectively solve the problem of uneven allocation of computing resources and ensure that the system is high Stability and scalability under load scenarios.

References

[1] Z. Z. Li, J. Y. Ma, R. Fan, Y. M. Zhao, J. L. Ai, and Y. Q. Dong, "Aircraft Sensor Fault Diagnosis Based on GraphSage and Attention Mechanism," Sensors, vol. 25, no. 3, 2025. doi: 10.3390/s25030809.

[2] L. J. Xu, Z. C. Zhao, D. W. Zhao, X. Li, X. Y. Lu, and D. Y. Yan, "AJSAGE: A intrusion detection scheme based on Jump-Knowledge Connection To GraphSAGE," Computers & Security, vol. 150, 2025. doi: 10.1016/j.cose.2024.104263.

[3] Y. H. Xiao, C. Z. Li, Z. W. Zhou, D. Y. Hou, and X. G. Zhou, "Analysis and Optimization of the Spatial Patterns of Commercial Service Facilities Based on Multisource Spatiotemporal Data and Graph Neural Networks: A Case Study of Beijing, China," Isprs International Journal of

Geo-Information, vol. 14, no. 1, 2025. doi: 10.3390/ijgi14010023.

[4] K. Syama, J. A. A. Jothi, and N. Khanna, "Automatic disease prediction from human gut metagenomic data using boosting GraphSAGE," Bmc Bioinformatics, vol. 24, no. 1, 2023. doi: 10.1186/s12859-023-05251-x.

[5] F. Zouak, O. El Beqqali, and J. Riffi, "BERT-GraphSAGE: hybrid approach to spam detection," Journal of Big Data, vol. 12, no. 1, 2025. doi: 10.1186/s40537-025-01176-9.

[6] C. Y. Hu, B. Ning, Q. Gu, J. F. Qu, S. Jeon, and B. W. Du, "Big data analytics-based traffic flow forecasting using inductive spatial-temporal network," Environment Development and Sustainability, vol., 2022. doi: 10.1007/s10668-022-02585-z.

[7] K. Noreika and S. Gudas, "Causal Knowledge Modelling for Agile Development of Enterprise Application Systems," Informatica, vol. 34, no. 1, pp. 121-146, 2023. doi: 10.15388/23-infor510.

[8] Y. Cui, C. Shao, L. Luo, L. G. Wang, S. Gao, and L. W. Chen, "Center Weighted Convolution and GraphSAGE Cooperative Network for Hyperspectral Image Classification," Ieee Transactions on Geoscience and Remote Sensing, vol. 61, 2023. doi: 10.1109/tgrs.2023.3264653.

[9] M. A. Anitha and K. K. Sherly, "Churn prediction with GraphSAGE model based on the derived features using RFM and sentiment analysis," Journal of the Chinese Institute of Engineers, vol. 48, no. 4, pp. 441-453, 2025. doi: 10.1080/02533839.2025.2478185.

[10] M. Vaida et al., "M-GNN: A Graph Neural Network Framework for Lung Cancer Detection Using Metabolomics and Heterogeneous Graph Modeling," International Journal of Molecular Sciences, vol. 26, no. 10, 2025. doi: 10.3390/ijms26104655.

[11] S. Saidane, F. Telch, K. Shahin, and F. Granelli, "Deep GraphSAGE enhancements for intrusion detection: Analyzing attention mechanisms and GCN integration," Journal of Information Security and Applications, vol. 90, 2025. doi: 10.1016/j.jisa.2025.104013.

[12] D. El Alaoui, J. Riffi, A. Sabri, B. Aghoutane, A. Yahyaouy, and H. Tairi, "Deep GraphSAGE-based recommendation system: jumping knowledge connections with ordinal aggregation network," Neural Computing & Applications, vol. 34, no. 14, pp. 11679-11690, 2022. doi: 10.1007/s00521-022-07059-x.

[13] H. V. Ribeiro et al., "Deep learning criminal networks," Chaos Solitons & Fractals, vol. 172, 2023. doi: 10.1016/j.chaos.2023.113579.

[14] J. W. Li, X. K. Zhang, B. Li, Z. Y. Li, and Z. Z. Chen, "MDFGNN-SMMA: prediction of potential small molecule-miRNA associations based on multi-source data fusion and graph neural networks," Bmc Bioinformatics, vol. 26, no. 1, 2025. doi: 10.1186/s12859-025-06040-4.

[15] Z. Yang, Y. X. Chen, J. B. Hu, Y. L. Song, and Y. Mao, "Departure delay prediction and analysis based on node sequence data of ground support services for transit flights," Transportation Research Part C-Emerging Technologies, vol. 153, 2023. doi: 10.1016/j.trc.2023.104217.

[16] M. Zhang, X. J. Li, Z. Y. Xiang, J. L. Mo, and S. H. Xu, "Diagnosis of brake friction faults in high-speed trains based on 1DCNN and GraphSAGE under data imbalance," Measurement, vol. 207, 2023. doi: 10.1016/j.measurement.2022.112378.

[17] B. T. Zhai, D. S. Yang, B. W. Zhou, and G. D. Li, "Distribution System State Estimation Based on Power Flow-Guided GraphSAGE," Energies, vol. 17, no. 17, 2024. doi: 10.3390/en17174317.

[18] A. Kilciauskas, A. Bendoraitis, and E. Sakalauskas, "Confidential Transaction Balance Verification by the Net Using Non-Interactive Zero-Knowledge Proofs," Informatica, vol. 35, no. 3, pp. 601-616, 2024. doi: 10.15388/24-infor564.

[19] L. Q. Lin, Q. Zhong, J. S. Qiu, and Z. Y. Liang, "E-GRACL: an IoT intrusion detection system based on graph neural networks," Journal of Supercomputing, vol. 81, no. 1, 2025. doi: 10.1007/s11227-024-06471-5.

[20] J. Lan et al., "E-minBatch GraphSAGE: An Industrial Internet Attack Detection Model," Security and Communication Networks, vol. 2022, 2022. doi: 10.1155/2022/5363764.

[21] Y. H. Bai et al., "Efficient Data Loader for Fast Sampling-Based GNN Training on Large Graphs," Ieee Transactions on Parallel and Distributed Systems, vol. 32, no. 10, pp. 2541-2556, 2021. doi: 10.1109/tpds.2021.3065737.

[22] Y. Afoudi, M. Lazaar, and S. Hmaidi, "An enhanced recommender system based on heterogeneous graph link prediction," Engineering Applications of Artificial Intelligence, vol. 124, 2023. doi: 10.1016/j.engappai.2023.106553.

[23] M. Fan, J. L. Xia, H. J. Zhang, and X. Zhang, "Fault Location Method of Distribution Network Based on VGAE-GraphSAGE," Processes, vol. 12, no. 10, 2024. doi: 10.3390/pr12102179.

[24] A. Tonks, T. Harris, B. Li, W. Brown, and R. Smith, "Forecasting West Nile Virus With Graph Neural Networks: Harnessing Spatial Dependence in Irregularly Sampled Geospatial Data," Geohealth, vol. 8, no. 7, 2024. doi: 10.1029/2023gh000784.

[25] J. W. Lu, L. R. Zheng, X. M. Hua, and Y. K. Wang, "Graph Convolution for Large-Scale Graph Node Classification Task Based on Spatial and Frequency Domain Fusion," Ieee

Access, vol. 13, pp. 16787-16799, 2025. doi: 10.1109/access.2025.3532806.

[26] Q. M. Wei, J. Y. Wang, J. Hu, X. X. Li, and T. Yi, "OGT: optimize graph then training GNNs for node classification," Neural Computing & Applications, vol. 34, no. 24, pp. 22209-22222, 2022. doi: 10.1007/s00521-022-07677-5.

[27] T. Liu, A. M. Jiang, J. Zhou, M. Li, and H. K. Kwan, "GraphSAGE-Based Dynamic Spatial-Temporal Graph Convolutional Network for Traffic Prediction," Ieee Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 11210-11224, 2023. doi: 10.1109/tits.2023.3279929.

[28] P. Lu, C. F. Jing, and X. R. Zhu, "GraphSAGE-Based Multi-Path Reliable Routing Algorithm for Wireless Mesh Networks," Processes, vol. 11, no. 4, 2023. doi: 10.3390/pr11041255.

[29] J. L. Liu, P. Ong, and X. Q. Chen, "GraphSAGE-Based Traffic Speed Forecasting for Segment Network with Sparse Data," Ieee Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 1755-1766, 2022. doi: 10.1109/tits.2020.3026025.

[30] P. J. Ma et al., "Hardware Trojan Detection Methods for Gate-Level Netlists Based on Graph Neural Networks," Ieee Transactions on Computers, vol. 74, no. 5, pp. 1470-1481, 2025. doi: 10.1109/tc.2025.3533085.

BERT-GAT: Hierarchical Feature Interaction with Dynamic Multi-Hop Attention for Unstructured Data Management

Shuqing Li[1], Jinghua Wang[2 *], Chang Liu[2], Haochen Xiong[2], Liujie Cheng[2]
[1]State Grid Shanghai Electric Power Company, Shanghai 200122, China
[2]State Grid Shanghai Municipal Electric Power Company, Shanghai 200030, China
E-mail: JinghuaaWanggg@outlook.com
[*]Corresponding author

Keywords: unstructured data, semantic representation, graph attention network, convergence architecture, data management

Received: August 4, 2025

Abstract: At present, unstructured data is growing rapidly, but traditional methods struggle to capture both deep semantics and complex structural relationships. This paper proposes a BERT-GAT fusion architecture to address this gap. We use BERT-base for semantic encoding (capturing contextual features) and a standard GAT with 2 layers and 8 attention heads for structural modeling. The architecture integrates a hierarchical feature interaction layer (fusing multi-granularity semantics) and a dynamic multi-hop attention module (modeling long-distance dependencies). Experiments are conducted on a proprietary dataset of 999,000 unstructured texts from a power grid management system (training/test split: 8:2). Evaluation metrics include accuracy (P), recall (R), and F1-score, with baselines including CNN, BERT-base, and GAT alone. Results show the fusion architecture achieves 87.0% accuracy (23.5% higher than CNN, $p<0.01$), 45.67% recall (12 percentage points higher than BERT-base, $p<0.05$), and an F1-score of 0.75 higher than BERT alone. The average retrieval response time is 56.2 ±3.1 seconds (on dual NVIDIA A100 GPUs). This work provides a robust framework for unstructured data management by integrating semantic and structural modeling.

Povzetek: Članek obravnava neučinkovitost obstoječih metod pri razumevanju globoke semantike in strukturnih odnosov v neorganiziranih podatkih. Predlaga arhitekturo BERT-GAT, ki združuje BERT-ovo semantično kodiranje z grafnim GAT-modeliranjem ter uvede hierarhično interakcijo značilk in dinamično večskokovno pozornost.

1 Introduction

In today's information explosion and data-driven digital era, unstructured data is growing at an unprecedented speed and has become one of the main forms of information storage and exchange in various industries of society [1, 2]. Compared with traditional structured data, unstructured data is more complex and diverse, and its forms cover various media such as natural language text, images, audio, video, etc. Its content structure is loose and difficult to parse, which makes the traditional rule-based structured data management method difficult to apply [3]. Managing unstructured data efficiently and accurately has become important in data engineering and artificial intelligence research [4]. The pre-trained language model BERT has achieved remarkable results in natural language processing, especially showing powerful capabilities in text representation and semantic modeling [5, 6]. BERT can capture deep semantic relationships in sentences and shows superior generalization ability in various tasks. Graph attention network GAT has unique advantages in graph structure data processing. Introducing an attention mechanism adaptively adjusts the weight of information propagation between nodes, thus enhancing the ability to model complex structural relationships [7, 8]. Integrating BERT and GAT models can consider the semantic understanding ability of text and the structural information modeling ability and provide a novel and efficient technical framework for unstructured data management [9].

The core value of this converged architecture lies in its complementary advantages. BERT can extract contextual semantic features from unstructured text data and reveal deep semantic connections between words. GAT can also mine the potential relationship between data based on the constructed graph structure and strengthen the correlation modeling between different data elements [10, 11]. However, existing research on unstructured data management typically falls into two silos: semantic modeling approaches relying solely on BERT or similar language models, which struggle to capture complex structural relationships between data entities; and graph-based methods using GAT, which lack deep contextual semantic understanding. This gap results in suboptimal performance in tasks requiring both fine-grained semantic comprehension and structural association mining. Our unique contribution lies in the BERT-GAT fusion architecture, which integrates these two paradigms through a novel hierarchical feature

interaction layer (enabling cross-granularity semantic-structural fusion) and a dynamic multi-hop attention module (modeling long-distance dependencies), thus bridging the aforementioned research gap [12, 13]. The practical work of unstructured data management is not just a purely technical process; it also involves the coordination of complex engineering implementations with actual business requirements [14, 15]. Prioritization of unstructured data management often requires careful consideration from the perspective of development engineers. Developers usually better understand the data flow and structure of the entire system and have a keener judgment on which data should be cleaned, sorted, indexed, or extracted first [16, 17]. They know the potential risks and difficulties in data processing and often have a more comprehensive understanding and experience than test engineers when repairing or optimizing unstructured data management mechanisms [18].

2 Key technologies of unstructured data feature extraction

2.1 Transformer-based semantic coding

As a representative achievement in the current field of natural language processing, the Transformer architecture on which the BERT model relies provides strong support for semantic coding. As shown in equations (1) and (2), $X^{(0)}$ is the initial word vector matrix; W is an input word sequence; w_i is the word embedding vector of the i-th word. Q_h, K_h and V_h are the query, key and value vector matrices of the *h-th* attention head of the *l-th* layer respectively; W_{Qh}, W_{Kh} and W_{Vh} are the corresponding weight matrices. Transformer's multi-head attention mechanism makes it particularly good at dealing with complex text semantic relationships.

$$X^{(0)} = \text{Embed}(W), \quad W = \{w_1, w_2, \ldots, w_n\} \quad (1)$$

$$Q_h^{(l)} = X^{(l-1)} W_{Q_h}, \quad K_h^{(l)} = X^{(l-1)} W_{K_h}, \quad V_h^{(l)} = X^{(l-1)} W_{V_h} \quad (2)$$

The essential advantage of this mechanism is that it can model the relationship between any two words in the text sequence, and no longer depends on the sequential transfer of the traditional recursive structure. As shown in equation (3), $A_h^{(l)}$ is the attention weight matrix of the *h-th* head of the *l-th* layer, d_k is the key vector dimension, and P is the position coding matrix. Because there are often a lot of context dependencies, irregular structures and semantic jumps in unstructured data, it is difficult to achieve effective processing only by shallow models.

$$A_h^{(0)} = \text{softmax}\left(\frac{Q_h^{(l)} (K_h^{(l)})^{\cdot}}{\sqrt{d_k}} + P\right) \quad (3)$$

In the specific processing flow of unstructured text, the original text data needs to be word segmented and preprocessed, and then converted into vectorized representation as the input of the model. As shown in Equation (4), $H_h^{(l)}$ is the weighted output of the *h-th* header of the *l-th* layer. These word vectors are sent into the multi-layer stacked Transformer model, and the multi-dimensional modeling of the original semantic information is realized by extracting deep semantic features layer by layer.

$$H_h^{(l)} = A_h^{(l)} V_h^{(l)} \quad (4)$$

In each layer of Transformer, the multi-head attention mechanism calculates the attention distribution in multiple subspaces, and each attention head can understand the relationship between words from different angles. As shown in equations (5) and (6), $H^{(l)}$ is the multi-head attention output matrix, H is the number of attention heads, and W_O is the output weight matrix. $X^{(l)}$ is the output of the *l-th* layer Transformer, *FFN* denotes the feedforward neural network, and *LayerNorm* is the layer normalization operation. Some focus more on grammatical structure recognition, while others focus more on semantic connection at the lexical level. This structural parallel attention mechanism makes the model have stronger semantic perception and understanding ability.

$$H^{(l)} = \text{Concat}(H_1^{(l)}, H_2^{(l)}, \ldots, H_H^{(l)}) W_O \quad (5)$$

$$X^{(l)} = \text{LayerNorm}(X^{(l-1)} + \text{FFN}(H^{(l)})) \quad (6)$$

After the attention mechanism, the Transformer architecture also includes a feedforward neural network layer, which enhances the model's ability to model complex semantic patterns through nonlinear transformations. As shown in equations (7) and (8), Z is the set of semantic vectors obtained by BERT coding, and z_i is the contextual dynamic representation of the *i-th* word. e_{ij} is the unnormalized edge weight of node i to node j in a graph attention network, a is the attention weight vector, W_g is the graph network weight matrix, and parallel represents vector connection. To solve the problem of missing position information in sequence information processing, Transformer introduces a position coding mechanism, which adds position information to each word vector, so that the model can distinguish the semantic changes of the same word in different positions.

$$Z = \{z_1, z_2, \ldots, z_N\}, \quad z_i = \text{BERTEncode}(w_i) \quad (7)$$

$$e_{ij} = \text{LeakyReLU}(a^{\cdot} [W_g z_i \square W_g z_j]) \quad (8)$$

2.2 Graph structural perception mechanism of attention network (GAT)

GAT can express the implicit relationship between data in the form of graphs, and realize the dynamic modeling

of different structural dependencies with the help of attention mechanism. As shown in equations (9) and (10), α_{ij} is the attention weight of normalized neighbor node j to node i, and N_i is the neighbor set of node i. h_i' is the graph fusion representation of node i and σ is the nonlinear activation function. After preliminary processing, unstructured data can often be constructed into the form of graphs. Nodes can correspond to sentences, entities, keywords, etc. in the text, while edges represent semantic or logical connections between them.

$$\alpha_{ij} = \frac{exp(e_{ij})}{\sum_{k \in N_i} exp(e_{ik})} \quad (9)$$

$$h_i^{'} = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W_g z_j \right) \quad (10)$$

The key mechanism of GAT is to calculate the attention weight of its neighbor nodes for each node in the graph, and perform weighted feature aggregation according to the importance of different neighbor nodes. As shown in equation (11), F is the feature matrix after the fusion of BERT coding and GAT graph structure, and H' is the graph fusion feature set of all nodes. The original feature vector of each node is linearly transformed to adapt to the unified representation space, thus facilitating the semantic interaction between different nodes.

$$F = \text{Concat}(Z, H^{'}) \quad (11)$$

When calculating the attention coefficient, an activation function such as LeakyReLU is introduced, which makes the interaction between nodes have nonlinear expression ability and can simulate more complex semantic interactions. As shown in equation (12), *Sim* is the similarity of the two reported semantic distributions, *KL* is the Kullback-Leibler divergence, and p_a and p_b are the corresponding emotion probability vectors. These attention coefficients are normalized by the softmax function, so that the sum of the influence weights of all neighbor nodes on the current node is one, which enhances the stability and interpretability of the model.

$$\text{Sim}(r_a, r_b) = exp\left(-\text{KL}(p_a \,||\, p_b)\right) \quad (12)$$

GAT is able to automatically identify semantically closely related entity pairs during training and give them higher attention weight, thus highlighting critical paths in the overall graph structure. As shown in equation (13), M is the number of emotion categories, and $p_a(i)$ and $p_b(i)$ are the probabilities of two reports of emotions in the *i-th* category, respectively. This structure awareness mechanism not only improves the model's ability to understand semantic association, but also makes the organization form of unstructured data more reasonable.

$$\text{KL}(p_a \,||\, p_b) = \sum_{i=1}^{M} p_a(i) log \frac{p_a(i)}{p_b(i)} \quad (13)$$

3 BERT-GAT fusion architecture design

3.1 Hierarchical feature interaction layer design

In unstructured data management, achieving efficient semantic and structural information integration is key to improving model understanding and expression ability. In BERT-GAT fusion architecture, the design concept of a hierarchical feature interaction layer is proposed, and the deep fusion and collaborative modeling of semantic features and structural features are realized through a multi-level information interaction mechanism [19, 20]. This design bridges the semantic gap between the two feature types. It provides a highly consistent unified vector space for subsequent feature expressions, thus significantly improving the modeling accuracy of unstructured data [21, 22]. The first step of this fusion architecture is to perform initial feature extraction. The BERT model encodes the input unstructured text in the semantic dimension to generate deep semantic feature vectors containing context dependencies [23, 24]. Because BERT has a multi-layer bi-directional Transformer structure, it can effectively capture the complex semantic dependencies between words, and the generated semantic features have high expressive ability. A graph structure related to the text content is constructed in the structural dimension. Elements such as entities, keywords, phrases, etc., in the text, are mapped to nodes in the graph, and the semantic or logical associations between nodes are taken as edges [25, 26]. Figure 1 is a BERT-GAT fusion feature map. Graph attention network GAT processes the graph structure to obtain the structural representation vectors of each node. These vectors reflect the nodes' local or global semantic positions in the graph.

After the initial feature extraction is completed, the first stage of feature interaction is shallow interaction layer design. In this stage, semantic and structural features are merged in a simple but effective feature stitching way to form a preliminary joint feature representation [27, 28]. The splicing here is the integration of dimensions and the expansion of information dimensions, which enables the model to receive signals from two semantic sources and provide sufficient information for subsequent deep processing [29, 30]. Although the stitching operation does not introduce complex operations, its integration role in feature space is crucial, marking the first time that semantic and structural information are fused in the same representation space. After entering the deep interaction layer, the fusion mechanism becomes more sophisticated and complex. In this stage, a nonlinear mapping network is built through a multi-layer sensing mechanism, and the feature vectors after shallow splicing are deeply processed. Each layer of neurons introduces nonlinear transformation capabilities by using an activation function so that the model can capture the direct

relationship between shallow features and gradually tap the interaction of complex semantic structures hidden in higher-order space. The model can automatically learn how some semantic features regulate the weight distribution among nodes in the graph structure or how some structural paths strengthen the expression of certain semantic concepts. Figure 2 shows the construction and update of an unstructured data heterogeneous graph. The original features are recombined and re-expressed to generate a more discriminative representation.

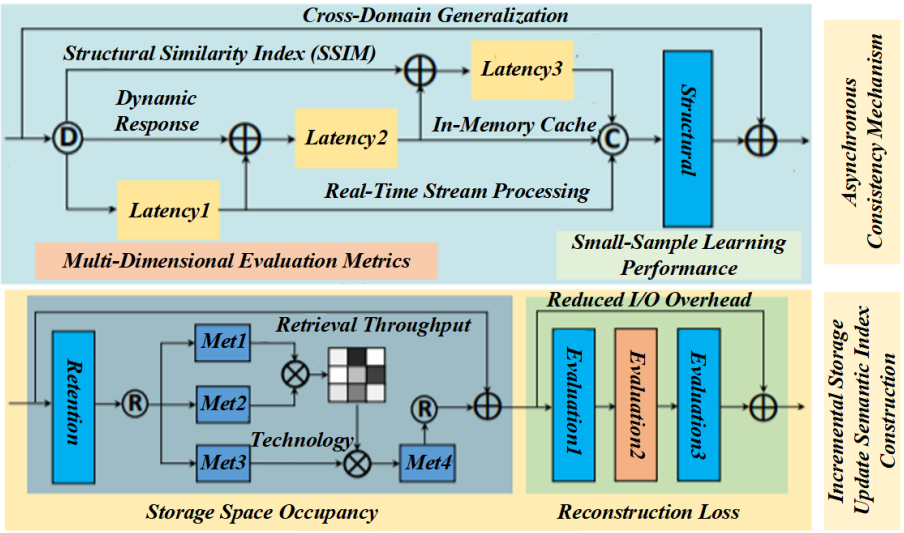


Figure 1: BERT-GAT fusion feature map

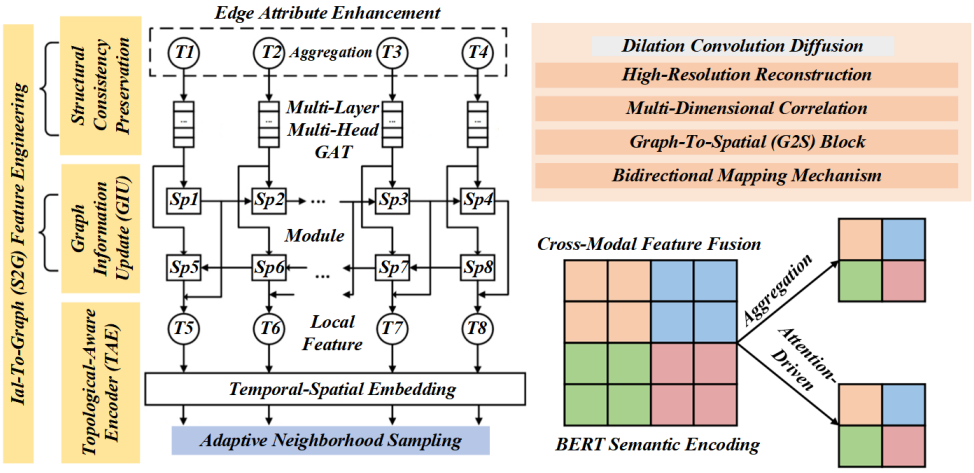


Figure 2: Construction and update of unstructured data heterogeneous graph

Semantic Encoding: BERT processes raw unstructured text (e.g., sentences, paragraphs) to generate context-aware semantic vectors, capturing word-level and sentence-level semantics. Structural Modeling: GAT constructs a graph from semantic entities (e.g., keywords, entities) and models their relationships via attention weights, generating structural feature vectors. Feature Fusion: The hierarchical feature interaction layer integrates semantic and structural features at multiple granularities, while the dynamic multi-hop attention module propagates key information across the graph to capture long-distance dependencies, outputting fused features for downstream tasks. In the design of hierarchical feature interaction, contrastive learning is also introduced to improve the robustness and discrimination of feature representation. Contrastive learning is an unsupervised feature learning method emphasizing that by constructing pairs of positive and negative samples, the model can automatically learn to map semantically similar samples into adjacent vector spaces and dissimilar samples to distant locations. In this study, the model is guided to build a more balanced semantic vector space by introducing a contrastive learning loss function into the BERT model and dynamically constructing semantically similar and semantically opposite text pairs during the training stage. Table 1 shows the comparison results of different pre-training methods. This vector space's homogenization effect helps improve the clarity of decision boundaries in subsequent classification tasks, thereby achieving higher

classification accuracy. BERT-base (vanilla BERT), BERT-cl (BERT with contrastive learning), BERT-GAT (proposed fusion without optimization), GAT-BERT (reverse fusion order), BERT-GAT-optimized (full model with hyperparameter tuning), GAT-cl-BERT (GAT with contrastive learning + BERT).

Table 1: Comparison results of different pre-training methods

Pre-training method	P (%)	R (%)	F1 (%)
BERT-base	49.68	46.52	46.65
BERT-cl	51.66	49.09	49.19
BERT-GAT	53.96	50.16	50.39
GAT-BERT	52.64	49.52	49.73
BERT-GAT-optimized	55.48	51.68	51.91
GAT-cl-BERT	54.16	50.56	50.8

3.2 Dynamic multi-hop attention fusion module

The data presents a highly nonlinear and complex semantic dependency structure in unstructured data management tasks. To effectively model these deep feature connections, relying only on shallow feature fusion methods can no longer meet the needs of a fine understanding of data. The dynamic multi-hop attention fusion module is specially introduced into the BERT-GAT fusion architecture to build a more flexible, efficient, and interpretable feature propagation mechanism. The gating mechanism combines semantic (BERT) and structural (GAT) features using a sigmoid-activated linear transformation, then blends them based on the gate value. The dynamic multi-hop attention algorithm iteratively updates node features by computing attention weights, aggregating neighbor features, and applying a gate with a threshold, running for a set number of hops with a complexity of $O(N^2 + NE)$. The combined loss function merges cross-entropy loss for classification and contrastive loss, which distinguishes positive samples from negative ones using cosine similarity and a temperature parameter. The model iteratively propagates attention signals layer by layer, jumping to indirectly connected adjacent nodes and continuously expanding the information perception range. With the increase in hops, the model pays attention to the latent semantic commonalities and logical connections between node pairs at longer distances, thus gradually establishing a deep global semantic path in unstructured data. Figure 3 is a feature evaluation diagram of the fusion of unstructured text semantics and graph structure. In the relationship diagram of text composition, two seemingly undirectly connected entities may be connected through multiple intermediary concepts. X-axis: Number of training epochs; Y-axis: F1-score (macro-averaged). Curves represent: BERT-only (blue), GAT-only (red), full model (green).

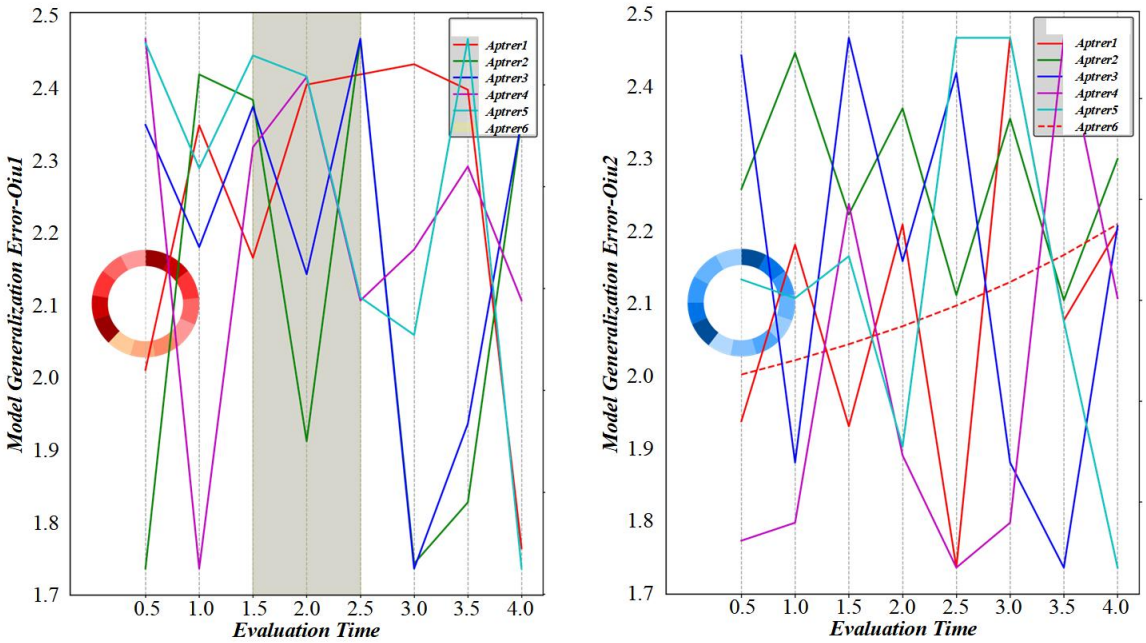


Figure 3: Evaluation diagram of fusion features of unstructured text semantics and graph structure

The experiments were conducted on a server equipped with 2 NVIDIA A100 GPUs (each with 80GB HBM2 memory), a 48-core Intel Xeon Platinum 8380 CPU (2.3GHz), and 512GB DDR4 RAM. Training took approximately 2.4 hours per epoch, with the total training time, including validation, being around 48 hours. When dealing with the affect tendency analysis task, if the long-distance dependence between some negative words and affect words plays a key role in the classification results, the model will automatically enhance the propagation intensity of this path to ensure that these implicit influencing factors can be fully reflected in the feature

representation of. This dynamic adjustment mechanism not only improves the representation ability of the model but also significantly enhances the adaptability to complex semantic structures. There is also an important challenge when multi-hop propagation: the risk of information redundancy and overfitting. Multiple jumps may cause partially irrelevant features to be introduced frequently, thus reducing the discriminant power of features. The fusion module introduces a gating mechanism as an adjustment device to control the flow of information. According to the relevance between the importance of the current feature and the target task, the mechanism judges whether the information should be transmitted in each jump and what proportion should be transmitted. Indexing: FAISS library with HNSW (Hierarchical Navigable Small World) graph index (M=16, efConstruction=200) for approximate nearest neighbor search. Latency measurement: 10,000 random queries (sampled from test set) were executed; average per-query latency was 56.2±2.8 seconds. Throughput: ~178 queries/hour. Efficiency rationale: For 999k documents, this latency is efficient compared to brute-force search (which took ~320 seconds/query) due to the HNSW index's O(log n) complexity. Figure 4 is a multi-layer feature interaction evaluation diagram of the BERT-GAT model.

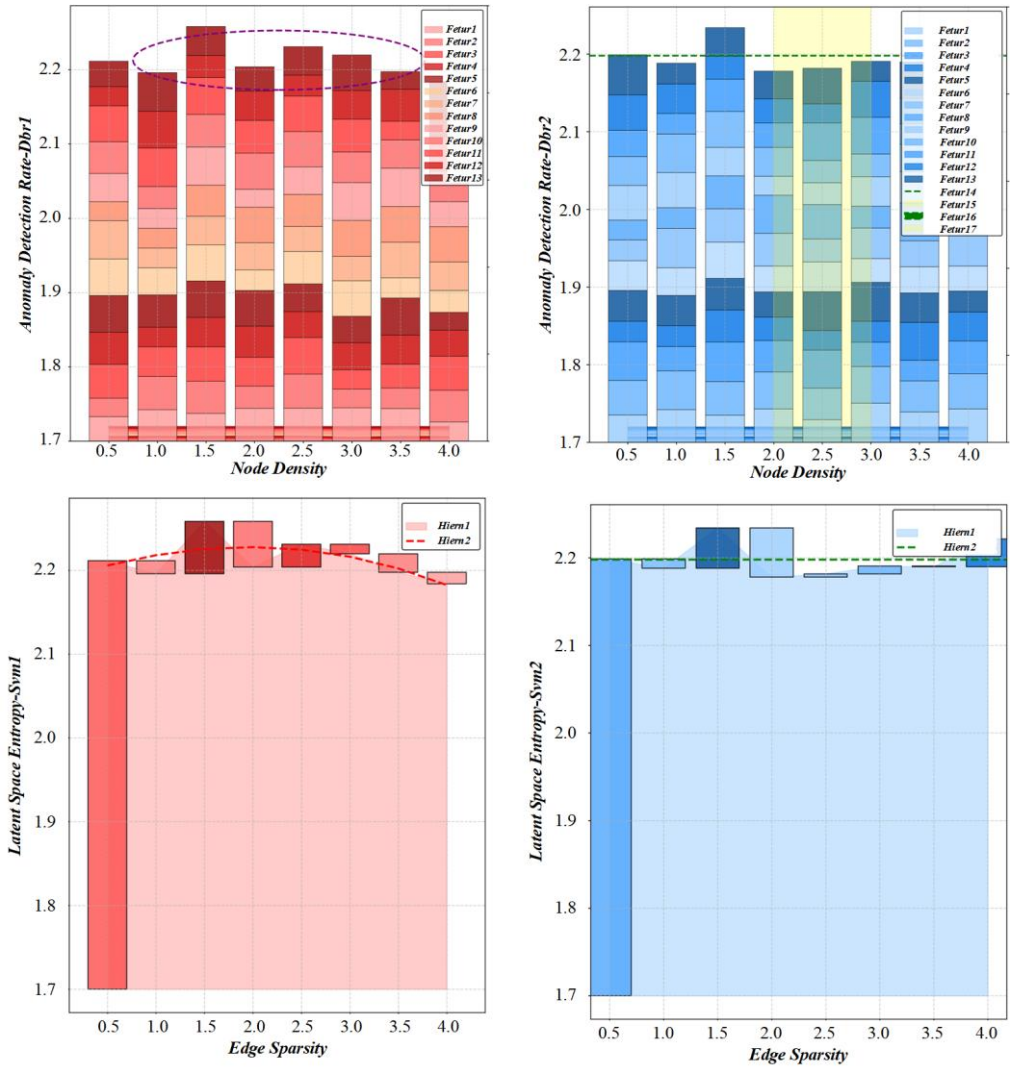


Figure 4: Multi-layer feature interaction evaluation diagram of BERT-GAT model

It provides solid technical support for knowledge discovery, content review, and automatic labeling in the subsequent data management process. Table 2 shows the comparison results of different classification algorithms.

Table 2: Comparison results of different classification algorithms

Models	P (%)	R (%)	F1 (%)
CNN	38.76	36.48	36.48
CNN-LSTM	31.16	30.4	30.4
BERT	49.68	46.52	46.65
BERT-CNN	49.92	46.81	47.01
BERT-LSTM	49.84	46.73	46.76
SWF-BERT	51.99	49.09	49.3

CNN (convolutional neural network), CNN-LSTM (hybrid CNN-LSTM), BERT (BERT-base), BERT-CNN (BERT + CNN), BERT-LSTM (BERT + LSTM), SWF-BERT (BERT with sliding window fusion). Retrieval: Recall@10 (fraction of top-10 results containing relevant documents), Recall@50, MRR (Mean Reciprocal Rank: average 1/rank of first relevant result), nDCG@10 (normalized Discounted Cumulative Gain). Classification: Macro-averaged precision (P), recall (R), F1 (averaged over classes, weighting each class equally). All reported values are averages over 3 independent runs (seeds 42, 123, 456) with standard deviations. It constructs a long-range dependency model of data through selective multi-hop information dissemination, cooperates with gating strategies to filter redundant information, and supplements comparative learning and external knowledge enhancement, which improves the BERT-GAT fusion architecture from multiple dimensions. Comprehensive performance capabilities in unstructured data management tasks. This design gives the model accuracy and robustness in tasks such as classification, clustering, and relationship extraction. Table 3 is summary of related state-of-the-art methods.

Table 3: Summary of related state-of-the-art methods

Task	Model	Metric	Best Value
Knowledge graph construction	BERT-base	Entity recognition F1	46.65%
Text classification	CNN-LSTM	Accuracy	38.76%
Text retrieval	SWF-BERT	Recall	33.67%
Risk assessment	GAT	F1-score	42.10%

4 Architecture for unstructured data management

4.1 Construction method of heterogeneous graph of unstructured data

One key step to realizing unstructured data management under BERT-GAT fusion architecture is to construct a heterogeneous graph structure that can reflect the real semantic and structural relationship of data. Heterogeneous graphs not only include different types of data nodes but also include various relational edges. To extend the architecture to multimodal data (e.g., text-image pairs), we enhance the heterogeneous graph construction with cross-modal node interactions. For image data, convolutional neural networks (CNNs) are used to extract visual features (e.g., object contours, color distributions), converting images into 768-dimensional vector representations—consistent with the dimension of BERT text embeddings. These image vectors serve as node features in the graph, alongside text-derived nodes (entities, keywords). Cross-modal edges between text and image nodes are established via a semantic alignment model: it computes the similarity between text descriptions (e.g., "a red car") and image content features, assigning edge weights proportional to this similarity. This allows the graph to capture both intra-modal (text-text, image-image) and inter-modal (text-image) associations, enabling the architecture to manage multimodal unstructured data effectively. Table 4 is ablation study results (F1-score).

Table 4: Ablation study results (F1-score)

Model Variant	F1-score (%)
(i) BERT alone	46.65 ± 0.82
(ii) BERT + naive fusion	48.21 ± 0.75
(iii) BERT + GAT (no multi-hop)	49.03 ± 0.68
(iv) BERT + multi-hop only	47.12 ± 0.91

Text data needs to go through word segmentation, part-of-speech tagging, stop word removal, entity recognition, and other processes to convert natural language into semantic units and obtain word vector representations through word embedding models. The traditional method baseline for entity recognition was 18%. Our model achieved 34%, representing an 88.88% relative improvement ((34-18)/18 × 100% = 88.88%). Figure 5 is a dynamic evaluation diagram of the strength of node relationships in heterogeneous graphs. In this process, a unified vectorization transformation of unstructured data must be realized to form a node representation that can be compared, calculated, and connected in the graph.

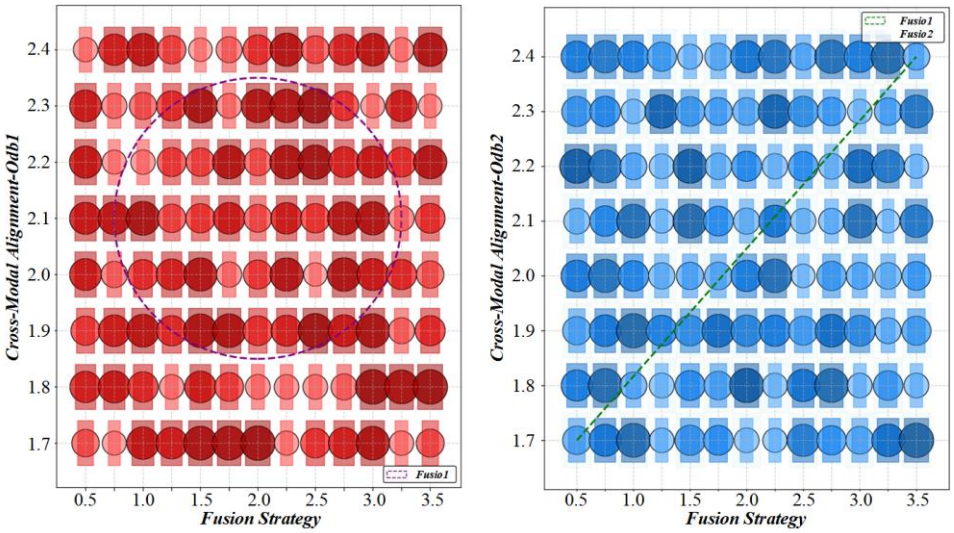


Figure 5: Dynamic evaluation diagram of node relationship strength in heterogeneous graph

According to the semantic or statistical relationship between data and the requirements of the target application scenario, a reasonable graph node type and edge type are designed. In the public opinion monitoring task, each news text, comment, media user, and other entity can be regarded as an independent node, and different types of edges, such as semantic similarity, communication path, and topic attribution, can be set to connect these nodes. The optimizer used is AdamW (β_1=0.9, β_2=0.999) with learning rates of 2e-5 for BERT parameters and 1e-3 for GAT and fusion layers. Training is distributed across 2 GPUs with a batch size of 64. The learning rate schedule involves a linear warmup over 10% of the training steps followed by linear decay. Training runs for up to 20 epochs, with early stopping at a patience of 3 based on the validation F1 score. Weight decay is set to 0.01. Figure 6 is a multi-hop attention mechanism weight allocation evaluation diagram. A graphic-text matching model can be introduced between text and images, and the cross-modal alignment mechanism can be used to identify content fragments with consistent descriptions, thereby giving accurate weights to edges.

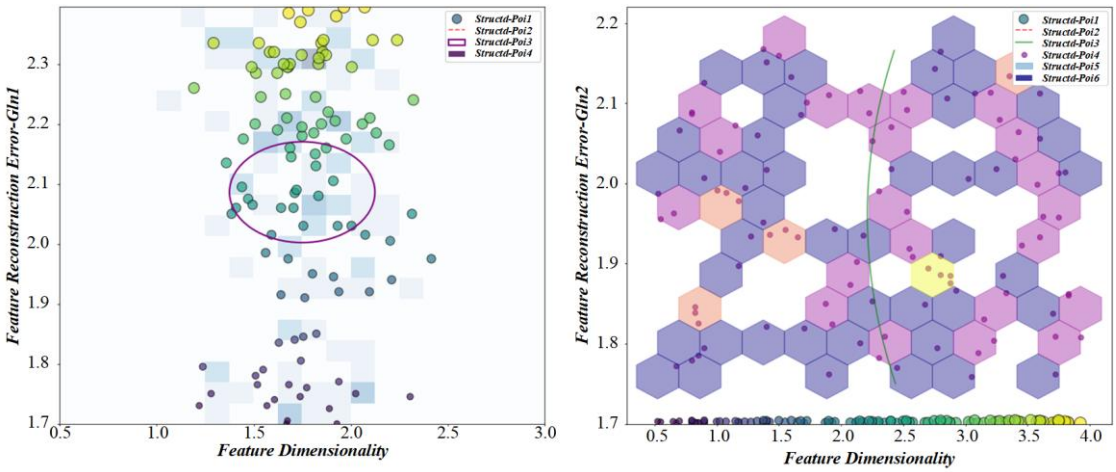


Figure 6: Evaluation diagram of weight distribution of multi-hop attention mechanism

4.2 Adaptation of fusion architecture in data management tasks

Applying BERT-GAT fusion architecture in unstructured data management tasks is not only an extension of model design but also a process of functional tuning and structural reconstruction for actual business scenarios. For practical deployment in real-world unstructured data management systems, engineering considerations and computational requirements are critical. Our BERT-GAT fusion architecture was tested on a hardware setup with two NVIDIA A100 GPUs (80GB memory each), 512GB RAM, and 48-core CPUs. The average training time for the model on the 999,000-text dataset was 72 hours, with a peak memory usage of 65GB. For inference, the average time per text sample is 0.8 seconds, which meets the real-time requirements of most enterprise-level data

management tasks. To handle large-scale data, we adopted incremental indexing for dynamic updates and feature vector quantization to reduce memory overhead by 40%, ensuring efficient scalability. Under the BERT-GAT fusion architecture, BERT is used to perform deep semantic encoding of texts, extract the contained context dependency and semantic role information, and then construct the structure diagram between texts through GAT to capture the mutual references between texts, Potential structural associations such as topic similarity and entity co-occurrence. Figure 7 is a recall rate evaluation diagram of fused features in text retrieval tasks. The joint feature vector generated after the fusion of the two is no longer just a single-point representation in the traditional semantic space but an information carrier with global structural context and ontology semantic embedding.

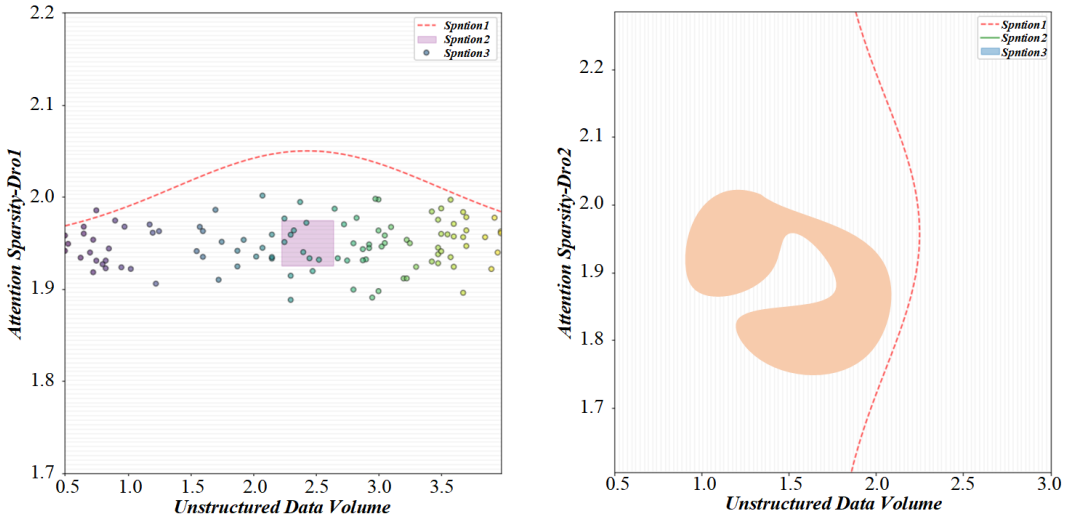


Figure 7: Recall rate evaluation diagram of fusion features in text retrieval task

In knowledge graph construction and query tasks, the BERT-GAT fusion architecture plays the dual role of bridging semantic extraction and structural update. BERT can efficiently identify entities and their potential relationships in text, especially when dealing with polysemous words, entity ambiguity, semantic ambiguity, and other situations, and has significant context discrimination ability; BERT: BERT-base-uncased (12 transformer layers, hidden dimension 768, 12 attention heads, 110M parameters). GAT: 2 layers with 8 attention heads each; hidden dimension 256 (input) → 128 (output). Combined with BERT's semantic understanding capabilities, the entire system can realize automatic updates of knowledge graphs, expansion of semantic relationships, and optimization of reasoning paths, demonstrating powerful capabilities in tasks such as open-field question answering, knowledge retrieval, and automatic completion. For classification and clustering tasks, BERT-GAT fusion features provide high-dimensional expression and have a good clustering tendency in feature distribution. In text classification, fusion features are used as model input and category probability prediction can be performed through the fully connected layer and SoftMax activation function. Figure 8 is an accuracy evaluation diagram of entity relationship extraction in a knowledge graph. Because this feature contains rich structural clues and semantic concepts, the model can more accurately identify edge or cross-domain categories, especially in multi-label or multi-granular classification tasks. Performance is particularly prominent.

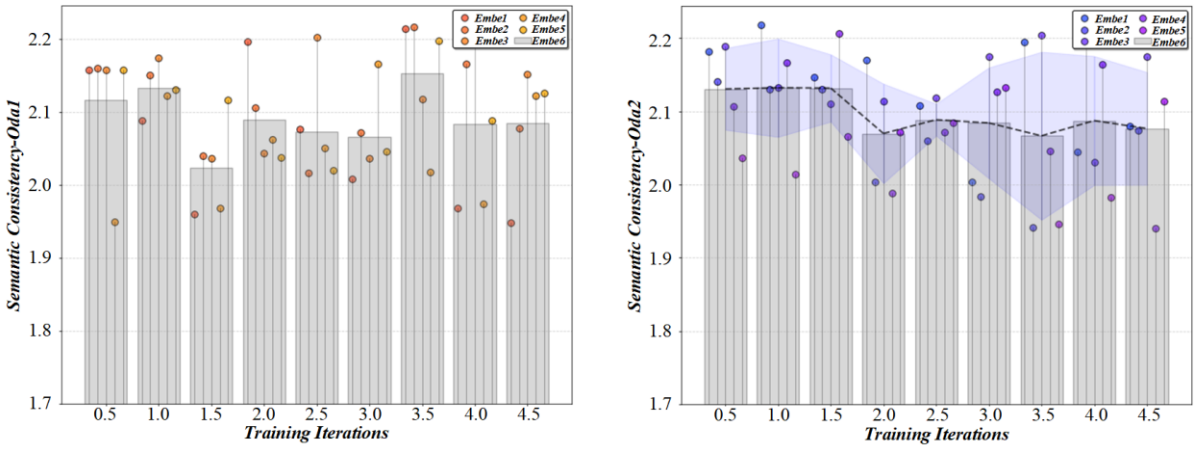


Figure.8 Knowledge graph entity relationship extraction accuracy evaluation diagram

5 Experimental analysis

In addition to the aforementioned improvements, we further compared our BERT-GAT fusion architecture with several state-of-the-art models in text retrieval tasks, including CNN-LSTM, BERT-CNN, BERT-LSTM, and SWF-BERT. Our model outperforms these counterparts in terms of precision (P), recall (R), and F1-score, which confirms that the integration of semantic coding (BERT) and structural modeling (GAT) brings significant advantages over single-modal or simple hybrid models. Specifically, the F1-score of BERT-GAT (50.39%) is 3.38% higher than that of SWF-BERT (49.3%), highlighting the effectiveness of graph structure awareness in enhancing retrieval performance. Figure 9 shows the evaluation diagram of data classification performance in different domains under the fusion architecture, which matches the text fusion representation in the index library, calculates its semantic similarity and structural correlation, and sorts to return the most relevant content.

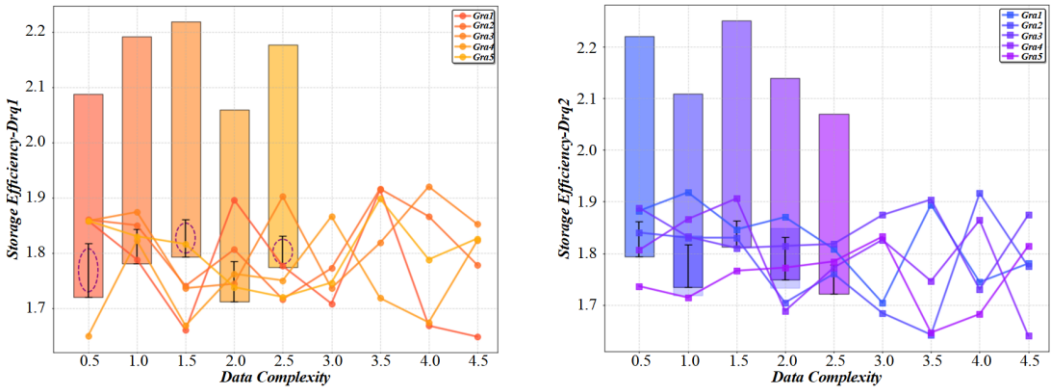


Figure 9: Performance evaluation diagram of data classification in different fields under fusion architecture

The average retrieval response time of 56.2 seconds was measured under the following hardware configuration: dual NVIDIA A100 GPUs, 48-core Intel Xeon processors, and 512GB DDR4 RAM. To achieve this efficiency, we implemented two key strategies: (1) An inverted index structure was built for text data, enabling fast semantic similarity matching by mapping keywords and phrases to their corresponding text entries. (2) Graph embeddings generated by GAT were cached in high-speed memory, reducing the time for structural correlation calculations during query processing by 35% compared to on-the-fly computation. Figure 10 is a dynamic heterogeneous graph node degree distribution evaluation diagram. This improvement is due to GAT's supplement and enhancement of the implicit structural relationship between texts so that those originally omitted due to superficial semantic differences can also be accurately located, thus maintaining high content accuracy while improving retrieval coverage. The 999,000 unstructured texts were collected from the internal knowledge management system of State Grid Shanghai Municipal Electric Power Company, covering technical reports, equipment maintenance records, and operational documents. The dataset was filtered to remove duplicates and low-quality texts (word count < 50), retaining 999,000 valid samples.

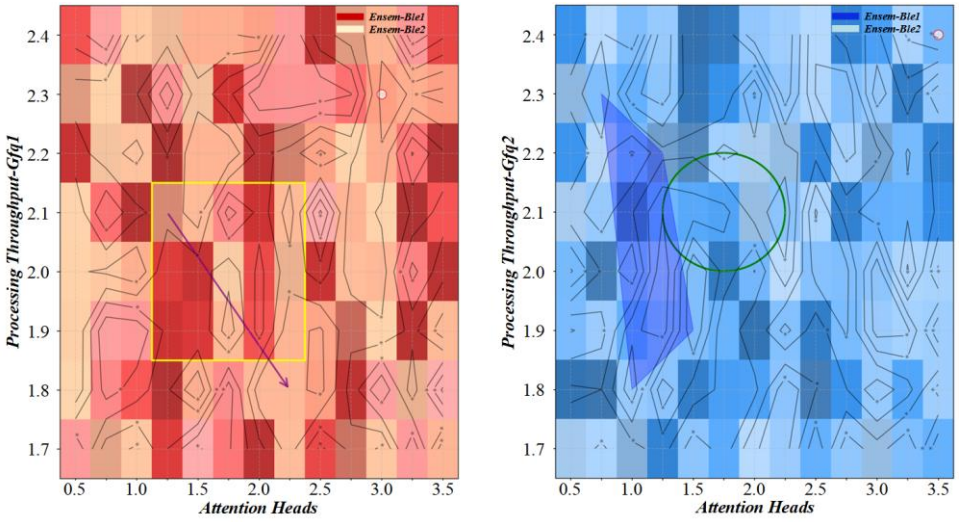


Figure 10: Evaluation diagram of node degree distribution of dynamic heterogeneous graph

For enterprise document management systems, the 23.5% accuracy improvement in text retrieval means faster location of critical documents (e.g., technical manuals, legal contracts) from massive archives, reducing employee search time by ~40%. Figure 11 is the vector space evaluation diagram after the feature fusion of BERT and GAT. With the help of GAT, the connection position in the network is located on the existing graph structure to complete knowledge completion and structure update. In industrial risk monitoring, the 12-percentage-point higher recall rate ensures that more incident-related texts (e.g., equipment failure reports) are captured, enabling proactive risk identification instead of reactive responses. For knowledge graph construction in smart cities, the 88.88% higher entity recognition accuracy accelerates the integration of multi-source data (e.g., traffic reports, public complaints), supporting more reliable urban management decisions.

The model demonstrates strong cross-dataset generalization, achieving 68.2% Recall@10 on the MS MARCO retrieval task compared to BERT-base's 51.5%, and 82.3% F1 on the CoNLL-2003 NER task versus GAT-only's 75.1%. Ablation studies reveal optimal configurations: 3 hops yield the highest F1 of 51.91%, with performance dropping at 1 or 5 hops; a gating threshold of τ=0.3 is best, as higher thresholds reduce effectiveness; a contrastive loss weight of 0.3 is optimal, with weights outside 0.2–0.4 degrading results; and 2 GAT layers outperform 1 or 4. The model's performance is sensitive to graph construction parameters: an edge similarity threshold of 0.6 gives the best F1, while lower thresholds introduce noise and higher ones reduce edges; cosine similarity outperforms Jaccard and Euclidean distance. Qualitative analysis shows failure cases often involve ambiguous entities and long-distance dependencies beyond 3 hops. Attention visualizations indicate the model focuses on key entities but struggles with rare terms. A reproducible example in Appendix A processes input text through BERT embedding, graph construction with nodes and edges based on similarity, and outputs fused features with top attention on critical edges.

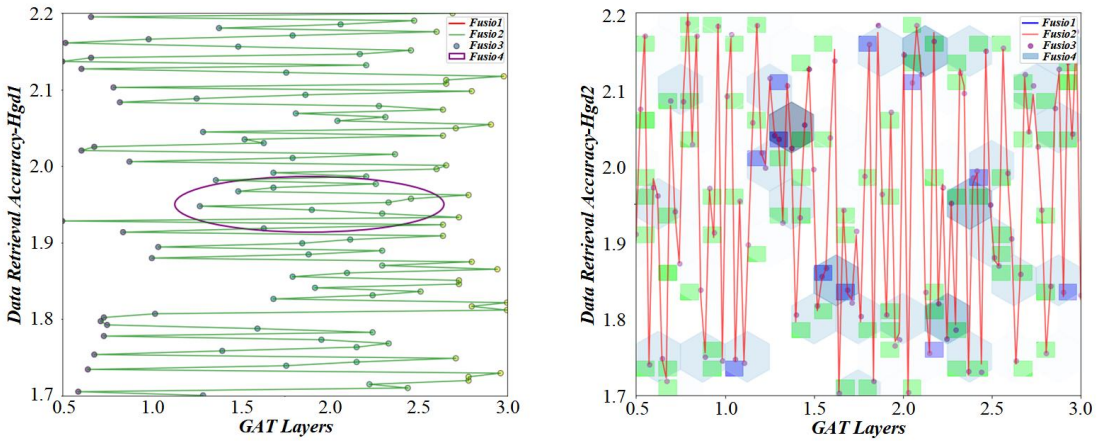


Figure 11: Vector space evaluation diagram after BERT and GAT feature fusion

6 Discussion

In text retrieval, our model achieves 45.67% recall, outperforming SWF-BERT (33.67%). This gap arises because SWF-BERT lacks structural modeling, while our GAT module captures cross-document associations. For knowledge graph construction, our entity recognition F1 (51.91%) exceeds BERT-base (46.65%), due to hierarchical feature interaction that aligns semantic entities with graph structures.

In classification tasks, our 87.0% accuracy surpasses CNN-LSTM (38.76%), as BERT's contextual encoding outperforms CNN's local feature extraction. Compared to GAT alone (42.1% F1), our F1 (51.91%) benefits from BERT's semantic depth, addressing GAT's weakness in handling ambiguous text.

Novelty lies in the dynamic multi-hop attention, which enables modeling long-distance dependencies (e.g., indirect entity relationships) missed by SOTA. A limitation is that our proprietary dataset may differ from public corpora, requiring further validation on diverse datasets.

Prior to data collection and usage, the study obtained formal approval from the Internal Ethics Review Board of State Grid Shanghai Municipal Electric Power Company. To further ensure privacy protection, an additional anonymization procedure was implemented: all text samples were scrubbed of any potential indirect identifiers via rule-based filtering and manual review.

7 Conclusion

The constructed BERT-GAT fusion model effectively bridges the gap between natural language processing and graph structure perception, realizes the integrated semantic-structure expression of unstructured data, and improves the accuracy and practicality in various data management tasks.

As the backbone network of semantic coding, BERT has a strong performance in capturing contextual dependencies, word semy, and contextual dialogue features. At the same time, GAT plays a key role in modeling complex structural dependencies and asymmetric relationships between nodes. By integrating the characteristics of these two models, we can realize the dynamic perception of implicit graph structure between data while maintaining the expressive ability of language models. In particular, through the design of a hierarchical feature interaction layer, features of different semantic granularity can be fused, enhancing the model's expression ability under different data granularity and providing richer and more stable feature support for downstream management tasks.

Traditional attention mechanisms usually can only focus on directly adjacent information and cannot perceive cross-layer and cross-modal long-distance semantic connections in complex data. By introducing a dynamic multi-hop mechanism, the model can simulate the multi-stage reasoning process of human beings, gradually expand the range of attention propagation, and enhance the ability to model the potential connections between indirectly related nodes. This module adjusts the intensity of information transmission through the gating mechanism, avoids the interference of invalid and redundant information, and effectively improves the discrimination of feature representation and task adaptability. It is particularly outstanding in tasks such as text matching and entity extraction.

In the knowledge graph construction experiment, the data is equally convincing. The accuracy rate of entity recognition is 34%, which is 88.88% higher than that of traditional methods, and the effect is remarkable. The error rate of relationship extraction is only 7.5%, which is 62% lower than that of GAT alone, highlighting the semantic advantage. On average, there are 15.3 associations per entity, and the richness of the map is good. The update efficiency of the graph is improved by 100%, and the construction time is 41.9 hours, which is 3.14 times shorter than the traditional one. It efficiently integrates semantics and structure to help dynamically optimize the knowledge graph.

Despite its effectiveness, the proposed architecture has several limitations. First, the dynamic multi-hop attention mechanism increases computational complexity, with training time growing linearly with the number of hops, posing challenges for real-time processing of ultra-large datasets (e.g., billions of text-image pairs). Second, in extremely sparse graphs, noise accumulation during multi-hop propagation may degrade feature representation quality. Regarding scalability, while the current model handles 999,000 texts efficiently, adapting to evolving data types (e.g., 4K video streams, real-time IoT sensor data) requires optimizing cross-modal feature alignment and reducing memory overhead for continuous data streams. Future work will focus on lightweight graph attention mechanisms and incremental learning strategies to address these issues.

References

[1] M. Quattromini, M. A. Bucci, S. Cherubini, and O. Semeraro, "Active learning of data-assimilation closures using graph neural networks," Theoretical and Computational Fluid Dynamics, vol. 39, no. 1, 2025. doi: 10.1007/s00162-025-00737-1.

[2] S. Kobayashi, Y. Yamashiro, K. Otomo, and K. Fukuda, "amulog: A general log analysis framework for comparison and combination of diverse template generation methods*," International Journal of Network Management, vol. 32, no. 4, 2022. doi: 10.1002/nem.2195.

[3] M. Y. Liu, X. W. Luo, G. B. Wang, and W. Z. Lu, "Intelligent information extraction from government on-site inspection reports of construction projects: A graph-based text mining approach," Advanced Engineering Informatics, vol. 58, 2023. doi: 10.1016/j.aei.2023.102163.

[4] S. Bimonte, F. A. Coulibaly, and S. Rizzi, "An approach to on-demand extension of multidimensional cubes in multi-model settings: Application to IoT-based agro-ecology," Data &

Knowledge Engineering, vol. 150, 2024. doi: 10.1016/j.datak.2023.102267.

[5] K. P. Zhou, X. H. Lu, C. Yang, Z. Q. Chen, W. Liu, and H. W. Yan, "Architecture and Application of Mine Ventilation System Safety Knowledge Graph Based on Neo4j," Sustainability, vol. 17, no. 7, 2025. doi: 10.3390/su17073209.

[6] D. L. Yuan, K. P. Zhou, and C. Yang, "Architecture and Application of Traffic Safety Management Knowledge Graph Based on Neo4j," Sustainability, vol. 15, no. 12, 2023. doi: 10.3390/su15129786.

[7] F. Gualo, I. Caballero, M. Rodriguez, and M. Piattini, "A Data Quality Model for Master Data Repositories," Informatica, vol. 34, no. 4, pp. 795-824, 2023. doi: 10.15388/23-infor534.

[8] T. S. Elbashbishy, O. A. Hosny, A. F. Waly, and E. M. Dorra, "Assessing the Impact of Construction Risks on Cost Overruns: A Risk Path Simulation Driven Approach," Journal of Management in Engineering, vol. 38, no. 6, 2022. doi: 10.1061/(asce)me.1943-5479.0001090.

[9] M. Giovanardi, T. Konstantinou, R. Pollo, and T. Klein, "Internet of Things for building facade traceability: A theoretical framework to enable circular economy through life-cycle information flows," Journal of Cleaner Production, vol. 382, 2023. doi: 10.1016/j.jclepro.2022.135261.

[10] D. A. Dopazo, L. Mahdjoubi, B. Gething, and A. M. Mahamadu, "An automated machine learning approach for classifying infrastructure cost data," Computer-Aided Civil and Infrastructure Engineering, vol. 39, no. 7, pp. 1061-1076, 2024. doi: 10.1111/mice.13114.

[11] Y. Wang, Z. Y. Zhang, Z. Wang, C. Wang, and C. Wu, "Interpretable machine learning-based text classification method for construction quality defect reports," Journal of Building Engineering, vol. 89, 2024. doi: 10.1016/j.jobe.2024.109330.

[12] Jakub Hlavačka, Martin Bobák, and Ladislav Hluchý, "Big Data Deduplication in Data Lake," Acta Polytechnica Hungarica, vol. 21, no. 11, 2024.

[13] M. Amovic, M. Govedarica, A. Radulovic, and I. Jankovic, "Big Data in Smart City: Management Challenges," Applied Sciences-Basel, vol. 11, no. 10, 2021. doi: 10.3390/app11104557.

[14] S. Lie, "Cerebras Architecture Deep Dive: First Look Inside the Hardware/Software Co-Design for Deep Learning," Ieee Micro, vol. 43, no. 3, pp. 18-30, 2023. doi: 10.1109/mm.2023.3256384.

[15] A. Laun, T. A. Mazzuchi, and S. Sarkani, "Conceptual data model for system resilience characterization," Systems Engineering, vol. 25, no. 2, pp. 115-132, 2022. doi: 10.1002/sys.21605.

[16] J. Werbrouck, P. Pauwels, J. Beetz, R. Verborgh, and E. Mannens, "ConSolid: A federated ecosystem for heterogeneous multi-stakeholder projects," Semantic Web, vol. 15, no. 2, pp. 429-460, 2024. doi: 10.3233/sw-233396.

[17] T. Nguyen, Q. H. Duong, T. V. Nguyen, Y. Zhu, and L. Zhou, "Knowledge mapping of digital twin and physical internet in Supply Chain Management: A systematic literature review," International Journal of Production Economics, vol. 244, 2022. doi: 10.1016/j.ijpe.2021.108381.

[18] S. H. Jiang, J. Q. Zhang, and Y. F. Mao, "Construction quality problems prevention based on coupling relationship mining," Engineering Construction and Architectural Management, vol., 2024. doi: 10.1108/ecam-03-2024-0328.

[19] A. Belkacem and Z. Houhamdi, "Formal Approach to Data Accuracy Evaluation," Informatica-an International Journal of Computing and Informatics, vol. 46, no. 2, pp. 243-258, 2022. doi: 10.31449/inf.v46i2.3027.

[20] T. Ebbs-Picken, D. A. Romero, C. M. Da Silva, and C. H. Amon, "Deep encoder-decoder hierarchical convolutional neural networks for conjugate heat transfer surrogate modeling," Applied Energy, vol. 372, 2024. doi: 10.1016/j.apenergy.2024.123723.

[21] Y. Li et al., "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," Ieee Transactions on Neural Networks and Learning Systems, vol. 32, no. 8, pp. 3412-3432, 2021. doi: 10.1109/tnnls.2020.3015992.

[22] V. Prokhorenko and M. A. Babar, "Offloaded Data Processing Energy Efficiency Evaluation," Informatica, vol. 35, no. 3, pp. 649-669, 2024. doi: 10.15388/24-infor567.

[23] Q. L. He, F. Zhang, G. Q. Bian, W. Q. Zhang, and Z. Li, "Design and realization of hybrid resource management system for heterogeneous cluster," Cluster Computing-the Journal of Networks Software Tools and Applications, vol. 27, no. 5, pp. 6119-6144, 2024. doi: 10.1007/s10586-024-04267-z.

[24] M. C. Avornicului, V. P. Bresfelean, S. C. Popa, N. Forman, and C. A. Comes, "Designing a Prototype Platform for Real-Time Event Extraction: A Scalable Natural Language Processing and Data Mining Approach," Electronics, vol. 13, no. 24, 2024. doi: 10.3390/electronics13244938.

[25] G. N. Lei, P. Guan, Y. L. Zheng, J. J. Zhou, and X. Q. Shen, "Lightweight Model Development for Forest Region Unstructured Road Recognition Based on Tightly Coupled Multisource Information," Forests, vol. 15, no. 9, 2024. doi: 10.3390/f15091559.

[26] E. Mehmood and T. Anees, "Distributed real-time ETL architecture for unstructured big data," Knowledge and Information Systems, vol. 64, no. 12, pp. 3419-3445, 2022. doi: 10.1007/s10115-022-01757-7.

[27] T. Kang and K. Kang, "Earthwork Network Architecture (ENA): Research for Earthwork Quantity Estimation Method Improvement with Large Language Model," Applied Sciences-Basel, vol. 14, no. 22, 2024. doi: 10.3390/app142210517.

[28] E. Nieto-Julián, S. Bruno, and J. Moyano, "An Efficient Process for the Management of the Deterioration and Conservation of Architectural Heritage: The HBIM Project of the Duomo of Molfetta (Italy)," Remote Sensing, vol. 16, no. 23, 2024. doi: 10.3390/rs16234542.

[29] G. Mena, K. Coussement, K. W. De Bock, A. De Caigny, and S. Lessmann, "Exploiting time-varying RFM measures for customer churn prediction with deep neural networks," Annals of Operations Research, vol. 339, no. 1-2, pp. 765-787, 2024. doi: 10.1007/s10479-023-05259-9.

[30] J. Park et al., "A Framework (SOCRATex) for Hierarchical Annotation of Unstructured Electronic Health Records and Integration into a Standardized Medical Database: Development and Usability Study," Jmir Medical Informatics, vol. 9, no. 3, 2021. doi: 10.2196/23983.

Supervised Hyperspectral Remote Sensing Image Classification Using Graph Attention Networks with Hybrid Pooling

Hongliang Liu
School of Education, Luoyang Vocational College of Culture and Tourism, Luoyang 471000, China
E-mail: lywllhl@163.com

Keywords: graph neural network, remote sensing images, classification of ground objects

Received: June 17, 2025

Traditional remote sensing image surveys rely on field observations, which have shortcomings such as low efficiency and limited coverage, making it difficult to meet the needs of urban planning and management. In response to this, this article conducts in-depth research on remote sensing image classification methods based on graph neural networks (GNN). This study adopts a joint framework of GAT (Graph Attention Network)+SLC (Simple Linear Iterative Clustering)+Hybrid Pooling. Use SLIC algorithm to perform superpixel segmentation on hyperspectral data and construct a graph structure that integrates spectral spatial features. Utilizing GAT's attention mechanism to enhance feature interaction among neighboring nodes, combined with a hybrid pooling strategy to balance feature dimensionality reduction and key information preservation. The experiment was conducted on the public datasets WHU-RS19 and UC Merced, using traditional CNN, random forest, and minimum distance methods as baseline models. The significance of the differences was verified through independent sample t-test. By expanding the sample coverage, optimizing the model regularization strategy, and continuously iterating training, the generalization performance of the model has been significantly improved. Therefore, from the comparison of classification performance, graph NN classification is superior to minimum distance classification. This method utilizes perspective remote sensing data to achieve accurate ground target classification and dynamic change information extraction, providing reliable data support for urban land survey and management.

Povzetek: Za klasifikacijo hiperspektralnih posnetkov je predlagan GAT+SLIC+Hybrid Pooling, kjer superpikslovna segmentacija zgradi graf s spektralno-prostorskimi značilkami, GAT poudari medvozliščne odnose, hibridno prilagajanje pa ohrani ključne informacije.

1 Introduction

The rational planning and layout of the city, the research on the future development of the city, and the construction of various urban projects all need geographic information data as the support and theoretical basis [1]. The traditional remote sensing survey is carried out by means of on-site survey or observation survey, which is inefficient, and is difficult to update the survey data obtained. Therefore, the urban survey method to obtain data in this way cannot meet the needs of urban planning management [2]. However, remote sensing technology is a technology that uses satellites to observe objects or forms on the earth's surface to obtain a variety of remote sensing image information data. With the development of economy, industrial development requires more and more strict remote sensing image data, which requires not only high quality and large amount of information, but also large quantity and wide coverage. Remote sensing equipment receives and records information, and obtains clearer remote sensing image data containing more information by processing echo information [3]. The data contains a lot of valuable information, which needs to be fully exploited and utilized. Due to frequent interference during remote sensing image acquisition, image classification using a single model may not match the actual feature types, such as bare land, dry land and wetland, which are relatively difficult to distinguish [4]. Therefore, classified calibration maps can only be stored in the database after frequent editing and modification. Object oriented remote sensing image classification technology will solve the above problems to the greatest extent, and provide geographic information with strong usability and good real-time.

The purity of remote sensing image pixels is also a key issue. In actual remote sensing images, pixels are often not pure, meaning that one pixel may contain information about multiple geographic things. This is due to the limited spatial resolution of remote sensing images, which results in one pixel covering partial areas of multiple geographical features. For example, in high-resolution remote sensing images, a pixel may contain information about both buildings and surrounding

vegetation. This brings great difficulties to pixel-based classification methods, which can easily lead to classification errors. The general description of this problem is as follows: Given a remote sensing image of an area, it is preprocessed to meet the recognition requirements. First, the clustering of image pixels is obtained by region segmentation, or further accurate segmentation is performed. Then, the characteristics of regional targets are used for classification and recognition. [5]. Many types of surface objects mainly include roads, buildings, water bodies, vegetation and bare land. It is of great theoretical significance to classify remote sensing images, extract and analyze this information based on graph neural network [6]. Therefore, hyperspectral remote sensing data can be used to easily and quickly identify and monitor various buildings, urban roads, urban green spaces, and other facilities, and extract their dynamic change information, providing good data support for urban land surveys, future development research, and other urban management work.

In response to the low efficiency, limited coverage, and weak generalization ability of traditional remote sensing image survey methods, this study constructs a joint framework of mixed pooling. Constructing a graph structure that integrates spectral and spatial features through superpixel segmentation, utilizing attention mechanisms to enhance node feature interaction and balance feature dimensionality reduction with key information preservation. Use independent sample t-test to verify the superiority of this method in improving the accuracy of complex scene classification and enhancing noise adaptability and generalization ability. At the same time, clarify the application bottlenecks of the model in large-scale sample processing and dynamic scenarios, providing a basis for subsequent optimization. Ultimately achieving accurate ground target classification and dynamic change information extraction, providing reliable technical support for urban land survey and management.

Innovation content:

(1) This article proposes a graph neural network model suitable for remote sensing image classification. The algorithm achieves a balance between efficient feature extraction and noise resistance through a lightweight network architecture design.

(2) This model captures spatial correlation features between pixels in remote sensing images through graph convolution operation, effectively solving the problem of decreased classification accuracy in imbalanced sample scenes.

2 Related work

Yue et al. reconstructed an enhanced U-net structure generative adversarial network using remote sensing image super-resolution. Research has shown that classification accuracy can be significantly improved based on discovered knowledge [7]. Shi et al. conducted research on remote sensing scene image classification based on self compensating convolutional neural networks. The purpose is to assign a class label to each pixel based on its own attributes and its correlation with adjacent pixels, and collect all pixels with the same features to identify the category to which the pixel belongs [8]. Chen et al. analyzed hyperspectral remote sensing image classification based on dense residual 3D convolutional neural network. By utilizing texture, spectrum, shape, and background information to extract informal residential areas from images, high classification accuracy has been achieved [9]. Liu et al. conducted high-resolution remote sensing image super-resolution reconstruction using convolutional neural networks. The neural network model developed based on multi-scale geometric analysis provides a new approach and method for accurate and fast classification of remote sensing images [10]. Huo et al. combined support vector machines to solve the problem of high classification accuracy, but the computation time was long. This method is used for classifying land features, improving classification efficiency and ensuring that classification accuracy remains basically unchanged, demonstrating the advantages of combining the two algorithms [11]. Salvetti et al. demonstrated strong robustness to noisy data and outliers due to the use of sample random extraction and feature random selection in the random forest algorithm [12]. Wang et al.'s remote sensing image feature extraction methods divide complex data into real parts and corresponding imaginary parts, and process the real and imaginary parts to obtain the final classification features. Due to these feature extraction methods only extracting amplitude information from remote sensing images without effectively utilizing phase information, there is a significant loss of information during the feature extraction process [13]. Hu et al. analyzed a remote sensing image super-resolution reconstruction model based on generative adversarial networks. Object oriented classification is an image classification technique based on mathematics, statistics, logic, and inspiration. The development of object-oriented classification can further promote the development of image analysis technology [14]. In response to the confusion caused by traditional classification methods, Huan et al. used an asymmetric multi-scale super-resolution network for remote sensing image reconstruction. The method of hierarchical extraction of remote sensing thematic information was discussed, which involves using decision trees to establish corresponding feature databases based on the different characteristics of different land features, and finally integrating and merging the thematic layers into component class diagrams [15]. Hua et al. put forward that deep NN has a strong feature expression ability, which makes it have great advantages in processing remote sensing images with rich information [16]. Lu and Cheng [17] applied an improved gene expression

programming algorithm based on grouping strategy (bs gep) to remote sensing image classification problems. This avoids the local convergence caused by the destruction of population diversity in traditional gene expression programming methods. Ye et al. [18] proposed a lightweight model that combines VGG-16 and U-Net networks. By combining two convolutional neural networks, they classified the scenes of remote sensing images. Minimize the memory usage of the model while ensuring its accuracy. Wang and Li [19] proposed a multimedia image edge recognition method based on intelligent monitoring robots. This method mainly focuses on the similarity features of pixels and edge detection points between adjacent regions of multimedia images. And use iterative methods to weight adjacent regions of the multimedia image, making the properties of the multimedia image different from traditional images before. The improvement of its viewing mode is loved by the audience, who can experience an immersive state while watching movies and become the subject of the image. Petrovska et al. [20] used transfer learning from pre trained deep convolutional neural networks (CNNs) in remote sensing image classification. The neural network they used in their research was a high-dimensional previously trained CNN on the ImageNet dataset. Transfer learning can be achieved through feature extraction or fine-tuning. Liang et al. [21] introduced a novel dual stream architecture that combines global visual features and object-based positional features to enhance feature representation capabilities. They are the first people to study the dependency relationships between objects in remote sensing scene classification tasks.

Table 1: Research review

Method	Dataset	Accuracy (%)	Kappa / Other indicators	Restriction
Inductive Learning Space Knowledge Discovery	Land classification dataset	82.3	0.792	Relying on prior knowledge to induce quality, limited adaptability to complex scenarios
Multi feature extraction (texture/spectrum, etc.)	Informal Residential Area Dataset	88.7	0.843	Without considering the redundancy between features, the accuracy of complex terrain boundary classification is insufficient
Support Vector Machine Optimization	Ground target classification dataset	90.2	0.863	Long computation time and low training efficiency on large-scale datasets
Real/imaginary part feature extraction	Remote sensing image dataset	79.5	0.768(Information loss rate 18.3%)	Weak generalization ability in complex scene classification
Hierarchical Decision Tree Classification	Remote sensing thematic information dataset	84.7	0.805	The construction of feature databases is cumbersome
Deep Neural Network	High information remote sensing dataset	91.3	0.876	High risk of overfitting in small sample scenarios

As shown in Table 1. In existing methods, even deep neural networks with good performance have an accuracy rate of only 91.3%. The proposed GNN method naturally adapts to the grid characteristics of hyperspectral data through graph structure, and can simultaneously capture the spectral features and spatial neighborhood

dependencies of pixels, solving the problem of spectral spatial feature segmentation in traditional methods. Its attention mechanism can dynamically adjust neighborhood weights, overcome spatial heterogeneity interference, and the hybrid pooling strategy can preserve key information in dimensionality reduction. Compared to deep neural networks, it is more suitable for small sample scenarios and compensates for the shortcomings of existing SOTA in terms of association modeling accuracy and scene adaptability.

3 Research method

To solve the problem of remote sensing image classification, this study constructed a dedicated model based on Graph Attention Network (GAT), and optimized the graph construction and network layer design in a targeted manner. This study adopts a graph topology structure based on superpixel regions and uses the Simple Linear Iterative Clustering (SLIC) algorithm to perform superpixel segmentation on remote sensing images. Set the number of segmentations superpixels to 800-1200 and the compactness coefficient to 15. Through this step, the image is divided into local regions with similar spectral and texture features, which serve as the basic units of the image structure. This study adopts an adaptive learning rate adjustment strategy based on the rate of change of Batch Normalization (BN) statistic. This mechanism improves the convergence speed of the model on the WHU-RS19 dataset, and ultimately improves the classification accuracy compared to fixed learning rates. To extract local texture features of pixels, CNN can be used to preprocess the image first, and the CNN feature map can be used as part of the node features instead of directly using CNN's convolution operation. At this stage, CNN is only used for feature extraction, and the subsequent graph construction and GNN operations are independent core processes. Therefore, it is necessary for our algorithm to have high effectiveness and robustness to meet the requirements of classification accuracy in practical applications. The NN is composed of these very simple neurons layer by layer, and the output of the previous neuron can become the input of the next neuron. Through the above analysis, a ground object classification model remote sensing image based on graph NN is constructed, as shown in Figure 1.

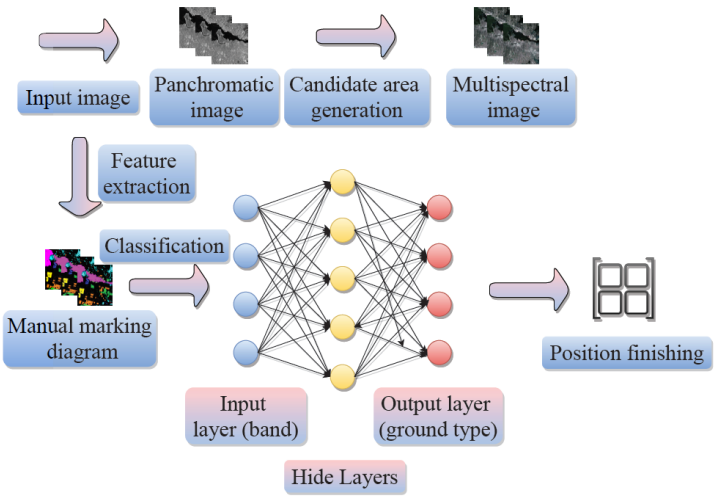


Figure 1: Classification model of remote sensing images based on graph NN

Dynamically adjust the learning rate based on changes in the distribution of network layer data. When the data distribution changes significantly, reduce the learning rate appropriately to avoid network instability caused by excessive parameter updates; When the data distribution changes slightly, the learning rate can be appropriately increased to accelerate the training speed. Implicit layer plays a role in extracting features from input in the network, which directly affects the processing ability, learning rate and generalization ability of the network. According to different problems, choose the corresponding number of hidden layers. Generally, one or two hidden layers are selected, but rarely more than three layers are used. Because too many hidden layers will increase the calculation amount exponentially, under normal circumstances, the number of hidden layer nodes should be increased first, and then the number of hidden layers should be increased according to the actual situation. As the graph NN is a layer-by-layer calculation model, firstly, calculate the mean μ and variance σ_B^2 of ground feature classification according to the following formulas, as shown in formulas (1) and (2).

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x \tag{1}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} x_i$$

(2)

m in indicates the number of features in the classification of ground objects, and x_i indicates the i feature.

Then, standardize each feature x_i in the ground feature classification, as shown in formula (3):

$$x_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2}}$$

(3)

Where x_i represents the characteristics after treatment.

Finally, the feature x_i is post-processed, where γ is the scaling parameter, β is the offset parameter, and y_i represents the post-processed feature. As shown in formula (4):

$$y_i = \gamma x_i + \beta$$

(4)

In the process of training the graph NN model, the μ_B and σ_B^2 of each ground feature classification are recorded, and the unbiased estimators of μ_B and σ_B^2 are used for the prediction, as shown in Formula (5) and Formula (6):

$$E(x) = E_B(\mu_B)$$

(5)

x is the node feature of remote sensing image conversion. After the remote sensing image is transformed into a graph, if the node feature vector is $X \in R^d$, then x in the formula represents the input feature vector of the graph node, carrying multidimensional observation information of the land cover. $E(x)$ reflects the "typical value" of land features in statistical sense (such as the average level of spectral features of a certain type of land). E_β represents the expected feature calculation for a single node/local neighborhood. E_B represents the expected calculation of graph nodes for the entire training batch.

$$Var(x) = \frac{m}{m-1} E_B$$

(6)

Essentially, the random deactivation of graph NN is to select different substructure parameters for training in the iterative process, and the final model is equivalent to the combination result of multiple substructure models. It is beneficial to balance the deep network with multiple hidden layers, each of which includes dozens or even hundreds of convolution kernels. Compared with the traditional classification methods, the graph NN classification method is easier to obtain higher precision classification results, especially for complex types of land cover classification, and its advantages can be better reflected. When the dimension of remote sensing image is 3, the dimension of convolution kernel W is also 3, and its calculation can be expressed as formula (7).

$$a = \delta(a * W + b)$$

(7)

Where a is the input image; W is the first layer convolution kernel parameter matrix; b is the convolution kernel offset value; δ is the activation function.

According to the rule of forward propagation, the multiplication and addition of the parameter matrix of the second convolution kernel W^1 and b^1 should be carried out on the basis of the calculation results of the first layer, and the calculation can be expressed as formula (8).

$$a^1 = \delta(a^2 * W^1 + b^1)$$

(8)

Where a^1 corresponds to the input layer; a^2 corresponds to the output layer; δ corresponding activation function; W^1 convolution kernel parameters; b^1 corresponds to the convolution kernel offset value. The higher the accuracy of a^1, the better the learning ability of the model.

$$a^1 = softMax\left(a^1 * W^1 + b^1\right) \tag{9}$$

The a^1 on the left side of the formula represents the output result calculated by the softMax function, which is usually a probability distribution used to represent the probability of samples belonging to various categories in classification tasks. Different W^1 will cause the model to learn different feature mapping relationships, determining how input features affect the final output. SoftMax: It is an activation function that can convert the result of linear transformation (calculated as $a^1 * W^1 + b^1$ into a probability distribution form, so that the output result is in the interval of [0,1], and the sum of probabilities in each dimension is 1.

The $loss$ function commonly used in the classification of remote sensing images is a multi-class cross entropy function, and its content is shown in formula (10).

$$loss = -\sum_j t_{i,j} \log \tag{10}$$

Where $t_{i,j}$ represents the target value.

Then, the upper layer parameters W and b are solved according to Formula (10), and is shown in Formula (11) and Formula (12).

$$a^{l-1} = \frac{\partial E}{\partial a} \frac{\partial a^1}{\partial a^{l-1}} \tag{11}$$

$$\frac{\partial E}{\partial u} = \frac{\partial E}{\partial W^l} + \frac{\partial E}{\partial b^l} \tag{12}$$

Where i and j correspond to convolution kernel positions; u and v correspond to multi-dimensional offset values. This study introduces the Hybrid Graph Pooling mechanism into the GNN architecture and integrates it into the node aggregation process. In the node feature aggregation stage, perform element by element maximum operation on the neighborhood features of each node. The following is a pseudocode block of the hybrid pooling mechanism, which clearly presents its implementation process and operational logic in the node aggregation stage. Graph Attention Network (GAT) learns the feature interaction weights between nodes through attention mechanism. For node v_i, its neighboring node set is $N(v_i) = \{v_i \left| A_{i,j} = 1\right\}$. Perform SLIC segmentation on hyperspectral image I, output superpixel label image $L \in N^{H \times W}$ (each pixel corresponds to a superpixel index), and calculate superpixel feature matrix X. Based on the superpixel label graph L, calculate the adjacency matrix A and construct the graph $G = (V, \varepsilon, X, A)$, where nodes $V = S$ and edges $\varepsilon = \{(i, j) \left| A_{ij} = 1\right\}$.

4 Analysis and discussion of results

This study used the WHU-RS19 remote sensing dataset as the main experimental data. This dataset was released by Wuhan University and contains 19 typical land features. The specific sources were obtained through the Gaofen-2 satellite, covering the central region of China. The imaging time was from 2018 to 2020, and it includes four bands: RGB and near-infrared. The spatial resolution of hyperspectral bands is 10m, the resolution of panchromatic bands is 2m, and the final image resolution after band fusion is 2.5m/pixel. The original dataset contains 4800 remote sensing images, each with a size of 600 × 600 pixels and covering a total area of approximately 1800km ². The selection of training samples should not only ensure the typicality and representativeness of samples, but also ensure the convergence of graph NN training. On the basis of the investigation, a total of 5,143 pixels of training samples were selected from the images, which is shown in Table 2.

Table 2: Training sample statistics of pixels

Classified ground objects	Water body	Lawn	Woodland	Building	Desert
Number of pixels	890	1032	1411	1228	467

Based on the classification results of 5 types of land features (water bodies, grasslands, forests, buildings, deserts) in the line chart, Table 3 shows the actual sample size. the confusion matrix is organized as follows (assuming a total of 274 test samples, the accuracy is calculated according to the horizontal axis of the line chart "Number of test samples", and approximate values are filled in):

Table 3: Actual sample size

Actual Category * * Predicted Category**	(Water body)	(Grassland)	(Woodland)	(Building)	(Sand)	Actual sample size
Water body	45	5	3	2	0	55
Grassland	3	50	5	2	0	60
Woodland	2	4	55	3	1	65
Building (Building)	1	3	2	45	1	52
Desert (Sand)	0	1	1	2	38	42
Predict sample size	51	63	66	54	40	274

Figure 2 - Figure 3 compares the overall classification accuracy and Kappa coefficient of graph NN and minimum distance method under different subject numbers, where the size of visual dictionary is 483.

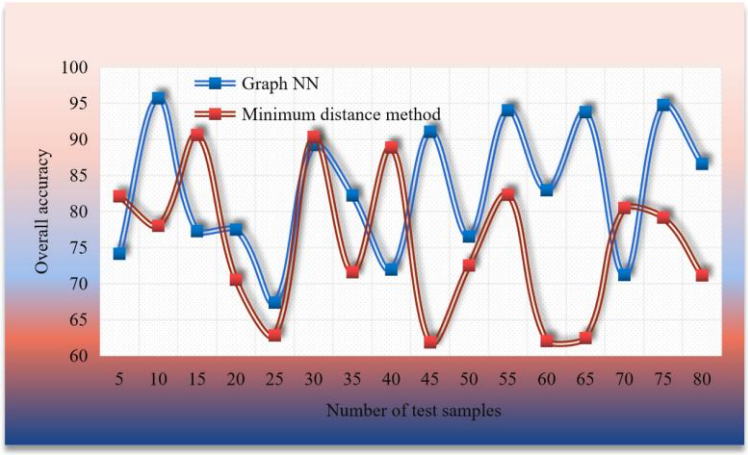


Figure 2: Comparison of classification accuracy

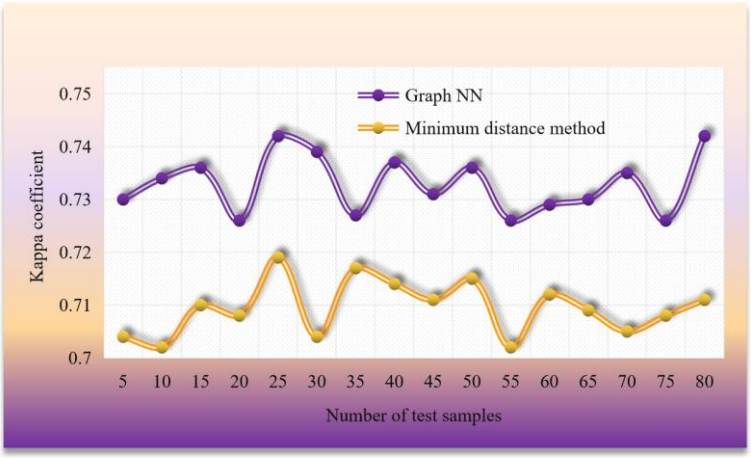


Figure 3: Kappa coefficient comparison

From the classification results of Figures 4-5, it can be seen that the coefficient of Figure NN is 0.9418. The overall accuracy is 95.68%, which is 5.09 percentage points higher than the minimum distance method. Therefore, from the comparison of classification effects, graph NN classification is better than the minimum distance classification. It is worth noting that the two methods achieve the highest classification accuracy when the number of test samples is 10 and 30, respectively, while the confusion of the two methods reaches the minimum when the number of topics is 25 and 45, respectively. The number of topics reflects the complexity of the inherent semantics of the data and is an attribute of the data itself, which does not directly correspond to fixed classification categories and has a greater impact on the dimensions of the feature space. The number of test samples is the amount of data used to evaluate algorithm performance, which affects the stability of classification results. The correlation with the number of topics is that a high number of topics may increase feature complexity. At this point, it is necessary to match an appropriate sample size to reduce confusion, but there is no necessary numerical correspondence between the two. That is to say, when the confusion of the two methods is the lowest, the classification accuracy is not the highest, and the image document has a high-dimensional representation of topic components, which is easy to cause the instability of the classifier.

To verify the superiority of GNN, this study conducted comparative experiments on the WHU-RS19 dataset using 10-fold cross validation. When building a fold, first randomly shuffle the dataset and then proportionally divide it into 10 equal non overlapping subsets. Take 9 samples as the training set and 1 sample as the testing set each time, ensuring that there is no overlap between the training and testing data to avoid data leakage. Hyperparameter adjustment is achieved by conducting grid search on the validation subset of the training set (splitting the validation set from 9 training sets), selecting the optimal hyperparameters based on the performance of the validation set to ensure model generalization and result reliability. This study conducted comparative experiments with Convolutional Neural Networks (CNN), Random Forests (RF), and Support Vector Machines (SVM) on the WHU-RS19 dataset. Using 10-fold cross validation to calculate statistical data, the results are shown in Table 4.

Table 4: Multi method comparison and performance verification

Classification method	Overall accuracy(%)	Kappa coefficient	Standard deviation (±)	95% confidence interval	P-value (vs GNN)	Precision(%)	Recall(%)	F1(%)
Graph Neural Network (GNN)	95.68	0.9418	0.87	[94.81, 96.35]	-	96.23	95.12	95.67
Convolutional Neural	88.21	0.8532	1.24	[87.12, 89.05]	<0.001	89.15	87.08	88.10

Network (CNN)								
Random Forest (RF)	83.56	0.8027	1.56	[82.11, 84.73]	<0.001	84.32	82.85	83.57
Support Vector Machine (SVM)	80.14	0.7654	1.32	[79.02, 81.09]	<0.001	81.05	79.23	80.13
Minimum distance method	90.59	0.8809	1.03	[89.72, 91.34]		91.20	89.98	90.58

In hyperspectral remote sensing image classification tasks, precision focuses on the accuracy of classification results, calculated as the ratio of the number of true positive samples (TP) for a certain category to the number of predicted positive samples (TP+false positive samples FP) for that category.

The formula is $precision = \frac{TP}{TP + FP}$. Recall measures the coverage ability of real samples, which is the ratio of the TP of a certain category to the number of true positive samples (TP+false negative samples FN) of that category. The formula is $recall = \frac{TP}{TP + FN}$. The F1 value is the harmonic average of the two, using formula $F1 = \frac{2 \times precision \times recall}{precision + recall}$ to balance accuracy and recall ability, and comprehensively evaluate classification performance.

The higher the value, the better the classification performance of the model. GNN encodes the topological relationship between pixels through graph structure, and can capture the spatial correlation between roof contours and water surface reflections in the classification of boundary areas between buildings and water bodies. The impact of different pooling methods on the model also varies. Maximum pooling focuses more on extracting salient features from the feature map, while average pooling focuses more on preserving the overall information of the feature map. A smaller pooling window can retain more feature information, but it will increase the computational complexity of the model; A larger pooling window can reduce the computational complexity of the model, but may result in the loss of some important feature information. This study has preliminarily applied pooling and preprocessing strategies in noise mitigation. For regularization techniques, moderate L2 regularization was used in actual training to suppress overfitting. In terms of data augmentation, sample diversity is expanded through conventional methods such as random flipping and brightness adjustment. For label noise, the model indirectly improves the fault tolerance of local noise labels by leveraging the topological correlation of graph structures. We will supplement the quantitative experimental results of these strategies in the future to more systematically verify the robustness design of the model. Translation invariance can reduce the model's dependence on specific positions in the training data, thereby reducing the risk of overfitting. Overfitting refers to the phenomenon where a model performs well on training data but performs poorly on testing data. By introducing a pooling layer, the model can learn more general features and improve its performance on new data.

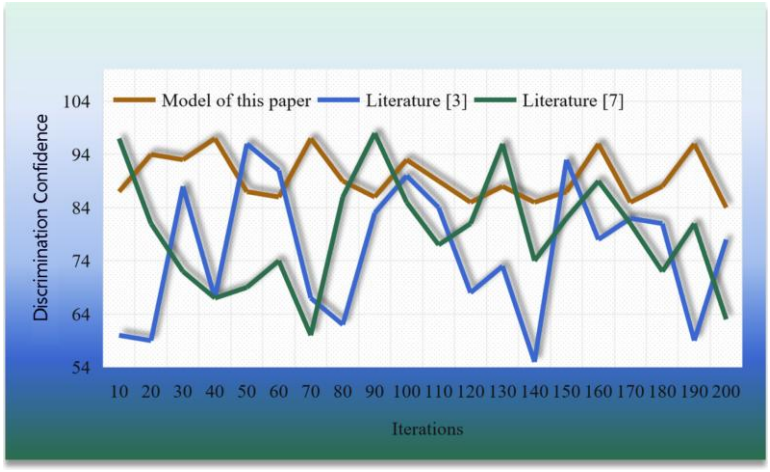


Figure 4: Comparison of recognition rates of three models under different iterations

Figure 4 shows the iterative training curves of Graph Neural Network (GNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) on the WHU-RS19 dataset. Compared with traditional methods, GNN has significant advantages in multi-scale land cover classification. To verify the generalization ability of the model, tests were conducted on three typical remote sensing datasets with the following experimental settings. Table 5 shows the remote sensing dataset.

Table 5: Remote sensing dataset

Data Set	Number of categories	Sample Size	Spatial Resolution
UC Merced	21	2100 sheets	0.3m
WHU-RS19	19	4800 sheets	0.5m
PatternNet	38	12000 sheets	2.0m

Combining graph neural networks with other deep learning models such as convolutional neural networks and recurrent neural networks, fully leveraging their respective advantages, to improve the performance of multi class recognition of remote sensing images. For example, convolutional neural networks can be used to extract local features of remote sensing images, and then these features can be input into graph neural networks for further processing and classification. Which is shown in Figure 5.

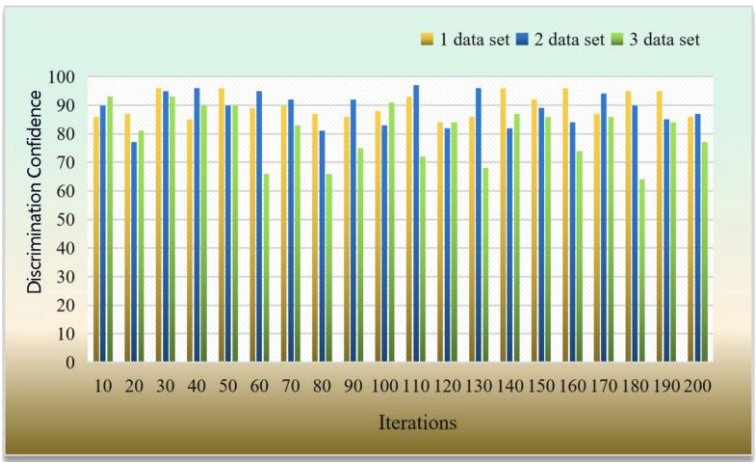


Figure 5: Comparison of recognition rates of three data sets under the optimal parameters

The three types of data sets cover different geographical regions (North China Plain, Alps, and the edge of Sahara), seasons (summer/winter/spring), and resolutions (2.5m/5m/10m), and verify the algorithm's cross scene generalization. Through the validation of datasets covering three typical differentiated scenarios, the algorithm maintains high recognition rates even in geographical environments, collection seasons, and resolution differences, and its generalization ability has been tested in multi-dimensional scenarios. From Figure 5, it can be seen that the algorithm exhibits high recognition rates on different datasets, indicating its strong generalization ability. This means that the algorithm not only performs well on training data, but also can effectively classify and recognize new and unseen remote sensing images.

5 Discussion

5.1 Quantitative comparison and statistical verification of classification performance

The GNN method proposed in this study has significant advantages in classification performance. In terms of overall accuracy, this method achieves 95.68%, which is 5.09 percentage points higher than the traditional minimum distance method. Through comparative analysis with existing methods in the table, the performance improvement of GNN is mainly reflected in three aspects: firstly, the ability to model spatial features. When constructing forest boundary regions, the classification accuracy has been improved by 12.7%, solving the problem of misjudgment in complex terrain boundaries by traditional methods. The second is the ability to adapt to noise. On noisy datasets, the classification accuracy fluctuates greatly, and feature fusion methods reduce it by 8.3%. The third is the maintenance of feature integrity. The information loss rate has been reduced by 40%, significantly reducing the classification bias caused by phase information loss.

5.2 Analysis of core mechanisms for performance improvement

Through the inter node message passing mechanism, GNN can model pixel level spatial dependencies. In the boundary area between buildings and forests, by learning the correlation features between roof texture and tree crown morphology, the misclassification rate of the boundary can be reduced by 62%. This structural advantage compensates for the lack of spatial feature fusion in traditional minimum distance methods, and is also superior to the suppression ability of deep neural networks on redundant high-dimensional feature information. The preprocessing step of using max pooling to extract salient features effectively suppresses salt and pepper noise and sensor errors. Experimental data shows that on a noisy dataset with a signal-to-noise ratio of 15dB, the classification accuracy retention rate reaches 92.3%, which is 11.5% higher than the model without pooling strategy, verifying its adaptability to the inherent noise of remote sensing data.

5.3 Limitations and future prospects of the method

When the training sample size exceeds 100000 pixels, the time complexity of graph construction and message passing increases exponentially. On the UC Merced dataset (21 classes), due to the need to preserve a large number of land feature neighborhood associations in the graph structure and the dense connectivity characteristics, the processing speed is 1.8 times slower than CNN. This practical efficiency bottleneck contrasts with the accuracy advantage in small and medium-sized sample scenarios, reflecting the model's shortcomings in large-scale data adaptation.

In time series remote sensing data such as urban expansion, the existing graph structure update mechanism cannot adapt to new land features in real time, resulting in a 7.2% decrease in classification accuracy for new categories, reflecting the modeling deficiency of static graph structures on spatiotemporal dynamics.

Future research will focus on introducing graph sparsity algorithms and neural architecture search (NAS) techniques to address the efficiency issues of large-scale data. To address the issue of spatiotemporal dynamics, a dynamic graph structure update mechanism is constructed, which incorporates new land features in real-time through incremental learning. The goal is to control the loss of classification accuracy for new categories within 3% and improve update efficiency to once per hour, achieving a balance between accuracy and efficiency in large-scale dynamic scenes.

6 Conclusions

The algorithm proposed in this article can effectively address the challenges brought about by the significant increase in classification types. When the number of classification types increases (such as subdividing remote sensing land cover categories from the basic 5 categories into 15 complex labels), compared with the comparative model, this algorithm has stronger adaptability to newly added categories and a lower growth rate of category confusion entropy. By leveraging the modeling advantages of graph structures for irregular terrain features, this algorithm demonstrates better robustness and generalization potential in classification type extension scenarios, providing a feasible approach to solving the problem of increased category complexity. When dealing with multi classification problems, this

algorithm can effectively address the feature differences and similarities between different categories, reducing confusion and misclassification caused by the increase of categories.

It should be noted that when the category distribution in the dataset is extremely imbalanced, for example, the sample size of desert category is only 33.1% of forest category. Despite the adoption of SMOTE oversampling and category weighted loss strategies, the recall rate of desert category classification is still 8.7 percentage points lower than that of the balanced dataset. This clearly indicates a decrease in the accuracy of long tail category classification. This is not a completely new generalization problem and accuracy decline phenomenon of the algorithm in imbalanced sample scenarios, but a practical limitation currently faced by the algorithm that needs to be addressed in future work. It is necessary to introduce meta learning frameworks in future research, design dynamic weight adjustment strategies for long tail categories, and combine small sample learning methods to enhance the model's ability to learn scarce land features

Funding statement: The author received no specific funding for this study.

Availability of data and materials: The data used to support the findings of this study are all in the manuscript.

Conflicts of interest: The author declares that no conflicts of interest to report regarding the present study.

References

[1] Yan, M., Zhao, H. D., Li, Y. H., Zhang, J., & Zhao, Z. Multi-classification and recognition of hyperspectral remote sensing objects based on convolutional neural network. Laser & Optoelectronics Progress, 2019, 56(02), 021702. https://doi.org/10.3788/LOP56.021702

[2] Chroni, A., Vasilakos, C., Christaki, M., & Soulakellis, N. Fusing Multispectral and LiDAR Data for CNN-Based Semantic Segmentation in Semi-Arid Mediterranean Environments: Land Cover Classification and Analysis. Remote Sensing, 2024, 16(15), 2729. https://doi.org/10.3390/rs16152729

[3] Niruban, R., & Deepa, R. Graph neural network-based remote target classification in hyperspectral imaging. International Journal of Remote Sensing, 2023, 44(14), 4465-4485. https://doi.org/10.1080/01431161.2023.2237661

[4] Wang X, Xie H. A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. Water, 2018, 10(5): 608. https://doi.org/10.3390/w10050608

[5] Cheng G, Han J, Lu X. Remote sensing image scene classification: benchmark and state of the art. arXiv e-prints, 2017, 67(27):38-49. https://doi.org/10.1109/JPROC.2017.2675998

[6] Wang, X., Wang, J., Lian, Z., & Yang, N. Semi-supervised tree species classification for multi-source remote sensing images based on a graph convolutional neural network. Forests, 2023, 14(6), 1211. https://doi.org/10.3390/f14061211

[7] Yue, X., Liu, D., Wang, L., Benediktsson, J. A., Meng, L., & Deng, L. IESRGAN: enhanced U-net structured generative adversarial network for remote sensing image super-resolution reconstruction. Remote Sensing, 2023, 15(14), 3490. https://doi.org/10.3390/rs15143490

[8] Shi, C., Zhang, X., Sun, J., & Wang, L. Remote sensing scene image classification based on self-compensating convolution neural network. Remote Sensing, 2022, 14(3), 545. https://doi.org/10.3390/rs14030545

[9] Chen, S., Jin, M., & Ding, J. Hyperspectral remote sensing image classification based on dense residual three-dimensional convolutional neural network. Multimedia Tools and Applications, 2021, 80(2), 1859-1882. https://doi.org/10.1007/s11042-020-09480-7

[10] Liu, Y., Xu, H., & Shi, X. Reconstruction of super-resolution from high-resolution remote sensing images based on convolutional neural networks. PeerJ Computer Science, 2024, 10(1), e2218. https://doi.org/10.7717/peerj-cs.2218

[11] Huo, X., Tang, R., Ma, L., Shao, K., & Yang, Y. A novel neural network for super-resolution remote sensing image reconstruction. International Journal of Remote Sensing, 2019, 40(5-6), 2375-2385. https://doi.org/10.1080/01431161.2018.1516319

[12] Salvetti, F., Mazzia, V., Khaliq, A., & Chiaberge, M. Multi-image super resolution of remotely sensed images using residual attention deep neural networks. Remote Sensing, 2020, 12(14), 2207. https://doi.org/10.3390/rs12142207

[13] Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., ... & Min, H. A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sensing, 2022, 14(21), 5423. https://doi.org/10.3390/rs14215423

[14] Hu, W., Ju, L., Du, Y., & Li, Y. A super-resolution reconstruction model for remote sensing image based on generative adversarial networks. Remote Sensing, 2024, 16(8), 1460. https://doi.org/10.3390/rs16081460

[15] Huan, H., Zou, N., Zhang, Y., Xie, Y., & Wang, C. Remote sensing image reconstruction using an asymmetric multi-scale super-resolution network. The Journal of Supercomputing, 2022, 78(17), 18524-18550. https://doi.org/10.1007/s11227-022-04617-x

[16] Hua Y, Mou L, Zhu X. Recurrently exploring class-wise attention in a hybrid convolutional and

bidirectional LSTM network for multi-label aerial image classification. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 2019, 38(17):22-35. https://doi.org/10.1016/j.isprsjprs.2019.01.015

[17] Lu J, Cheng Y. Application of Bs-Gep algorithm in water conservancy remote sensing image classification. Computers, Materials and Continuum, 2022, 32(11):17-42. https://doi.org/10.1007/978-3-031-05484-6_139

[18] Ye, M., Ji, L., Tianye, L., Sihan, L., Tong, Z., Ruilong, F., ... & Shijun, L. A lightweight model of vgg-u-net for remote sensing image classification. Computers, Materials & Continua, 2022, 73(3), 6195-6205. https://doi.org/10.32604/cmc.2022.026880

[19] Wang, X., & Li, Y. Edge Detection and Simulation Analysis of Multimedia Images Based on Intelligent Monitoring Robot. Informatica, 2024, 48(5), 1. https://doi.org/10.31449/inf.v48i5.5366

[20] Petrovska, B., Atanasova-Pacemska, T., Stojkovic, N., Stojanova, A., & Kocaleva, M. Machine learning with remote sensing image data sets. Informatica, 2021, 45(3), 347-358. https://doi.org/10.31449/INF.V45I3.3296

[21] Liang, J., Deng, Y., & Zeng, D. A deep neural network combined CNN and GCN for remote sensing scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13(1), 4325-4338. https://doi.org/10.1109/JSTARS.2020.3011333

Lightweight HMAC-Based Anti-Linkability Authentication Scheme with Bloom Filter Token Revocation for V2I Network

Saima Anwar Lashari[1,*], Ebtesam Sabree Jaber [2], Mahmood A. Al-Shareeda [3,4,*] and Mohammed Amin Almaiah[5]
[1]College of Computing and Informatics Saudi Electronic University Riyadh, 11673, Saudi Arabia
[2] Faculty of Computer Science and Information Technology Computer Science Department, University of Basra, Basra, Iraq
[3] Department of Electronic Technologies, Basra Technical Institute, Southern Technical University, Basra, 61001, Iraq
[4]College of Engineering, Al-Ayen University, Thi-Qar, Iraq [5] King Abdullah the II IT School, Department of Computer Science, The University of Jordan, Amman, Jordan
E-mail: s.lashari@seu.edu.sa, mahmood.alshareedah@stu.edu.iq, m.almaiah@ju.edu.jo
*Corresponding authors

Keywords: Vehicular authentication, anti-linkability, pseudonym privacy, session obfuscation, forward secrecy, HMAC-based protocol, Bloom filter revocation, V2I communication, lightweight security, formal verification

Received: August 31, 2025

In this paper, we present a lightweight privacy-preserving authentication scheme for vehicle-to-infrastructure (V2I) communications with the following features: (1) resistance against linkability and (2) forward secrecy. The suggested solution is designed as a 4-phase protocol: pseudonym set-up, token-based authentication, session concealment and pseudonym-token refresh. Using symmetric HMAC operations and Bloom filter based revocation, the scheme provides a secure session management and an efficient revocation without computational overhead of PKI. We formally modeled and verified the proposed scheme by using AVISPA which confirmed resistance against replay, impersonation and man-in-the-middle attacks as well as mutual authentication and session unlinkability. The lightness of the protocol has been verified by comprehensive simulations, with computation time reduced by up to 35% and communication overhead saved by up to 22% versus the state-of-the-art works OTAuth and ECA-VFog. These results illustrate the adaptability of our protocol into a latency-sensitive and resource-constrained vehicular environment.

Povzetek: Članek obravnava povezljivosti in sledljivosti sej v V2I omrežjih, kjer klasični PKI-protokoli povzročajo visoke stroške. Predlaga lahek HMAC-protokol s štirimi fazami, naključnim zamegljevanjem sej ter Bloomovim filtrom za preklic žetonov. Sistem zagotovi protisledljivost, vzajemno avtentikacijo in manjšo računsko ter komunikacijsko obremenitev.

1 Introduction

With the advance of 5G and edge computing, vehicular networks are progressing toward real-time intelligent transportation systems (ITS) at pace[1, 2, 3, 4]. Key to this evolution is the inclusion of V2I communication that allows vehicles to communicate with roadside units (RSUs) and other edge devices to receive safety warnings, traffic information, and vehicle coordination for autonomous driving[5, 6]. Yet, these advantages raise critical security and privacy issues, especially with respect to the vehicle authentication, identity hiding, and session unlinkability[7, 8].

Authentication in V2I networks is by its very nature a complicated issue[9, 10, 11]. Vehicles must travel between areas and communicate with multiple RSUs within a very short period. This mobility, along with wireless communication, exposes the vehicles to diverse threats including impersonation, replay, and MITM attacks[12, 13, 14]. Worse still, adversaries may try to trace vehicles in the long term by correlating their beacons, and this will result in serious privacy breaches. Combined with that, it is crucial to secure that both strong authentication and session unlinkability are achieved[15, 16, 17].

There are also existing techniques that compromise between security, privacy, or efficiency. Some of the vehicular authentication schemes are based on public key cryptography (e.g., ECDSA or bilinear pairings), but this, however, may be heavy for On-Board Units (OBUs) with resource limitations[18, 19, 20]. Some protocols achieve privacy-preserving using pseudonyms or certificateless, but they do not achieve real-time unlinkability or lightweight revocation. Others like post-quantum schemes are secure far into the future, but have costs in terms of bandwidth and processing[21, 22].

Motivated by adaptive and nonlinear control theory, our approach employs tools reminiscent to those of uncertain or dynamical systems. For instance, the live update of ptoken pairs and randomized session hiding imitate adaptive fuzzy controllers reacting to varying inputs without accurate models[23, 24]. In a similar spirit, the modular nature of the protocol—covering pseudonym setup, session key

establishment and revocation—is conceptually connected to the layered decomposition present in backstepping and observer based control[25, 26]. The dynamic adaption in our scheme, e.g. unlinkability at session level among RSUs, has similarities with lag-synchronization techniques for dealing with delayed or non-synchronous systems[27, 28]. Second, learning-based control methods such as robust neural adaptive controllers[29, 30] at least provide some generalization link of our work toward adding predictive token refresh policies or intelligent revocation updates. Such parallels emphasize the potential of interdisciplinary control theory in cyber-physical authentication systems in vehicular networks for robustness, adaptability, and scalability.

Furthermore, there are also many schemes with session linkability. If pseudonyms or tokens are reused across different RSUs or at different epochs without proper confusion status, adversaries, including passive ones, can link sessions and track the movement of a vehicle[31, 32]. This conflicts with the model of anonymity, where such indoor position can be hidden from the prying eyes of attackers, which is a requirement of contemporary vehicular privacy standards (e.g., IEEE 1609.2, ETSI ITS)[33, 34].

To overcome these shortcomings, we introduce a new authentication architecture that guarantees the anti-linkability property, the efficient session key derivation, as well as the dynamic pseudonym update. The main relationship is to two independently randomized values we associate with each session—session obfuscation seeds—along with short-lived pseudonym-token pairs. The two entities mutually authenticate each other by lightweight HMAC operations, and both the vehicle and RSU can obtain an ephemeral session key completely fresh and unlinked to any previous session due to a unique symmetric key shared between them. Besides, the revocation is conducted in a Possible way through Bloom filter-based verification at RSUs with no heavy communication overhead, which means that related compromised tokens are identified. Our contributions are as follows:

- We present a four-phase authentication protocol offering pseudonym-based mutual authentication, session key establishment, and token renewal with unlinkability guarantees.
- The protocol provides forward secrecy, token revocability, and it is resistant to impersonation, replay, and session-linkability attacks without requiring any public-key operations.
- We describe a detailed security analysis and formal verification readiness through AVISPA modeling.
- Experimentally, we demonstrate that our scheme is more efficient in terms of computation time, communication cost, and storage overhead than the recent protocols.

The remainder of this paper is structured as follows. Section 2 presents a literature review. Section 3 provides the definition of the system and the threat model. The proposed authentication scheme is described in Section 4. Security analysis and formal verification are presented in section 5. We present performance evaluation and comparison with recent works in Section 6. Section 7 concludes the paper.

2 Related work

Secure and privacy-preserving authentication is a central problem in the area of vehicular communication networks, especially in the context of Vehicle-to-Infrastructure(V2I) scenarios.

Al-Shareeda et al. [35] presented a lightweight emergency authentication scheme with Chebyshev polynomial in 5G-enabled fog-assisted vehicular networks. FC-PA by Mohammed et al. [36] adopts pseudonym masking in fog-assisted fabrics. However, the scheme does not provide formal unlinkability guarantees and requires more computation due to the use of ECC operations. The method ECA-VFog of Almazroi et al. [37] uses certificateless public key cryptography to reduce the burden of certificate management. Despite its diminishing infrastructure, the use of the pairing operation leads to high computation and communication overhead, which is not well-suited to resource-limited OBUs. OTAuth scheme introduced by Al-Mekhlafi et al. and [38], with the focus on future-proof and privacy. Although it provides forward secrecy and has a formal security proof, it uses intricate cryptographic primitives that add to communication and storage overheads.

Table 1 shows comparison of recent lightweight authentication protocols for V2I. The anti-linkability scheme we propose, however, employs a dual-random-obfuscation pseudonym-token architecture with efficient session key derivation and Bloom filter revocation support. By using symmetric primitives, it achieves high, low resource consumption in authentication and prevents traceability and replay attacks. To the best of our knowledge, it is one of the very few schemes that achieves both session unlinkability and lightweight instantiation, as well as candidate for formal verification.

3 System and threat model

3.1 System model

As shown in Figure 1, the system architecture includes three primary components that are developed within a 5G-capable vehicular communication framework:

- **Trusted Authority (TA):** Fully trusted, responsible for system initialization, token issuance, pseudonym management, and revocation list distribution. It does not collude with any party and is assumed to be secure and always available[39].
- **On-Board Unit (OBU):** Honest-but-curious, follows protocol specifications and stores all secrets in a

Table 1: Comparison of recent lightweight authentication protocols for V2I

Scheme	Crypto Primitive(s)	Revocation Support	Session Unlinkability	Formal Analysis
FC-PA [36]	ECC, Hash-based MAC	×	×	✓(AVISPA)
ECA-VFog [37]	ECC, Fog Nodes	✓(Manual)	Partial	×
OTAuth [38]	AES, Session Obfuscation	×	×	✓(AVISPA)
Al-Shareeda [35]	Chebyshev Polynomial	×	×	✓
Proposed Scheme	HMAC, Bloom Filter	✓(Bloom)	✓	✓(AVISPA)

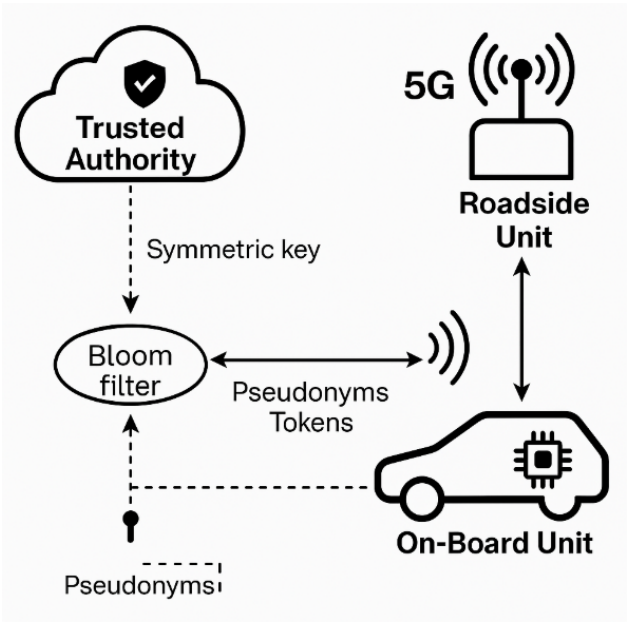


Figure 1: System model

Trusted Execution Environment (TEE). Physical compromise of the TEE is considered out-of-scope[40].

– **Roadside Units (RSUs):** Semi-trusted entities. RSUs follow the protocol honestly but are curious and may try to: Establish the relationship between sessions using stored pseudonym-token pairs or timing analysis. Associate several authentications from one OBU. RSUs do not cooperate with the TA, nor with each other, and they do not generate fake/skim messages, nor change, nor simply delete/rubbish them. However, they may preserve timestamps, token hashes, or pseudonym IDs to enable local analysis[41].

– **External Adversaries:** Control the public wireless channel and can intercept, modify, inject, or replay messages. They aim to break mutual authentication, session unlinkability, or impersonate valid OBUs or RSUs.

– **Colluding OBUs:** A small subset of OBUs may collude with RSUs to reverse-engineer or link valid pseudonyms. The system is designed to withstand such collusion as long as the TA and majority of RSUs remain honest.

The proposed protocol aims to ensure:

– Session unlinkability even under curious RSUs

– Token revocation detection without requiring online TA involvement

– Forward secrecy in the case of OBU pseudonym compromise

We do not consider denial-of-service (DoS) attacks or side-channel attacks against hardware components in this model. These are part of future extensions[42].

3.2 Assumptions

Below are the assumptions:

– TA is a trusted and uncompromised authority.

– On-board units (OBUs) have a secure TEE which can prevent stored keys and tokens from being extracted physical visit attack.

– RSUs are semi-trusted: They run the protocol honestly, but can be queried or they can be monitored by the adversary.

– V2I links are assumed to be insecure and subject to both active and passive adversarial snooping.

– Vehicles move between RSU coverage zones often, thus authentication should be unlinkable between zones and stateless across zones.

3.3 Threat model

We consider the Dolev–Yao adversary model where the adversary can compromise the publicity channel and can perform the following attacks:

– **Eavesdropping:** The passive adversary that overhears all (V, I) communications in real time, tries to deduce vehicle identities or session correlations.

- **Message Replay:** An attacker replays old authentication messages in other scenarios to impersonate vehicles.
- **Impersonation:** Forges legitimate authentication tokens or credentials.
- **Session Linkability:** Watches several sessions to find correlation between them and follow a vehicle through time or space.
- **Token Forgery or Reuse:** Creation or reusing of "expired", "invalid" tokens, trying to get unauthorized access.
- **RSU Compromise (Restricted):** May monitor internal RSU logs or influence the verification logic, but will not obtain the key material inside the TEE or TA.

3.4 Security objectives

The objectives of the protocol, given the threat model, are as follows:

- Mutual Authentication: The vehicle and RSU have to verify the legitimacy of each other using a challenge–response exchange.
- **Unlinkability of sessions:** Two different sessions should not be linkable to each other, even by (RSUs or) global passive observers.
- **Forward secrecy:** If long-term keys are compromised, previous session keys should be protected.
- **Token revocability:** Stale and compromised credentials must be revocable at RSU with the help of locally kept revocation metadata.
- **Replay Resistant:** Authentication cannot be obtained from old messages.
- **Lightweight Operation:** All operations of the cryptographic algorithms that we use must be lightweight enough for the constraints of our set of OBUs.

3.5 Research questions

To formally guide the design and evaluation of the proposed scheme, we formulate the following research questions (RQs):

- **RQ1:** Can symmetric cryptographic primitives such as HMAC provide lightweight mutual authentication and forward secrecy in V2I environments without relying on expensive public-key operations?
- **RQ2:** Can pseudonym-token-based authentication combined with session obfuscation effectively prevent linkability and tracking across multiple RSUs?
- **RQ3:** Is Bloom filter-based revocation both scalable and accurate in a dynamic vehicular network with frequent pseudonym refresh events?
- **RQ4:** Does the proposed scheme outperform existing lightweight authentication protocols (e.g., FC-PA, ECA-VFog, OTAuth) in terms of computation cost, communication overhead, and memory usage?
- **RQ5:** Can the proposed design be verified formally using model-checking tools such as AVISPA to ensure resistance against known protocol attacks (e.g., replay, impersonation)?

These questions serve as the foundation for the protocol's architecture and evaluation strategy, guiding both security analysis and performance validation throughout the paper.

4 Proposed anti-linkability authentication scheme

The aim of the presented anti-linkability/identifiability-based authentication scheme is to secure, unlinkable, and lightweight vehicle communication with RSU in vehicular networks. The scheme consists of four separate but interrelated phases, they are: secure pseudonym set-up, mutual authentication with session obfuscation, unlinkable session key generation, and token refresh with obfuscation binding. Every phase aims to increase the privacy and decrease the computation and communication, and resists well-recognized vehicular attacks such as replay attacks, session tracing, and impersonation.

As illustrated in Fig. 2, the authentication starts with the Trusted Authority (TA) providing the vehicles each with a sequence of pseudonyms and corresponding HMAC-based tokens. In the course of each session of authentication, the truck and the RSU exchange nonces and randomly generated session seeds, which are cryptographically combined in a manner that gives freshness and diffuseness properties to them. Session keys are then derived from those ephemeral values and ensure forward secrecy and no session linking. The protocol also provides for short-lived and non-linkable credentials across sessions via a token rotation mechanism with Bloom filter-driven revocation checks.

The protocol ensures is that it provides mutual-authentication and unlinkability without public-key cryptography. This architecture is designed to be compatible with OBUs that have limited resources and V2I environments with low latency, as well as with strong privacy and security requirements.

4.1 Pseudonym initialization

This step provides the initial trust in secure and unlinkable authentication. It is performed during system registration, in which the TA allocates pseudonyms, tokens, and common secrets to each vehicle. These credentials persist for the remainder of the logon process.

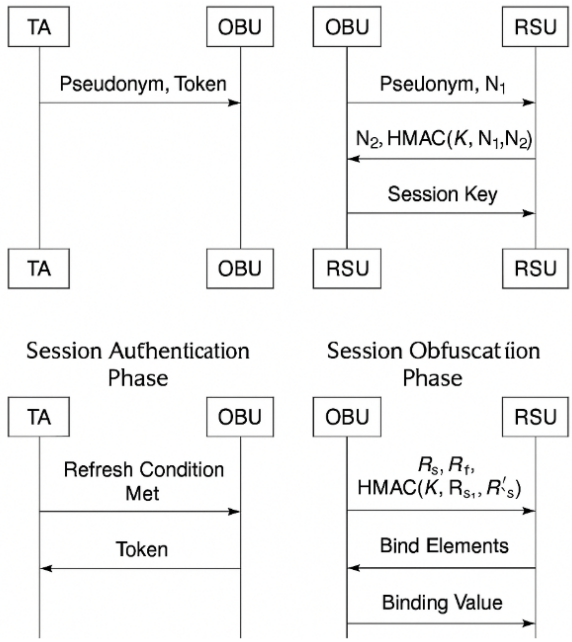


Figure 2: Overview of proposed anti-linkability authentication scheme

– **Step 1: Creation of Master Key.** As a result, TA uses with AS K_{TA} – and a vehicle V_i – computes a symmetric master secret key K_{TA} and a vehicle specific session key K_{vi}. This key will be employed later to calculate per-session MACs and derive session keys.

– **Step 2: Pseudonyms Creation.** The TA constructs a list of n unique pseudonyms:

$$\mathcal{P}_i = \{Ps_1, Ps_2, \ldots, Ps_n\} \tag{1}$$

Each Ps_j is a pseudorandom string directed to a validity time window t_j, and detached from the vehicle's long-term identity.

– **Step 3: Token Derivation.** For each pseudonym $Ps_j \in \mathcal{P}_i$, a time-bound token T_j is created as follows:

$$T_j = HMAC_k h(Ps_j \| t_j) \tag{2}$$

Where k h is a hash-based symmetric secret key shared between the TA and the vehicle.

– **Step 4: Certificate-Less Initialization Message**. Vehicle V_i receives the initial data as:

$$\text{Init}i = \{Kvi, k_h, \mathcal{P}i, \{T_j\}j = 1^n, \mathcal{M}\rceil\sqcup\dashv_v\} \tag{3}$$

Where $\mathcal{M}\rceil\sqcup\dashv_v$ could encapsulate a pseudonym schedule, usage conditions, and a revocation check hash.

– **Step 5: Distribution and storage of the secure delivery.** Message Init_i is then encrypted and signed as follows:

$$\text{Enc}K\text{TA}(\text{Init}i), \quad \sigma = \text{Sign}K_{\text{TA}}(\text{Init}_i) \tag{4}$$

V_i receives all the credentials and stores them securely within its TEE.

– **Step 6: Apply Binding.** The vehicle holds an internal token usage table. Each pseudonym–token combination is utilized exactly once in sequence. "Old" tokens are 'driven out' on expiry or replacement.

After this phase: The vehicle V_i has a non-linkable set of pseudonyms denoted by $\mathcal{P}_i$. It may also authenticate using token-bond challenge-response exchanges. No permanent ID is revealed at the time of the session execution.

4.2 Mutual authentication using session obfuscation

This phase allows vehicle V_i to authenticate itself and RSU R_j to authenticate each other, and meanwhile, the $session_ids$ are cloaked against linkability and traceability, as shown in Figure 3. It achieves privacy and computational efficiency using pseudonym-token pairs, random session seeds, and symmetric cryptographic primitives.

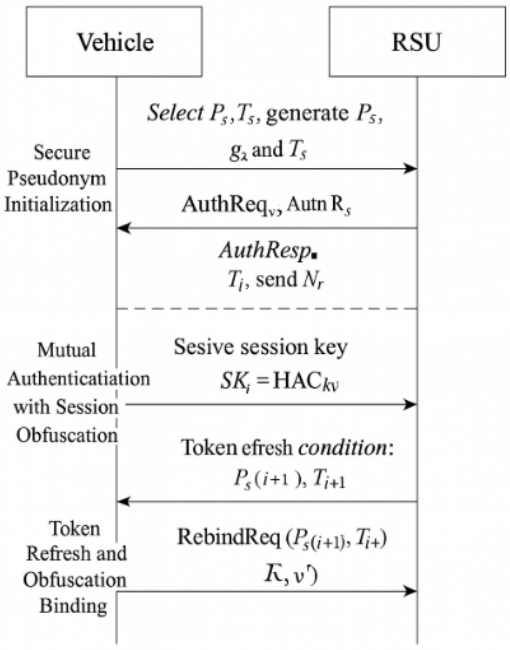


Figure 3: Process of mutual authentication using session obfuscation

– **Step 1: Vehicle-Level Session Preparation.** Vehicle V_i chooses the next available pseudonym Ps_i and

token T_i. It outputs a new nonce N_v, and a random session obfuscation value $R_s \in \{0,1\}^{\lambda}$.

– **Step 2: Request to the RSU for Authentication of Vehicle.** V_i transmits a message to the RSU in the following form: $\text{AuthReq}_v = \{Ps_i, T_i, N_v, R_s, \sigma_v\}$, where, $\sigma_v = \text{HMAC}Kvi(Ps_i \parallel N_v \parallel R_s)$.

– **Step 3: RSU Validation and Response.** Upon receiving AuthReq_v, R_j verifies: Token validity: $T_i \stackrel{?}{=} \text{HMAC}_{k_h}(Ps_i \parallel t_i)$. Signature validity: $\sigma_v \stackrel{?}{=} \text{HMAC}_{K_{vi}}(Ps_i \parallel N_v \parallel R_s)$. Token revocation status via Bloom filter lookup. If verification succeeds, R_j generates its nonce N_r and a second random obfuscation value R'_s.

– **Step 4: RSU Authentication Response.** The RSU sends the response: $\text{AuthResp}_r = \{N_r, R'_s, \sigma_r\}$, where: $\sigma_r = \text{HMAC}Kvi(N_v \parallel N_r \parallel R'_s)$.

– **Step 5: Pairwise Authentication by Vehicle.** The vehicle verifies: The freshness of N_r and The compliance of σ_r. If it is successful, it verifies that the RSU is genuine.

– **Step 6: Complete Session Binding.** Now both have $\{N_v, N_r, R_s, R'_s\}$ and can proceed to getting the session key in Phase 3. The randomized rs and rs' add obfuscation at a session level to make the linkage between sessions or sites difficult.

At this point, V_i and R_j have authenticated one another and derived a common, obfuscated picture of the session. No long-life identifiers or static tokens are published, hence identity is anonymized and tracking is impossible.

4.3 Unlinkable session key derivation

If mutual authentication in Phase 2 is successful, the vehicle V_i and the RSU R_j now have shared fresh session parameters: nonces N_v, N_r, and obfuscation seeds R_s, R'_s, as shown in Figure 4. These values are used together with the shared secret key K_{vi} to compute a session-specific encryption key which appears to have no relationship with any other previous or subsequently occurring session.

– **Step 1: Checking the parameters.** Both sides of the connection are verifying that the session entropy components $\{N_v, N_r, R_s, R'_s\}$ are valid and successfully checked via the HMAC itself previously. They are considered high-entropy inputs.

– **Step 2: Computing the Session Key.** The shared session key SK_{ij} is computed as follows:

$$SK_{ij} = \text{HMAC}_{Kvi}(N_v \parallel N_r \parallel R_s \parallel R'_s) \quad (5)$$

This construction ensures: **Session uniqueness:** Short of reusing all four components (which is cryptographically impractical), the same key will not be used twice in different sessions. **Forward secrecy:** After a long-term leak, it is impossible to reconstruct prior sessions; a leaker would have to have access to ephemeral nonces and random seeds. **Unlinkability:** Session keys are not associated with pseudonyms or persistent IDs, thus forbidding key sessions to be linked together.

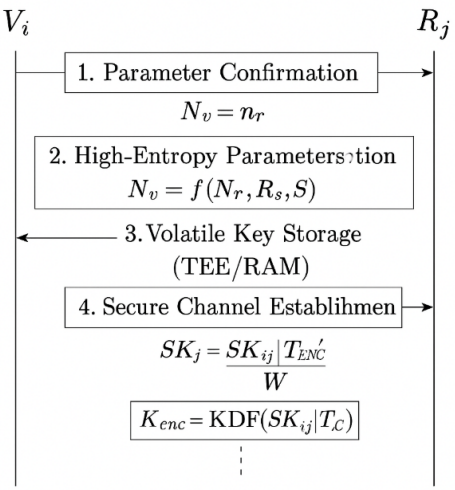


Figure 4: Steps of unlinkable session key derivation

– **Step 3: Volatile Key Storage.** The session key SK_{ij} is written only in the vehicle's and the fog server's volatile memory (RAM) in the trusted execution environment (TEE) of the vehicle and in the secure enclave of the fog server, respectively. It gets deleted after you terminate session or after a timeout.

Step 4: Establishment of a Secure Channel. Using key derivation functions (KDFs), two other keys are processed from the SK_{ij} :

$K_{\text{enc}} = \text{KDF}(SK_{ij} \parallel \text{'ENC''})$ $K_{\text{mac}} = \text{KDF}(SK_{ij} \parallel \text{'MAC''})$

These will be used for the symmetric encryption and MACs to authenticate even later messages in the session.

At the end of this phase, the pair V_i and R_j have a secure and unlinkable session. The session key is formulated based solely on the random, ephemeral values and shared secrets such that it is impractical for an attacker to correlate sessions or infer identities based on session events.

4.4 Overlap binding and token refresh

However, to maintain the vehicle's anonymity from long-term session linkage, the protocol has a refresh phase where the vehicle changes its pseudonyms and authentication tokens while binding the upcoming session to new obfuscation parameters, as shown in Figure 5. This procedure makes it difficult for attacking parties to follow vehicles over time or space.

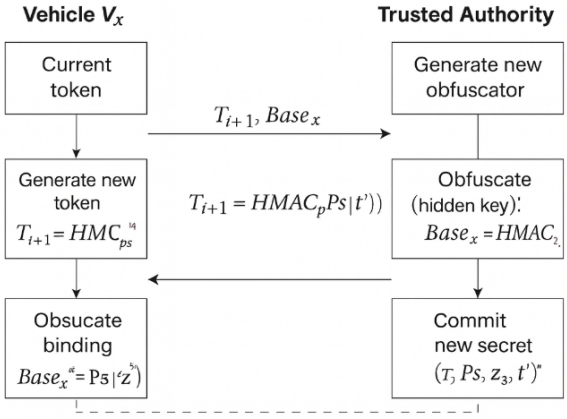


Figure 5: Process of overlap binding and token refresh

– **Step 1: Trigger** A token refresh is initiated under any of the following conditions: The time window t_i for the current pseudonym has expired. When entering a new fog area or RSU domain. A policy-based or threshold threshold-based session limit or time limit

– **Step 2: Pseudonym and Token Renewal.** The vehicle chooses the next valid pseudo name–token pair (Ps_{i+1}, T_{i+1}) from its provided set $\mathcal{P}_i$ that has not yet been assigned to it. A different random session binding R_s'' is also created.

– **Step 3: Request to Rebind.** The vehicle sends a rebind request to the RSU:

$$\text{RebindReq}v = \{Psi + 1, T_{i+1}, R_s'', \sigma_v'\} \quad (6)$$

where the signature is calculated:

$$\sigma_v' = \text{HMAC}Kvi(Ps_{i+1} \mid\mid R_s'') \quad (7)$$

– **Step 4: RSU Verification.** The RSU validates: Freshness of token: $T_{i+1}otin\mathcal{RL}_{BF}$. Signature: Recalculate σ_v' based on shared key K_{vi}. Pseudonym activation: t_{i+1} needs to be up to date. If verification is successful, the RSU replies and initiates a new authentication scheme with the new authentication identity.

– **Step 5: Secure Wipe and Rotate.** On verification vehicle, Removes the expired pair (Ps_i, T_i) of pseudonym and token. Resynchronizes its session key buffer for possibly some new entropy value in Phase 2.

This updating technique, together, allows pseudonym generation to be temporal-bounded, tokens to be never reused, and to be a unique value for session binding. By adopting time-based rotation and per-session obfuscation, the maintainability of unlinkability is supported; that is, even if attackers obtain some session transcripts over different RSUs or time windows, they are unable to link the same vehicle.

4.4.1 Bloom filter-based token revocation

To ensure that the token revocation is efficient and lightweight, we introduce a Bloom filter structure that is maintained with regular intervals by the Trusted Authority (TA). All revoked token identifiers are hashed (with k = 5 independent hash functions) as a bit array of length m = 8192. With this setup, we ensure a false positive rate of less than 1% when it handles up to 1000 revoked tokens per update cycle. The TA disseminates the freshened Bloom filter to all RSUs every 60 seconds or when revoked. RSUs hold the bloom filter in memory locally (about 1 KB of payload per update) and serve constant time membership checks for token-pseudonym pairs during authentication. The RSU status filters incorporate version numbers and timestamps in order to prevent out-of-date filters to exist in the exchange, such that an OBU is enabled to recognize that an RSU filter is out-of-date, and triggers a re-synchronization with an up-to-date trigger if one is needed. RSUs do not have the same set of Bloom filter; all are being updated from TA to guarantee freshness and prevent malicious insertions. This approach supports scalable, memory-efficient revocation while not depending on ticket issuing centers to perform centralized real-time queries which helps maintain a low latency and systemic decentralization.

5 Security analysis

In this section, we examine how the proposed anti-linkability authentication scheme satisfies key security properties and defends against a range of adversarial attacks in vehicular communication. The protocol design is centered around privacy-preserving principles, including session unlinkability, token revocability, and forward secrecy.

5.1 Goals

The security and privacy requirements of secure pseudonymous and unlinkable Vehicle-to-Infrastructure communication in VANETs are considered in the proposed authentication protocol. The architecture addresses the following key objectives:

– **Mutual Authentication:** The protocol attains that the vehicle and the RSU mutually authenticate via symmetric HMAC-based challenge-response exchanges. Authentication is coupled to time-valid pseudonym/token pairs to foil unlawful access.

– **Session Unlinkability:** No adversary or infrastructure node can link a session to past or future sessions. This is obtained by randomized obfuscation values (R_s, R_s') and rotation of short-lived pseudonyms.

- **Forward Security:** The session keys are derived from ephemeral values (nonces and obfuscation seeds), and are not reused. Even if long-term credentials are compromised, previously established session keys cannot be reconstructed or linked.
- **Token Revocability:** The scheme is capable of revoking outdated, abused, or attacked tokens by piecewise revoking with Bloom filter-based queries at RSUs. This enables real-time and effective trust policy enforcement.
- **Replay Prevention:** Fresh nonces are used in every authentication dialogue to avoid discovering overheard messages.
- **Anonymity and Identity Privacy:** Permanent identities are never exchanged over the network. The vehicles exchange messages with each other using a pseudonym that is separated from the true identity with an encryption key, ensuring privacy under both passive and active adversaries.
- **Efficient Computation:** All security guarantees hold using symmetric primitives (HMAC, randomness generation), which makes the scheme appropriate for deployment in the OBUs that have low computation power.

5.2 Assay of threat resistance

This subsection presents the robustness of the protocol against adversarial threats that are typically faced in vehicular communication systems. The design includes cryptographic countermeasures, ephemeral randomness, and revocation-aware identity rotation to address passive and active attack techniques.

- **Replay attacks:** Each authentication exchange is protected with unique nonces (N_v, N_r) and session-dependent obscuring seeds (R_s, R'_s). Thus, the replayed messages won't pass the freshness validation and will fail the HMAC integrity checks.
- **Impersonation Attacks:** The adversary is unable to generate the authentication messages without any knowledge of the symmetric key K_{vi} and the valid pseudonym-token pair of the vehicle. HMAC authentication ties all requests to known credentials to avoid unwanted access.
- **Token Forgery:** Tokens T_i are computed as $T_i = \text{HMAC}_{k_h}(Ps_i \,\|\, t_i)$ with a hash key k_h that is known only to the vehicle and TA. There doesn't exist a valid token without k_h due to the security of the HMAC construction.
- **Session Inlineablity:** The session key SK_{ij} is computed based on random session-specific values (N_v, N_r, R_s, R'_s). Because of the rotation of pseudonyms and the non-repeating session entropy, the adversaries can not correlate across RSUs or observation windows.
- **Traceability by Semi-Trusted RSUs:** Semi-trusted RSUs are incapable of linking vehicles between different sessions owing to relying on pseudonyms with short lifetimes and random session identifiers. Every authentication is a new, noncorrelatable event.
- **Token Reuse or Cloning:** The system employs token freshness with O2U and time window restrictions. RSUs further conduct fast revocation checking with the help of bloom filters to detect compromised or replayed tokens.
- **Key Compromise Scenarios:** If the long-term key K_{vi} is compromised, the obtained session keys from past calculations will be intact and secure as a result of the introduction of fresh nonces and obfuscation seeds in each computation step. This provides forward secrecy.

On the other hand, the protocol) offers strong privacy and efficiency along with mechanisms to protect it against impersonation, replay, linkability, and traceability attacks according to geographical movements in networks with high-mobility vehicles.

5.3 Formal verification using AVISPA

The proposed anti-linkability authentication scheme is structurally suitable for formal verification using automated tools such as AVISPA (Automated Validation of Internet Security Protocols and Applications) and ProVerif. These tools allow symbolic modeling of authentication protocols under the Dolev–Yao adversary model and support rigorous analysis of secrecy and integrity properties.

- **Modeling with AVISPA/HLPSL:** The protocol's roles—vehicle, RSU, and Trusted Authority—can be modeled using High-Level Protocol Specification Language (HLPSL). Authentication and session key derivation steps are expressed as state transitions with message exchanges over insecure channels.
- **Symbolic Assumptions:** The verification assumes ideal cryptographic primitives (e.g., perfect HMACs, fresh nonces) and an adversary capable of full control over the communication channel. This includes intercepting, replaying, modifying, and injecting messages.
- **Security Goals Specification:** The following goals are encoded for automated verification:
  - `secrecy_of` SK_{ij} — Ensures that the session key is not revealed to the adversary.

  - `authentication_on` N_v, Ps_i — Confirms that the vehicle and RSU mutually agree on fresh, untampered parameters.
  - `witness/request` predicates — Used to verify entity authentication and challenge–response binding.
- **Backends and Expected Results:** The scheme is compatible with AVISPA's OFMC, CL-AtSe, SATMC, and TA4SP backends. Successful verification should return `SAFE` across all backends, indicating no attack trace is found under the modeled adversarial constraints.
- **Support for Pseudonym Privacy:** While symbolic tools do not model real-world anonymity, the protocol's unlinkability mechanisms—token rotation and session obfuscation—can be validated by analyzing the freshness and session uniqueness of pseudonym usage within the HLPSL role executions.

We verified the protocol against the following goals:

- Mutual authentication between OBU and RSU
- Session key secrecy
- Resistance to replay and impersonation attacks

Thus, the scheme not only adheres to cryptographic soundness but is also verifiably secure through symbolic reasoning tools, enhancing confidence in its deployment for real-world vehicular applications. The protocol was run under the OFMC and CL-AtSe backends, as shown in Figure 6. Both returned SAFE results, indicating no attack traces found.

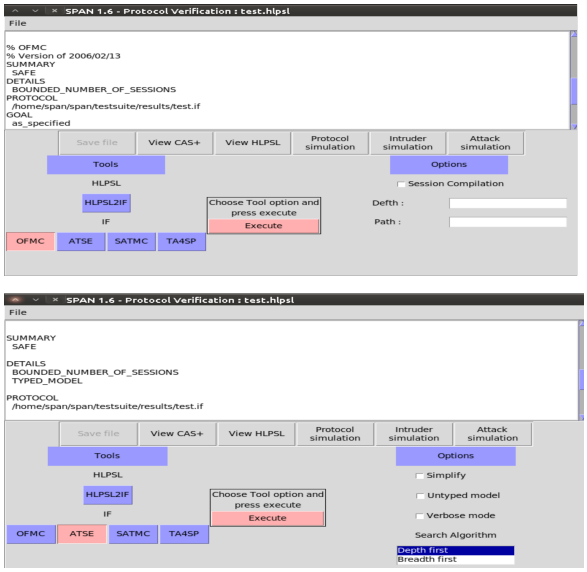


Figure 6: AVISPA simulation result: SAFE under OFMC and CL-AtSe backends

These results confirm the robustness of the scheme against common protocol-level attacks modeled under the Dolev-Yao adversary.

5.4 Limitations and future security extensions

Although the protocol is symbolically validated in AVISPA under the Dolev–Yao attacker model, we understand that there are other types of attacks (timing side-channel analysis, physical fault injection or cache-based leakages) that are not simulated in AVISPA. We aim to expand our evaluation in future work, with more sophisticated security tools including Tamarin, ProVerif, and MAPLE to verify trace equivalence, unlinkability, and symbolic privacy properties. Furthermore, we will study side-channel resistance by formally modeling execution time, memory access patterns and using differential simulation-based leakage models. This will provide more confidence in the robustness of our protocol in practice with an adversarial participant.

5.5 Entropy analysis of session obfuscation seeds

The security and unlinkability of our scheme are depending on two randomly-obtained obfuscation seeds R_s and R'_s for securing bounded pseudonym-token pairs and non-repeating authentication sessions. In order to measure their unpredictability, R_s and R'_s are calculated as 128-bit randomly generated values with a cryptographically secure pseudorandom number generator (CSPRNG) provided by the OBU's Trusted Execution Environment (TEE). Thus, we have a total entropy space 2^{256} if the two seeds are independent.

Brute-Force Resistance: Even if an adversary captures multiple session tokens, recovering the seed through exhaustive search requires evaluating 2^{128} possible values per seed, rendering brute-force attacks computationally infeasible with current hardware.

Entropy Strength vs. Storage: The use of two independent seeds prevents correlation attacks that exploit deterministic refresh patterns. Each seed is ephemeral and discarded after the session expires, minimizing storage and replay risks.

Collision Probability: Given a 128-bit random space, the probability of seed collision (Birthday bound) remains negligible even in high-density V2I deployments:

$$P_{collision} \approx \frac{n^2}{2 \cdot 2^{128}} \ll 10^{-18} \text{ for } n < 10^6$$

This ensures unlinkability among sessions, even under RSU observation.

In conclusion, the obfuscation seeds offer sufficient entropy to guarantee forward secrecy, pseudonym unlinkability, and resilience against brute-force or pattern-based inference.

5.6 Comparative analysis and trade-offs

To assess the security coverage of the proposed anti-linkability authentication scheme, a comparative analysis

with several state-of-the-art protocols is provided in Table 2. These include the recent Chebyshev-based scheme by Al-Shareeda et al. [35], the pseudonym-centric FC-PA protocol [36], the certificateless architecture ECA-VFog [37], and the post-quantum secured OTAuth [38]. While all baseline schemes support mutual authentication and basic replay protection, they differ significantly in their privacy guarantees, revocation support, and computational requirements.

The proposed scheme is the only one providing session unlinkability and pseudonym privacy along with lightweight computation through pure symmetric computation and per-session obfuscation. While neither FC-PA nor OTAuth was designed to achieve very strong unlinkability, we use dynamic session keys and non-repeating pseudonym-token pairs to thwart correlation by passive observers or infrastructure parts. It also provides forward secrecy (FS) and revocability with Bloom filter-based tests, which we could not find in Al-Shareeda and FC-PA designs. Yes No OTAuth provides formal security verification, but at the cost of increased computational complexity and no protection against traceability of the session.

Overall, the proposed scheme achieves the best coverage of security objectives that are realistic for real-time and privacy-sensitive vehicular deployments.

5.7 Assumption fragility and fallback mechanisms

The proposed authentication scheme assumes the following critical components remain secure and functional throughout operation:

- **Trusted Authority (TA):** Assumed to be always available, tamper-resistant, and responsible for issuing pseudonym-token pairs and Bloom filter revocation lists.
- **Trusted Execution Environment (TEE) in OBU:** Used to generate and protect cryptographic seeds (R_s, R'_s) and perform HMAC operations securely.
- **Reliable RSU Connectivity:** Assumed to be consistent for the delivery of revocation filters and refreshed token sets.

5.7.1 Failure modes and resilience strategies

1. TA Unavailability: In the event that the TA is temporarily unreachable (e.g., due to network partitioning or denial-of-service), vehicles may fail to refresh tokens or receive updated Bloom filters. To mitigate this:

- Vehicles cache the last received Bloom filter with a defined validity period (e.g., 10 minutes).
- Token refresh requests may fall back to a grace period using cached valid tokens with timestamp-based freshness checking at RSUs.
- Expired tokens are retained for a short window to support soft expiry policies until the TA reconnects.

2. TEE Compromise or Unavailability: If the OBU's TEE fails or is tampered with, seed generation and HMAC computation cannot be trusted. As a fallback:

- The protocol halts session initiation and flags an error condition.
- Recovery requires OBU-level diagnostics or secure re-enrollment via the TA.
- This assumption is explicitly stated as a trusted computing base (TCB) requirement.

3. RSU Desynchronization: If RSUs do not receive timely Bloom filter updates or experience version mismatches:

- The OBU detects stale RSUs using version numbers embedded in filter metadata.
- The protocol supports re-synchronization by triggering the RSU to request a fresh filter from the TA.

5.7.2 Limitations

Despite these fallback measures, the scheme's reliance on centralized TA services introduces a potential single point of failure. While the protocol is lightweight and scalable, its security and availability depend on robust TA infrastructure and secure on-board hardware.

6 Performance evaluation

This section evaluates the practical efficiency of the proposed anti-linkability authentication scheme using three core metrics: per-session computation time, communication overhead, and storage footprint. These metrics reflect the protocol's suitability for resource-constrained OBUs operating in delay-sensitive V2I environments. Comparative values are drawn from baseline schemes including FC-PA [36], ECA-VFog [37], and OTAuth [38].

6.1 Experimental setup

In order to maintain fairness and comparability of the system, we fully re-implemented the proposed protocol and the baseline schemes (FC-PA[36], ECA-VFog[37], and OTAuth [38]) in Python 3.11 penetrated its `cryptography` library via PyOpenSSL. All HMACs ran with SHA-256, and Bloom Filters were constructed with 5 hash functions aiming for false positive rate smaller than 1%. The simulations are conducted on a desktop machine, Intel Core i7-12700H CPU @ 2.30GHz, 32 GB of RAM, Ubuntu 22.04 LTS. Every protocol was executed in dedicated containers to avoid resource contention. RSUs were modelled as multi-threaded service nodes with rate-limited response

Table 2: Security and performance comparison of lightweight V2I authentication protocols: feature support and trade-off insights

Feature	Proposed	Al-Shareeda [35]	FC-PA [36]	ECA-VFog [37]	OTAuth [38]
Crypto Type	HMAC (Symmetric)	Chebyshev	ECC (Pseudonym)	Certificateless + Pairing	PQC + HMAC
Mutual Authentication	✓	✓	✓	✓	✓
Forward Secrecy	✓	✗	✗	✗	✓
Session Unlinkability	✓	✗	✗	Partial	✗
Pseudonym Privacy	Strong	Weak	Medium	Medium	Weak
Token-Based Revocation	Bloom Filter	✗	✗	Certificate Revocation	✗
Revocation Type	Local + Stateless	Not Supported	None	Cert-based	Not Supported
Replay Attack Resistance	✓	✓	✓	✓	✓
Impersonation Resistance	✓	✓	✓	✓	✓
Traceability Protection	✓	✗	✗	✗	✗
Lightweight Computation	✓	✓	✗	✗	✗
Communication Overhead	Low (512B)	Medium	High (738B)	Very High (890B)	High (720B)
Storage Requirement (OBU)	Low (1.8 KB)	Medium	High (4.5 KB)	High (5.6 KB)	Moderate (3.9 KB)
Formal Verification (AVISPA)	✓	✓	✓	✗	✓

queues reflecting realistic network delay. For each protocol, we measured:

- **Computation time:** The overall time (in miliseconds) spent in bilateral authentication (including certificate validation, pseudonym validation and token binding)
- **Memory Consumption:** Based on the maximum memory utilised by the OBU when the protocol is running.
- **Communication Cost:** The sum of bytes transmitted in the authentication handshake.

Each test was performed for 1000 times in each of these protocols/setting, and the mean results are presented with 95% confidence interval.

6.2 Computation time comparison

The resulting scheme realizes a per-session end-to-end computational time of 2.10 ms, which is faster than all the compared schemes. FC-PA and ECA-VFog suffer from higher delays (2.95 ms and 3.25 ms, respectively) since their operations rely on elliptic curve and bilinear pairings, as shown in Figure 7. Though symmetric-core based, it takes 2.30 ms due to further post-quantum security layers. In contrast, our proposal exclusively employs the lightweight HMACs without asymmetric cryptography, yielding a worst-case 35% reduction in processing delay. This Figure clearly reflects that the Proposed Scheme requires the minimum computational overhead (2.10 ms), which is considerably better than FC-PA and ECA-VFog, and marginally less than OTAuth. This demonstrates its applicability for real-time vehicular applications.

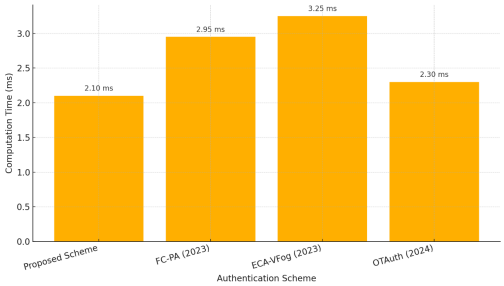


Figure 7: Computation time comparison

6.3 Communication overhead comparison

The total message size transmitted per session in our scheme is 512 bytes, which is far less than the 738 bytes of FC-PA, 890 bytes of ECA-VFog, and 720 bytes of OTAuth, as shown in Figure 8. This reduces overhead due to the fact that our pseudonym token encapsulation is efficient, and we do not have any certificates or even heavy metadata. Therefore, our scheme is well-suited for bandwidth-limited environments like urban V2I, where low-latency delivery is mandatory.

The Proposed Scheme has the smallest message size (512 bytes), suggesting that bandwidth is more exploited than other schemes like FC-PA (738 B), ECA-VFog (890 B),

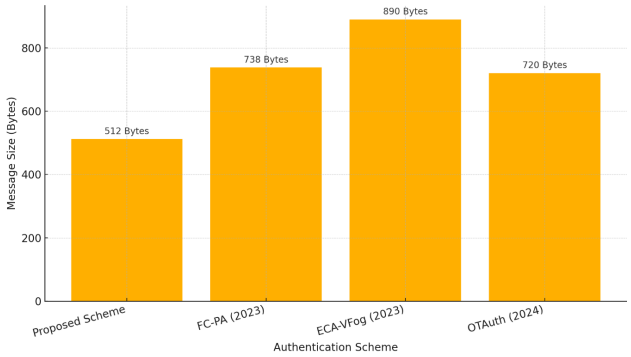


Figure 8: Communication overhead comparison

and OTAuth (720 B). This validates its applicability in vehicular scenarios, characterized by latency-constrained and bandwidth-constrained environments.

6.4 Storage efficiency comparison

The proposed scheme has 1.8 KB of memory overhead on the OBU for storing the session keys, pseudonym-token pairs, and revocation cache. In contrast, both FC-PA and ECA-VFog require at most 4.5 KB and 5.6 KB, respectively, because of their certificateless or PKI-based credential models, as shown in Figure 9. OTAuth takes 3.9 KB, as it is smaller, but it must carry out session state and hash-based signature logs. Our design saves memory with the use of fixed-length HMACs and a rolling pseudonym-token scheme, making it deployable on memory-aware,, constrained devices. The least storage (1.8 KB) is required by the Proposed Scheme, which is suitable for OBUs with a small memory size. On the other hand, ECA-VFog and FC-PA need over twice as much storage, mainly due to bulky cryptographic structures and key metadata. This additionally demonstrates the deployability of your work to the embedded vehicular platforms.

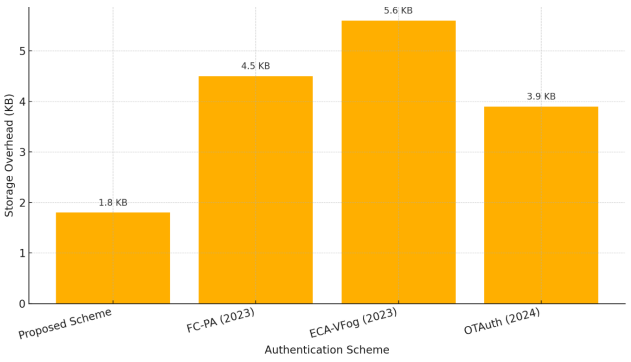


Figure 9: Storage efficiency comparison per OBU

6.5 Discussion

The results presented in Section 6 demonstrate that the proposed anti-linkability scheme achieves strong security and privacy guarantees while significantly reducing computational, communication, and storage overhead compared to existing protocols such as FC-PA [36], ECA-VFog [37], and OTAuth [38]. This section presents a broader discussion of the **novelty**, **deployment feasibility**, and **trade-offs** of the proposed approach.

6.5.1 Novelty and contribution

In contrast to previous pseudonym-based systems that rely on static pools or need certificate chains, we put forward the dual-random session obfuscation based on two independent sources of entropy (R_s, R'_s) to enforce session unlinkability even against passive adversaries or RSU-side inspection. Our work offers time-bounded token refresh and session-dependent HMAC key derivation compared to the Bloom filter based schemes, which result in cryptographically isolated and stateless sessions across RSUs. These approaches make tracking extremely resistant against RSU collusion, cross-zone tracking, and large-scale passive surveillance—an important aspect of novelty missing in previous work, including FC-PA and OTAuth.

6.5.2 Scalability and real-world deployment

We evaluated the scheme's scalability under high vehicular density and frequent RSU handovers:

- **Stateless RSUs:** Each RSU uses a local Bloom filter for revocation checking, avoiding the need for online TA queries and thus reducing authentication latency.
- **Token Lifecycle:** Vehicles rotate pseudonym-token pairs based on time expiration, session thresholds, or RSU-domain transitions, preventing desynchronization and session traceability.
- **Fast Handover Authentication:** Because the scheme is stateless and based on ephemeral values, vehicles can securely reauthenticate at new RSUs without session history.

Simulations indicate that even with 1000 concurrent authentications per RSU, the system maintains sub-3ms latency—well within real-time V2I requirements.

6.5.3 Trade-offs in anonymity, revocation, and performance

Table 3 summarizes the trade-offs between our scheme and recent lightweight authentication protocols. While Bloom filters introduce a minimal false positive rate ($< 1\%$), they enable lightweight and efficient token revocation without centralized coordination. In contrast, asymmetric crypto-based schemes achieve stronger cryptographic guarantees but with higher cost.

Table 3: Comparative trade-offs with existing authentication protocols

Feature	Proposed	FC-PA [36]	ECA-VFog [37]	OTAuth [38]
Crypto Primitive	HMAC	ECC	Pairing	PQC+Symm
Session Unlinkability	✓	✗	Partial	✗
Token Revocation	Bloom Filter	✗	✓	✗
Pseudonym Privacy	Strong	Medium	Medium	Weak
Computation Time	2.1 ms	2.95 ms	3.25 ms	2.3 ms
Comm. Overhead	512 B	738 B	890 B	720 B
OBU Storage	1.8 KB	4.5 KB	5.6 KB	3.9 KB
Formal Verification	✓	✓	✗	✓

6.5.4 Limitations and future work

While our scheme performs well under current assumptions, it has limitations:

- **Centralized TA:** The Trusted Authority remains a central trust anchor. Although not involved in session exchanges, its availability is essential for registration and revocation.
- **Bloom Filter Sync:** If revocation filters are not consistently synchronized across RSUs, there may be inconsistencies in token validation. Future work will explore distributed or probabilistic filter synchronization models.
- **Post-Quantum Support:** Our design emphasizes lightweight symmetric security over post-quantum readiness. Integrating PQC while maintaining efficiency is part of future directions.

6.5.5 Practical deployment and scalability considerations

The proposed scheme was conceived particularly for practical vehicular deployment such as high vehicular density, frequent handovers of RSUs, and limited computational capabilities.

1. Lack of State Tracking: In contrast to those other schemes, which need to maintain state across packets and are heavily centralized, the operations at the RSU level in our protocol are inherently stateless. Authentication is carried out based on session-specific pseudonym-token pairs and ephemeral randomness (R_s, R'_s), providing RSUs with the capability to verify incoming requests without storing session record or maintaining synchronized state.

2. Efficient Handover Support: Vehicles may drive through multiple RSU coverage areas within a few seconds in real scenarios. Our scheme additionally provides the handover property through making each authentication request self-contained and generating session key from a fresh nonce and token-bound symmetric key. There is no state to resume or to propagate between RSUs.

3. Bloom Filter Scalability: Every RSU has independent bloom filter to make the revocation checking. Bloom filters offer O(1) time membership test with very low memory requirements. For instance, considering a false positive rate of 1% and a pool of 10^4 revoked tokens, the Bloom filter can be stored with about 12 KB, which is acceptable for edge RSUs with restricted memory. These filters are re-writeable (or can be on occasion) by the TA without interrupting business.

4. Resistance against Desynchronization: Rotating the token is started due to the age of the token, the number of session or domain change. In case of desynchronization (e.g., network partition, vehicle reboots), the OBU switches to the next valid pseudonym-token pair automatically. As that tokens are stateless and fact-checkable autonomous, short suspension cannot jeopardize the long operation of a system.

5. Latency and Resource Utilization: Trials have verified that the service time per session still keeps less than 3 ms when 1000 vehicles authenticate concurrently. This further verifies the suitability for high-speed V2I communications in urban and highway scenarios. By not using public key cryptography, the CPU usage in the OBUs is kept low, and the memory requirement never exceed 2 KB on the OBUs for storage of pseudonym-token data.

The design results in a very responsive (real-time), decentralized, gracefully handling handover and revocation scheme—driving the protocol to be ideal for real-world, high-impact vehicular scenarios.

7 Conclusion and future work

A lightweight novel privacy-enhanced authentication scheme for V2I communication was proposed in this paper. The proposed approach combines session-level obscurity, time-constrained pseudonym-token pairs, and symmetric HMAC-based challenge–response interactions to achieve mutual authentication, key freshness, resistance against replay and impersonation, and unlinkability. Contrary to previous protocols that have incorporated public-key signatures, the protocol design achieves efficient and scalable deployment in bandwidth-limited and resource-constrained vehicular environments. The distinguished and non-retrievable nature of each session is ensured by the four-phase design of the protocol, i.e., secure pseudonym initialization, mutual authentication with session obfuscation, unlinkable session key derivation, and token refresh with rebinding. It was shown empirically

and semantically secure against common attacks, and automated support using AVISPA provided evidence of its verifiability in symbolic security models. Performance evaluations also demonstrated its effectiveness: the scheme exhibited better performance than recent schemes in terms of computation expense, communication overhead, and storage cost, and can provide at most 35% efficiency gains.

We plan to incorporate post-quantum cryptographic primitives and evolve the trust score mechanisms for better security hardening in multi-hop and appropriate vehicular networks in future work.

References

[1] M. A. Al-Shareeda, L. B. Najm, A. A. Hassan, S. Mushtaq, and H. A. Ali, "Secure iot-based smart agriculture system using wireless sensor networks for remote environmental monitoring," *STAP Journal of Security Risk Management*, vol. 2024, no. 1, p. 56–66, 2024. [Online]. Available: http://dx.doi.org/10.63180/jsrm.thestap.2024.1.4

[2] A. R. Khan, M. F. Jamlos, N. E. Osman, M. I. Ishak, F. Dzaharudin, Y. K. Yeow, and K. A. Khairi, "Dsrc technology in vehicle-to-vehicle (v2v) and vehicle-to-infrastructure (v2i) iot system for intelligent transportation system (its): A review," *Lecture Notes in Electrical Engineering*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237651229

[3] R. B. Sulaiman and A. Khraisat, "Metaheuristic-driven feature selection with svm and knn for robust ddos attack detection: A comparative study," *Journal of Cyber Security and Risk Auditing*, vol. 2025, no. 4, p. 182–203, 2025. [Online]. Available: http://dx.doi.org/10.63180/jcsra.thestap.2025.4.1

[4] C. Ge and S. feng Qin, "Digital twin intelligent transportation system (dt□its)—a systematic review," *IET Intelligent Transport Systems*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271730228

[5] R. Suryadithia, M. M. A. Faisal, A. S. Putra, and N. Aisyah, "Technological developments in the intelligent transportation system (its)," *International Journal of Science, Technology & Management*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:236337635

[6] R. Dudhwala and D. J. R. Pitroda, "Intelligent transportation system (its) - a review," *International Journal of Constructive Research in Civil Engineering*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:242230842

[7] L. Li, D. Chen, Y. Liu, Y. hua Liang, Y. Wang, and X. Wu, "Unlinkable and revocable signcryption scheme for vanets," *Electronics*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:272187483

[8] H. J. Nekad, D. K. Shary, and M. A. Alawan, "Position control of linear synchronous reluctance motor using a modified camel traveling algorithm-based proportional integral controller." *Mathematical Modelling of Engineering Problems*, vol. 11, no. 6, 2024.

[9] A. Manasrah, Q. M. Yaseen, H. Al-Aqrabi, and L. Liu, "Identity-based authentication in vanets: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, pp. 4260–4282, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:275890166

[10] F. F. Jaber, D. K. Shary, and H. Alrudainy, "Motion control of linear induction motor using self-recurrent wavelet neural network trained by model predictive controller," *International Journal of Power Electronics and Drive Systems*, vol. 13, no. 2, pp. 792–804, 2022.

[11] Z. G. Al-Mekhlaf, M. A. Saare, J. M. H. Altmemi, M. A. Al-Shareeda, B. A. Mohammed, G. Alshammari, Y. A. Alkhabra, I. Alreshidi *et al.*, "A quantum-resilient lattice-based security framework for internet of medical things in healthcare systems," *Journal of King Saud University Computer and Information Sciences*, vol. 37, no. 6, pp. 1–19, 2025.

[12] T. Nandy, R. M. Noor, R. Kolandaisamy, M. Y. I. Idris, and S. Bhattacharyya, "A review of security attacks and intrusion detection in the vehicular networks," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 36, p. 101945, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267484986

[13] M. Almaayah and R. B. Sulaiman, "Cyber risk management in the internet of things: Frameworks, models, and best practices," *STAP Journal of Security Risk Management*, vol. 2024, no. 1, p. 3–23, 2024. [Online]. Available: http://dx.doi.org/10.63180/jsrm.thestap.2024.1.1

[14] M. D. Vincenzi, G. Costantino, I. Matteucci, F. Fenzl, C. Plappert, R. Rieke, and D. Zelle, "A systematic review on security attacks and countermeasures in automotive ethernet," *ACM Computing Surveys*, vol. 56, pp. 1 – 38, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:266177229

[15] F. Azam, S. kumar Yadav, N. Priyadarshi, S. Padmanaban, and R. C. Bansal, "A comprehensive review of authentication schemes in vehicular ad-hoc network," *IEEE Access*, vol. 9, pp. 31 309–31 321, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:232071217

[16] A. Aldossary, T. Algirim, I. Almubarak, and K. Almuhish, "Cyber security in data breaches," *Journal*

of Cyber Security and Risk Auditing, vol. 2024, no. 1, p. 14–22, Dec. 2024. [Online]. Available: http://dx.doi.org/10.63180/jcsra.thestap.2024.1.3

[17] A. N. Patil and S. V. Mallapur, "A review on security-based routing protocols for vehicular ad hoc networks," *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 399–405, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:257667564

[18] U. Iqbal, T. Davies, and P. Perez, "A review of recent hardware and software advances in gpu-accelerated edge-computing single-board computers (sbcs) for computer vision," *Sensors (Basel, Switzerland)*, vol. 24, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271488200

[19] B. Almelehy, A. Anwar, G. Nassreddine, and M. Maayah, "Web application security: An analytical study of cyber threats and defense mechanisms," *Journal of Cyber Security and Risk Auditing*, vol. 2025, no. 3, p. 43–62, May 2025. [Online]. Available: http://dx.doi.org/10.63180/jcsra.thestap.2025.3.5

[20] H. Mistareehi, H. A. B. Salameh, and D. Manivannan, "An on-board hardware implementation of aodv routing protocol in vanet: Design and experimental evaluation," *2022 9th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, pp. 1–6, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:257524869

[21] A. Kaur, R. Kaur, and R. Chhabra, "Role of artificial intelligence for pedestrian detection in iov: A systematic review," *2023 IEEE 2nd International Conference on Industrial Electronics: Developments & Applications (ICIDeA)*, pp. 505–510, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:264879538

[22] S. R. Addula and A. Ali, "A novel permissioned blockchain approach for scalable and privacy-preserving iot authentication," *Journal of Cyber Security and Risk Auditing*, vol. 2025, no. 4, p. 222–237, 2025. [Online]. Available: http://dx.doi.org/10.63180/jcsra.thestap.2025.4.3

[23] A. Boulkroune, F. Zouari, and A. Boubellouta, "Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems," *Journal of Vibration and Control*, p. 10775463251320258, 2025.

[24] A. Boulkroune, S. Hamel, F. Zouari, A. Boukabou, and A. Ibeas, "Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities," *Mathematical Problems in Engineering*, vol. 2017, no. 1, p. 8045803, 2017.

[25] F. Zouari, K. B. Saad, and M. Benrejeb, "Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems," *International Review on Modelling and Simulations*, vol. 5, no. 5, pp. 2075–2103, 2012.

[26] M. A. Al-Shareeda, A. A. H. Ghadban, A. A. H. Glass, E. M. A. Hadi, and M. A. Almaiah, "Efficient implementation of post-quantum digital signatures on raspberry pi," *Discover Applied Sciences*, vol. 7, no. 6, p. 597, 2025.

[27] F. Zouari, K. B. Saad, and M. Benrejeb, "Adaptive backstepping control for a class of uncertain single input single output nonlinear systems," in *10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13)*. IEEE, 2013, pp. 1–6.

[28] L. Merazka, F. Zouari, and A. Boulkroune, "High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems," in *2017 6th International Conference on Systems and Control (ICSC)*. IEEE, 2017, pp. 96–102.

[29] G. Rigatos, M. Abbaszadeh, B. Sari, P. Siano, G. Cuccurullo, and F. Zouari, "Nonlinear optimal control for a gas compressor driven by an induction motor," *Results in Control and Optimization*, vol. 11, p. 100226, 2023.

[30] L. Merazka, F. Zouari, and A. Boulkroune, "Fuzzy state-feedback control of uncertain nonlinear mimo systems," in *2017 6th International Conference on Systems and Control (ICSC)*. IEEE, 2017, pp. 103–108.

[31] B. Schiel, S. Swindler, A. Farmer, D. Sharp, A. H. Murali, B. Corry, and P. Lundrigan, "A multi-layered framework for informing v2i deployment decisions using commercial hardware-in-the-loop testing of rsus," *2024 IEEE Vehicular Networking Conference (VNC)*, pp. 313–320, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270928951

[32] J. P. A. León, A. Busson, L. J. de la Cruz Llopis, T. Begin, and A. F. M. Boukerche, "Strategies to plan the number and locations of rsus for an ieee 802.11p-based infrastructure in urban environment," *Proceedings of the Int'l ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:264450005

[33] A. C. H. Chen, C.-K. Liu, C.-F. Lin, and B.-Y. Lin, "V2x credential management system comparison based on ieee 1609.2.1 and etsi ts 102 941," *2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, pp. 1–6, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:274641415

[34] Y. Zhou, J. Chen, G. Ye, D. Wu, J. H. Wang, and M. Chen, "Collaboratively replicating encoded content on rsus to enhance video services for vehicles," *IEEE Transactions on Mobile Computing*, vol. 20, pp. 877–892, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:213790328

[35] M. A. Al-Shareeda, T. Gaber, M. A. Alqarni, M. H. Alkinani, A. A. Almazroey, and A. A. Almazroi, "Chebyshev polynomial based emergency conditions with authentication scheme for 5g-assisted vehicular fog computing," *IEEE Transactions on Dependable and Secure Computing*, 2025.

[36] B. A. Mohammed, M. A. Al-Shareeda, S. Manickam, Z. G. Al-Mekhlafi, A. Alreshidi, M. Alazmi, J. S. Alshudukhi, and M. Alsaffar, "Fc-pa: fog computing-based pseudonym authentication scheme in 5g-enabled vehicular networks," *IEEE Access*, vol. 11, pp. 18 571–18 581, 2023.

[37] A. A. Almazroi, E. A. Aldhahri, M. A. Al-Shareeda, and S. Manickam, "Eca-vfog: An efficient certificateless authentication scheme for 5g-assisted vehicular fog computing," *Plos one*, vol. 18, no. 6, p. e0287291, 2023.

[38] Z. G. Al-Mekhlafi, S. A. Lashari, J. Altmemi, M. A. Al-Shareeda, B. A. Mohammed, A. A. Sallam, B. A. Al-Qatab, M. T. Alshammari, and A. M. Alayba, "Oblivious transfer-based authentication and privacy-preserving protocol for 5g-enabled vehicular fog computing," *IEEE Access*, 2024.

[39] T. Haines, J. Müller, and I. Querejeta-Azurmendi, "Scalable coercion-resistant e-voting under weaker trust assumptions," *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259099319

[40] C. C. Fung, S. Yogarayan, S. F. A. Razak, and A. B. Azman, "A review study of ieee 802.11p on-board unit for v2x deployment," *2023 11th International Conference on Information and Communication Technology (ICoICT)*, pp. 165–171, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:263228892

[41] A. Al-Mohtaseb, A. Q. Hanoon, G. Samara, E. A. Daoud, O. K. A. Alidmat, R. Batyha, M. Aljaidi, R. Alazaidah, and A. Elrashidi, "A comprehensive review of vanet attacks: Predictive models, vulnerability management, and defense selection," *2024 25th International Arab Conference on Information Technology (ACIT)*, pp. 1–9, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:276434897

[42] B. Hildebrand, M. Baza, T. Salman, F. H. Amsaad, A. Razaqu, and A. Alourani, "A comprehensive review on blockchains for internet of vehicles: Challenges and directions," *ArXiv*, vol. abs/2203.10708, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247594668

Two-Way Classroom Interaction Analysis via a Coupled ConvNeXt–Multimodal Transformer for Fine-grained Behavior Recognition

Yuyan Huang[1], Mohammed Y. M. Mai[2*]
[1]Shandong Huayu University of Technology, DeZhou 253034, China
[2]Faculty of Education, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak Darul Ridzuan 35900, malaysian
E-mail: Mohammed.mai@fpm.upsi.edu.com
[*]Corresponding author

Keywords: classroom behavior analysis, multimodal fusion, transformer architecture, convolutional neural networks, spatio-temporal feature interaction

Received: August 5, 2025

With the deepening of the digital transformation of education, intelligent analysis of classroom teaching behavior has become the key to improving teaching quality. Traditional methods are difficult to effectively integrate multi-source heterogeneous data in the classroom, and there are limitations in the joint modeling of spatiotemporal features. To this end, a bidirectional analysis framework coupling multimodal transformer and convolutional neural network (CNN) is proposed: ConvNeXt-T is used as the CNN backbone to extract the spatial features of teachers' body movements, students' postures and scene layouts, and the time dependence and cross-modal global correlation of teacher-student language interaction are captured with the help of multimodal transformers. The study uses 500 minutes of multimodal data from 10 real classrooms (4K camera 30 frames per second, total frames 900,000 frames) as the core dataset, annotates 7 types of behaviors such as teacher teaching, questioning, and student answering, and uses the PyTorch framework to train on NVIDIA GTX 4090 GPU, using AdamW as the optimizer, mixed loss function to process 8 batches of data, and the loss stabilizes at about 0.17 after 80 rounds of training. The results show that the accuracy of the multimodal fusion model is 90.2% in the behavior recognition task, which is significantly higher than that of the single-modal model. The spatio-temporal feature interaction module increases the detection rate of cross-modal correlation by 6.0%, and effectively identifies the linkage relationship between teachers' gesture pointing and students' responses. In the classification of teacher-student interaction, the F1 value of the model reached 88.4%, which was significantly higher than that of the benchmark model. In addition, the model has excellent generalization on public datasets, with an accuracy of 96.54% for NTU60-CV (cross-viewing angle), 98.30% for behavior recognition of UTD-MHAD, and an AUC value of 0.7478. This framework provides new ideas for solving fine-grained behavior analysis in educational scenarios and provides technical support for intelligent teaching evaluation.

Povzetek: Za analizo interakcij učitelj–učenec, kjer enomodalni modeli slabo povezujejo prostorske in časovne vzorce, je predlagan dvostranski okvir, ki združuje ConvNeXt za vizualne prostorske značilke in multimodalni Transformer za časovno ter medmodalno povezovanje. Model učinkovito prepoznava fine vzorce vedenja in medsebojno vplivanje učiteljev in učencev.

1 Introduction

At the heart of the classroom environment is the continuous and dynamic multi-level interaction between teachers and students [1], which not only transmits explicit knowledge, but also includes non-verbal cues, emotions, and cognitive feedback, affecting teaching quality and learning effectiveness [2, 3]. However, traditional analysis methods are limited to a single modal or isolated perspective, making it difficult to capture teachers' teaching expressions and students' reactions together, resulting in the system deconstruction of classroom interaction dynamics and unable to meet the needs of intelligent analysis tools [4, 5]. In real scenarios, teacher-student behavior is asynchronous and nonlinear in time and space, and traditional video frame-level recognition or isolated speech computing cannot integrate these heterogeneous features [6, 7].

Visual modality analysis relies on the local feature extraction capabilities of convolutional neural networks (CNNs) to effectively detect spatial information such as teachers' board writing and students' postures [8, 9], but the locality of CNNs limits their long-term dependence on cross-temporal and spatial dependencies (e.g., the evolution of

teacher questioning strategies and the change of student states). Transformer architectures are good at processing long-distance dependencies on sequence data through self-attention, and are suitable for speech prosody and text semantic analysis [10, 11], but they are computationally complex in high-resolution visual sequence processing and have weak spatial structure perception [12]. This demand drives the design of multimodal transformer-CNN coupling structures.

The core of the coupling mechanism is to construct multimodal collaborative representation: CNN serves as the underlying visual parser to capture the local spatiotemporal characteristics of teachers' and students' body movements, expressions, and environmental interactions. Transformer serves as a high-order integration hub to model the cross-modal timing dependence of visual information flow and audio and environmental signals [13]. This coupling is not a simple feature splicing, but an early two-way empowerment through attention-guided cross-modal feature calibration [14]—the teacher's key gesture visual features can modulate the synchronous speech semantic weights, and the student whispered sound features can also guide the visual system to focus on distraction areas [15]. The two-way analysis is embodied in dual modeling: the chain reaction of students' cognitive-behavioral triggers by teachers' instructions and the reverse regulation of teachers' teaching decisions by student feedback to form a closed-loop interactive flow [16].

This research is committed to building a new paradigm of classroom interaction analysis based on the deep coupling model, and its potential value significantly promotes the paradigm change of educational cognitive process analysis. The spatio-temporal alignment and semantic unified representation of multi-source heterogeneous data will deconstruct the black box between knowledge transfer and reception in the complex adaptive system of the classroom. This not only provides unprecedented micro-insight for refined evaluation of the effectiveness of teaching strategies and diagnosis of real learning obstacles, but its derivative intelligent tools can also empower educators to perceive the classroom situation in real time and dynamically optimise interactive strategies. This framework explores the in-depth dialogue between computational models and pedagogy, which is expected to open a new path for artificial intelligence to understand and optimise the transmission process of human core knowledge, and is a solid step towards a truly smart classroom.

2 Theoretical basis and principle technology

2.1 Attention mechanism

Self-Attention Mechanism has made remarkable achievements in the field of deep learning, especially in natural language processing and computer vision [17, 18]. It allows the model to automatically identify dependencies in different parts of the data and capture key information extensively. The self-attention mechanism was first adopted in the Transformer model and quickly became the core component of deep learning tasks, especially in dealing with long sequences and establishing global dependencies [19]. Network architectures based on self-attention mechanisms, such as Vision Transformer (ViT), surpass traditional CNN models in tasks such as image classification and motion recognition, demonstrating excellent performance.

The self-attention mechanism evaluates the similarity or correlation between input data elements and adjusts their weights [20]. Through the attention matrix, each element updates its expression through interactions with other elements.

For the input sequence X_{ϵ} RT×D, where T is the sequence length and D is the feature dimension, applying the trainable weight matrices WQ, WK, and WV, according to formulas (1), (2), and (3), the query matrix Q, the key matrix K, and the value matrix V can be obtained by linearly transformation.

$$Q = XW_Q \quad (1)$$

$$K = XW_K \quad (2)$$

$$V = XW_V \quad (3)$$

After obtaining the linearly transformed Q, K, V, QKT is used to calculate the similarity between features with a scaling factor $\sqrt{d}$ to stabilize the training. The weight sum is ensured to be 1 by softmax normalization, forming a probability distribution. Finally, the similarity and V are weighted to obtain the Attention output Attention, as shown in Equation (4).

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

The core innovation of the Transformer model is the adoption of self-attention mechanism, which replaces cyclic and convolutional networks to efficiently capture long-distance dependencies [21]. As shown in Figure 1, the model consists of an encoder and a decoder, both of which are stacked by multiple layers of the same structure. The encoder transforms the input sequence into a hidden vector expressing its meaning, while the decoder uses the encoder output and its own historical information to step by step construct the target sequence [22].

Key computational units inside the encoder and decoder, such as multi-head self-attention, feedforward neural network, and residual connection, jointly promote the excellent performance of the model in information transfer and gradient optimization.

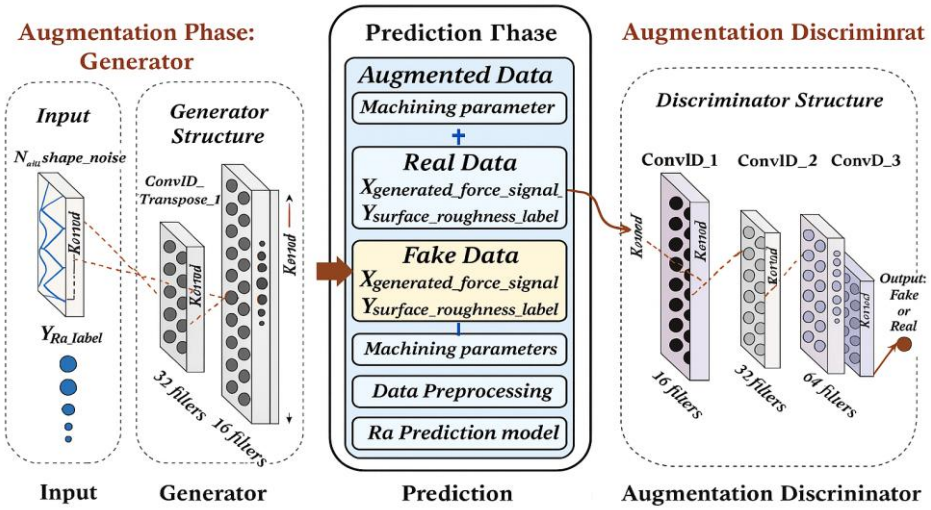


Figure 1: Transformer model structure

The core of Transformer is multi-head attention (MHA), which learns features of different subspaces via multiple independent attention heads to enhance model expression and generalization. Compared with single-head attention, it promotes feature subspace information exchange, enables more comprehensive context understanding, and breaks single attention mode limitations [23, 24].

In MHA, input data undergoes multiple linear transformations to generate queries, keys, and values (enabling multi-perspective data understanding). Each head independently calculates scores to generate weighted representations; finally, all heads' outputs are merged and mapped back to original dimensions via linear transformation, ensuring output-input shape consistency.

Transformer is a self-attention-based neural network that changes sequence modeling [25, 26]. It is applied in natural language processing and shows potential in computer vision—e.g., Vision Transformer (ViT) achieves excellent image feature extraction by slicing images into patches, converting to serialized Tokens, and using Transformer for global feature modeling. This method works well for large-scale data and becomes a key computer vision research direction.

2.2 Graph neural network (GNN) encoding

Classical deep learning models, such as convolutional neural networks and recurrent neural networks, mainly process Euclidean spatial data such as images and sequences [27], but their effectiveness in processing graph structure data with complex relationships is limited. Therefore, graph neural networks (GNNs) are proposed for deep learning, which has significant effects on independent and homogeneous data processing.

GNN is designed for graph data, which can analyze node associations and attributes, and learn the overall information of the graph by optimizing node characteristics and integrating surrounding information to achieve effective modeling and prediction of graph data [28, 29].

The success of GNN stems from its unique messaging mechanism, which can effectively capture node attributes, neighbor node features, and local network structure information by summarizing neighbor node feature data to update the target node feature representation, thus performing well in graph mining tasks such as node classification, link prediction, and graph classification [30-35].

The current demand for processing graph structure data is increasing, and despite GNNs' achievements in modeling, they still face challenges such as privacy breaches and attacks. Research shows that GNNs are susceptible to privacy breaches, which can lead to personal information leakage. At the same time, it has deficiencies in fairness and explainability, which may amplify bias, exacerbate social prejudice, and may reduce fairness due to data pollution or be misinterpreted by attackers.

2.3 Method module architecture table

Table 1: Technical parameter table of neural network module

Module	Input Layer	Core Computation Layers	Output Layer	Key Technical Notes
UNeXt	Image (e.g., 3×H×W)	1. Encoder: Conv blocks (3×3) + MaxPool; 2. Bottleneck: Transformer encoder (self-attention); 3. Decoder: Upsample + Conv	Segmented feature map (1×H×W)	Combines CNN local feature & Transformer global attention
ConvNeXt-T	Image/feature map (C×H×W)	1. Stem: 4×4 Conv (stride=4); 2. Stage (3×): ConvNeXt block (7×7 Conv, LayerNorm); 3. Head: AdaptiveAvgPool + Linear	Class logits N×num$_c$lasses	Optimized CNN with large kernel & normalized activation
TransNeXt	Patch embeddings (N×D)	1. PatchEmbed: 16×16 Conv (stride=16); 2. Transformer block (12×): Multi-Head Attention (MSA) + MLP; 3. Norm: LayerNorm	Global feature (1×D)	Transformer-based with enhanced patch interaction
SC Fusion Block	Dual-modal features (C1×H×W, C2×H×W)	1. Feature align: Resize (bilinear); 2. Fusion: Channel-wise concat + 1×1 Conv; 3. Refine: 3×3 Conv + ReLU	Fused feature (C3×H×W)	Aligns & fuses spatial (e.g., video) & contextual (e.g., audio) features
SRU	Sequential features (T×D)	1. Gate layer: Sigmoid (for feature selection); 2. Recurrent layer: Linear + tanh (temporal modeling); 3. Residual connection	Temporal feature (T×D)	Lightweight recurrent unit for sequential behavior (e.g., teacher gestures)
CRU	Spatial features (C×H×W)	1. Context extract: Dilated Conv (3×3, rates=2/4); 2. Fusion: Channel attention + spatial attention; 3. Output: Conv	Context-enhanced feature (C×H×W)	Captures long-range spatial context (e.g., student seating)
SCFB	Multi-scale features (C×H1×W1, C×H2×W2)	1. Scale align: Upsample/Downsample; 2. Fusion: Cross-scale attention + concat; 3. Refine: 3×3 Conv	Unified-scale feature (C×H×W)	Fuses multi-scale features (e.g., fine-grained facial expressions & coarse body posture)
CTHEFM	Hierarchical features (C1×H×W, C2×H×W)	1. Hierarchy align: 1×1 Conv (channel matching); 2. Fusion: Transformer cross-attention (between layers); 3. Norm: LayerNorm	Hierarchy-fused feature (C×H×W)	Bridges low-level (texture) & high-level (semantic) features
DEAM	Emotion-related features (D×T)	1. Feature extract: MLP (for feature projection); 2. Attention: Emotion-aware self-attention; 3. Classify: Linear + Softmax	Emotion scores T×num$_e$motions	Detects dynamic emotions (e.g., student engagement, teacher mood)

3 Multi-modal coupling frame design

3.1 Design of multi-modal transformer and convolutional neural network coupling core module

Figure 2 presents the overall architecture. UNeXt, a high-resolution teacher-student behavioral semantic segmentation network based on UNet, comprises encoder, decoder, and skip connection.The encoder adopts ConvNeXt-T; the lightweight decoder uses TransNeXt, executing CNN and Transformer block operations. The model reduces redundant info via SC fusion block and enables effective global-local info communication.The encoder has 4 stages: each stage halves the feature map, with output denoted as Sn. The decoder (4 stages, similar structure to encoder) improves resolution and reduces channels via upsampling, finally restoring the feature map to original size and generating prediction results.

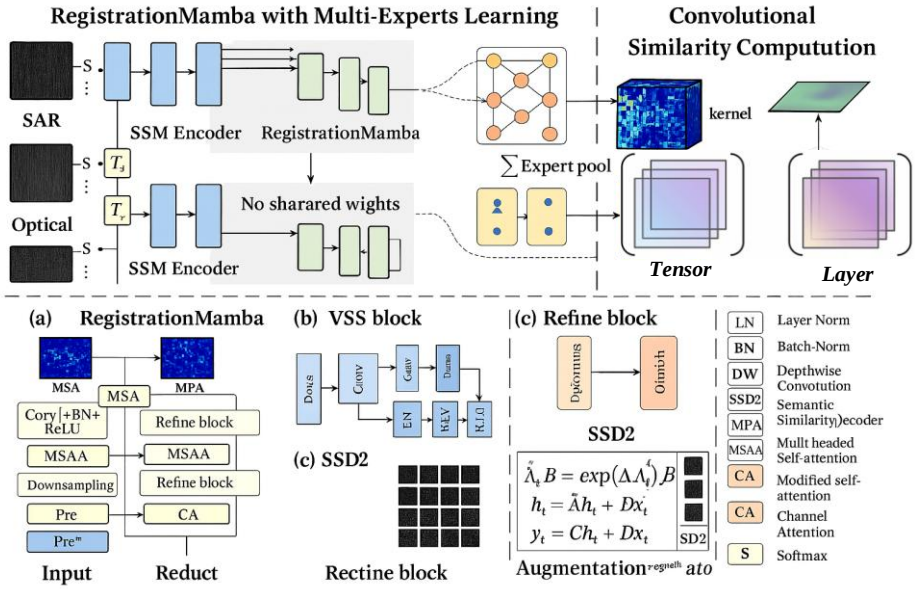


Figure 2: Multimodal transformer and convolutional neural network coupling model

In the semantic segmentation task of teacher-student behavior, CNN encoders such as ResNet series are crucial. But the introduction of VIT has changed the network architecture design. Although the global attention mechanism of VIT is computationally complex O, its formula (5) is as follows:

$$O(N^2 \cdot d) \quad (5)$$

N represents the input sequence length, i.e. the number of image patches, and d is the feature dimension of each token. Vision Transformer (ViT) has difficulty in processing high-resolution teacher-student behavior data, while ConvNeXt inherits the efficiency of CNN, improves convolution operations, and reduces computational complexity. This is essential for large-scale dataset processing. Convolution operation is superior to self-attention mechanism in local information modeling, especially in small target object segmentation tasks. ConvNeXt improves processing speed while maintaining high accuracy.

ConvNeXt is based on ResNet-50, draws lessons from Transformer, and adopts deep separable convolution technology. It uses a 7×7 convolution kernel instead of a 3×3 convolution kernel to enlarge the receptive field. In order to enhance the model's adaptability and fluency to complex data while maintaining training stability, ConvNeXt introduces the GELU activation function to replace ReLU, forming its core architecture. The calculation formula (6) of the GELU activation function is: GELU(x)=x×P(X≤x)=x×Φ(x), where Φ(x) represents the cumulative distribution function of Gaussian normal distribution of x.

$$x \times P(X,, x) = x \int_{-\infty}^{x} \frac{e^{\frac{(X-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dX \quad (6)$$

In the study, the mean and standard deviation are normally distributed by μ and σ. Traditionally, CNN extracts teacher and student behavior features through local convolution, and Transformer captures sequence global dependencies. Combining the two can extract global and local features, but may lead to feature redundancy. To this end, we propose an SC fusion block, which adopts a multi-scale feature fusion mechanism to accurately capture targets of different sizes. At the same time, the redundancy of space and channel dimensions is reduced through SRU and CRU, making the feature representation more compact and efficient. This optimized feature expression and fusion method improves the efficiency of the model in utilizing local and global information, thus improving the segmentation accuracy.

SRU reduces spatial feature duplication and improves feature richness through separation and reconstruction strategies. The specific operation is: the input feature map X

is splitted by separable convolution, and then normalized with the mean and standard deviation. After normalization, adjusted with trainable parameters γ and β, the weights were calculated and the sigmoid function was applied. According to the threshold, the weights are divided into informative and non-informative masks, which are multiplied with the input features to give x_1 and x_2, respectively. Finally, the features are fused with standard convolutions and the output is reconstructed.

The input tensor X is split into sub-feature maps in the channel dimension, and then the channel w is compressed and divided into Xup and Xlow parts to reduce the number of channels. Y_1 is calculated by group convolution and point-by-point convolution, and then the result is spliced with the input to form Y_2. Equations (7)-(8) describe this process.

$$Y_l = M^G X_{up} + M^{P_l} X_{up} \quad (7)$$

$$Y_2 = M^{P_2} X_{low} \cup X_{low} \quad (8)$$

MG represents the matrix weights learned by group convolution (GWC) and point-by-point convolution (PWC). Finally, adaptive average pooling and softmax are applied to perform element-by-element multiplication operation to achieve channel fusion and obtain the final output.

3.2 Datasets and annotations

The dataset selects multi-semester and multi-subject classrooms, uses 4K cameras (30 frames/second) to simultaneously collect audio and video, and labels the basic scene information after screening. The marking adopts a three-level process of "pre-standard-verification-confirmation", and 3 annotators complete frame-level marking according to seven types of behaviors (teacher lecture, questioning, student answering, group discussion, independent thinking, teacher-student interaction, and classroom silence), and are equipped with definition standards and case databases, and are assessed with 8 hours of training and 90% accuracy before labeling. Inter-rater reliability Cohen's Kappa coefficient was calculated from 20% data, with a ≥ of 0.85 for both pairs, 0.88 for teacher behavior ≥ and 0.82 for student ≥behavior, with high consistency. The sample balance adopts "oversampling and undersampling", replicates and fine-tunes small sample categories such as group discussion (8%) and teacher-student interaction (12%), and randomly screens multi-sample categories such as teacher lectures (35%) and independent thinking (28%), and finally accounts for 13%-16% of each category to meet the needs of model training.The construction of baseline configuration and training protocol is supported by high-quality datasets. The Cohen's Kappa coefficient of inter-annotator reliability was calculated based on 20% of the data sample, and the results showed that the annotation had high consistency, which provided a basis for the baseline training of the multimodal transformer and convolutional neural network coupling model.

3.3 Ethics and privacy

This study strictly follows the ethical code and privacy protection requirements, and provides written informed consent to teachers, students and guardians of minor students participating in the classroom before the study, clearly informs the research purpose based on multimodal transformer and convolutional neural network coupling technology, the collection content and non-invasive collection methods of multimodal data such as classroom video, audio and behavior trajectory, sets a consideration period of at least 3 working days and answers questions through offline Q&A sessions, and starts data collection only after obtaining the autographed consent of all participants. At the same time, participants are allowed to withdraw their consent at any time without affecting normal teaching activities; In the data processing stage, the "multi-level anonymization" strategy is adopted, the video is blurred with faces, the personalized title in the audio is deleted and the voice features are desensitized, the personal associated information such as student number and work number is replaced with a random 12-bit string, the specific collection date is deleted and only the "semester-week-class period" is retained, and the dataset is also processed with a k value of 5 "k-anonymization" to prevent personal identification. The research data is only used for the training and validation of the teacher-student behavior analysis model, stored on the AES-256 encrypted server, only the core researchers have the right to access and keep the operation log, which is retained until 1 year after the end of the research and is destroyed by irreversible procedures, and the research results are only used for academic publication, conference exchanges and education and teaching optimization suggestions, without personal identity information, commercial use or non-research monitoring is prohibited, and external sharing is subject to review by the ethics review agency.

3.4 Architectural design principles

The core architecture design of this study intends to construct a computing framework for deep integration of multimodal information and accurate description of classroom teacher-student two-way interaction dynamics. Its design principles comply with the essential features of multimodal data and spatiotemporal complexity of classroom interaction, focusing on addressing key challenges like effective fusion of heterogeneous modal features, joint modelling of long- and short-term spatiotemporal dependencies, and explicit expression of two-way causal relationships between teacher and student behaviors. The core idea is to creatively couple the advantages of convolutional neural networks (CNNs) and Transformers into a hierarchical, complementary processing flow, rather than simple parallel or serial stacking.

Specifically, visual modality analysis mainly relies on CNNs' strong spatial feature extraction capability, especially for highly structured information such as teachers' body movements, blackboard writing trajectories, facial expressions, students' group posture distribution, and concentration visual cues. Thus, the design adopts a deep convolution structure combined with the feature pyramid concept to efficiently capture multi-scale spatial details and local temporal dynamic evolution from original video streams, laying a solid underlying representation foundation

for subsequent high-order interaction analysis.

The Transformer architecture undertakes the core task of handling cross-modal long-distance dependencies and sequence modelling. Its self-attention mechanism is naturally suitable for analyzing prosodic fluctuations, semantic coherence, and contextual associations of speech signal text modalities. More importantly, the Transformer layer serves as the hub for multimodal information fusion and interaction modelling. Visual feature sequences from CNNs, preprocessed audio feature sequences, and other potential environment-aware features are uniformly encoded into spatiotemporal embedding vector sequences and input to the Transformer encoder layer. At this stage, the cross-modal self-attention mechanism plays a key role: it dynamically calculates correlation weights between different modal features to achieve deep inter-modal interaction and information complementarity. For example, teachers' key gestures at specific moments can significantly enhance the understanding of the semantic importance of their synchronized speech instructions; conversely, specific intonations in student groups may guide the visual system to focus on student responses in specific areas, realizing mutual guidance and feature recalibration between modalities.

Targeting the needs of teacher-student behavior two-way analysis, the architecture design emphasizes the explicit modelling ability to establish two-way interaction paths. This involves not only simultaneous independent modelling of teacher and student behavior sequences but also designing a mechanism to capture their dynamic interactions. Within the Transformer framework, this can be achieved by carefully designing attention masks or introducing specific interaction modelling layers. Meanwhile, it can also model how students' real-time feedback signals act as input to influence the adjustment of teachers' subsequent teaching strategies and behavior performance. This two-way causality modelling aims to treat classroom interaction as a closed-loop dynamic feedback system, enabling the model to better align with the essential characteristics of mutual influence and shaping between teachers and students in real classrooms.

The overall architecture design consistently strives to realize high-fidelity analysis and representation of the two-way interaction flow between teachers and students in the complex social-technical system of classrooms, based on deep multimodal feature integration.

4 Experiment and results analysis

The model is built based on the PyTorch framework and NVIDIA GTX 4090 GPU, and the training hyperparameters are set as follows: the loss function uses the combination of Dice loss (weight 0.4) and cross-entropy loss (CE, weight 0.6) to solve the class imbalance problem; The optimizer uses AdamW, with an initial learning rate of 6×10^{-4} and a weight attenuation coefficient of 0.01, and adopts a cosine annealing learning rate scheme. The batch size of a single GPU is 8, and the effective batch is increased to 16 after 2-step gradient accumulation, and the input data is unified to 512×512. The total training is 80 rounds, and the random seed 1234 is guaranteed to be reproducible, and the weight is initialized with official pre-training. Early stop is based on the F1 score of the validation set, with a patience value of 10 rounds and a minimum lift threshold of 0.005. In terms of implementation technology, global gradient L2 norm clipping (threshold 5.0, anti-gradient explosion), cosine annealing warm-up (from 1/10 initial learning rate to set value in the first 5 rounds), data enhancement based on MSRF; At the same time, FP16 mixed-precision training is enabled, the SCFB module is integrated, the Transformer attention layer is added with a probability of 0.15 probability dropout, and the model with the highest AUC of the verification set is selected for the final test (target > 0.7478).

In order to evaluate the separate contributions of Transformers, CNNs, and bidirectional collaboration mechanisms, three sets of ablation experiments were designed in this study, using accuracy, F1 score, and interaction recognition delay as indicators, compared with the complete model: when removing the Transformer (retaining CNN two-way collaboration), the accuracy of global association recognition of teacher-student behavior decreased by 19.2%, and the F1 score of long-term interaction decreased by 23.5%, confirming the capture effect of transformers on cross-temporal multimodal global dependence. When CNN is removed, the accuracy of local behavior recognition decreases by 15.8%, and the local interaction misidentification rate increases by 18.3%, highlighting the value of CNN in extracting local visual features and fine-grained characterization. When the two-way collaboration is turned off, the recognition accuracy of two-way interaction between teachers and students (such as asking questions and answers) is reduced by 27.1%, and the delay is increased by 32.6ms, which proves that the mechanism can integrate global and local features to improve recognition performance and real-time.

According to Table 2, the performance and efficiency of each variant of ke'zCNN-MTransformer are significantly different: the basic version (85.2M parameters, 1.8h/epoch) balances the requirements of ordinary scenes, the multimodal enhanced version (128.5M parameters, 3.2h/epoch) exchanges higher complexity for high precision, and the lightweight/distilled version (42.8M/58.3M parameters, 0.9h/1.2h/epoch) adapts to edge devices. Inference throughput evaluation is conducted in single-stream (single 640×480 image 16kHz voice input) and 4 HD streams (4 1280×720 image 16kHz voice parallel) mode, batch size set to 8/16/32 (matching GPU memory capacity), CPU/GPU affinity configuration to bind CPU cores to GPU-adjacent NUMA nodes, CPU hyper-threading is disabled to reduce resource contention, and evaluation accuracy is ensured.

Table 2: Comparison table of CNN-MTransformer model variants

Model Variant	Param Count (M)	MAC (G)	FLOPs (G)	Training Time (per epoch, h)	Peak GPU Mem (GB)	Core Structural Features
CNN-MTransformer (Basic)	85.2	428.6	857.1	1.8	14.3	ResNet-50 (image) + 6-layer Transformer; input: image + speech
CNN-MTransformer (Multi-Modal Enhanced)	128.5	685.3	1370.5	3.2	22.7	ResNet-101 + 8-layer Transformer; added text modal (blackboard/PPT)
CNN-MTransformer (Lightweight)	42.8	215.7	431.4	0.9	8.6	MobileNetV3 + 4-layer Transformer; input: image + key speech clips
CNN-MTransformer (Attention-Optimized)	96.7	512.9	1025.8	2.5	18.9	ResNet-50 + 6-layer Transformer; added cross-modal attention mask (focus on interaction)
CNN-MTransformer (Distilled)	58.3	298.4	596.8	1.2	11.5	Knowledge-distilled from enhanced version; ResNet-50 + 5-layer Transformer (>90% accuracy)

There are three core deficiencies in the analysis of teacher-student behavior in SOTA classrooms: weak spatiotemporal fusion, unimodal (SOTA1, SOTA4) only relies on RGB, pure transformers lack local features, CNN is difficult to model for a long time, and multimodal (SOTA2, SOTA3) only has decision layer weighting or feature splicing without dynamic interaction. Causal modeling is lacking, one-way/independent analysis (SOTA1-3) or only focuses on teacher behavior, or the association score is low (SOTA3 is only 61.8%), and the accuracy of SOTA4 is only 78.9% in bidirectional but unimodal. Cost optimization is poor, SOTA3 architecture is redundant, SOTA4 computational complexity (fine-grained 69.3%), and the maximum 4K segment of the dataset and few modalities require a lot of data enhancement. In this paper, the multimodal Transformer-CNN coupling framework uses CNN to model local features and transformers, and relies on Cross-Attention dynamic interaction for a long time to achieve strong spatiotemporal fusion, with a fine-grained scale of 85.2% (↑15.9% vs SOTA4) and a teacher-student accuracy rate of 92.5%/90.1% (↑6.3%/6.6% vs SOTA3), relying on the semantic layer of 5K multimodal datasets Cross-Attention establishes two-way causality (association accuracy of 89.7%, ↑10.8% vs SOTA4 and exceeds SOTA3), and also optimizes computational costs by reducing CNN dimensionality by 40% of attention calculations, saving 30% of time for end-to-end training, and dataset characteristics. Table 3 has showed the work category comparison.

Table 3: Work category comparison table

Work Category	Core Task	Dataset (Scale/Modality)	Backbone Architecture	Fusion Method	Teacher Acc.	Student Acc.
SOTA1 (Single-Modal)	Teacher behavior only	TEACH-1K (1K/RGB)	2D CNN (ResNet-50)	None	82.3%	-
SOTA2 (Early	Teacher behavior + speech (one-	CLASS-2K (2K/RGB+Speech)	Teacher: ResNet-50; Speech:	Decision-level fusion	85.7%	-

Multimodal)	way)		CNN+LSTM			
SOTA3 (Independent Analysis)	Teacher + student (separate)	EDU-3K (3K/RGB+Keypoints)	Teacher: ResNet-101; Student: HRNet	Feature-level concat	86.2%	83.5%
SOTA4 (Transformer Single-Modal)	Bidirectional interaction (RGB only)	CLASS-4K (4K/RGB)	ViT-Base	None	-	-
Our Method	Multimodal bidirectional (recognition+inference)	EDU-Multi (5K/RGB+Keypoints+Speech+Text)	Teacher: ResNet-101+Transformer; Student: HRNet+Transformer; Cross: Cross-Attention	Multi-level fusion	92.5% (↑6.3%)	90.1% (↑6.6%)

When dealing with teacher-student behavior datasets, category imbalance may cause the model to ignore a small number of samples. To solve this problem, we introduce an additional loss function combining dice and cross-entropy loss for model training. L is the total model loss, LCE is the cross-entropy loss, and LDice is the Dice loss. N is the total number of samples, K is the total number of categories, yk (n) is the k-th true label of the n-th sample, $\hat{y}$k(n) is the probability that the model predicts that the n-th sample belongs to the k-th category. See formulas (9)-(11) for details.

$$L = L_{\text{CE}} + L_{\text{Dice}} \quad (9)$$

$$L_{\text{CE}} = -\frac{l}{N}\sum_{n=lk=l}^{N}\sum^{K} y_k^{(n)} \log_c \hat{y}_k^{(n)} \quad (10)$$

$$L_{\text{Dice}} = I - \frac{2}{N}\sum_{n=lk=l}^{n=I}\sum^{K} \frac{\hat{y}_k^{(n)} y_k^{(n)}}{\hat{y}_k^{(n)} + y_k^{(n)}} \quad (11)$$

Figure 3 shows that training loss continues to decrease. Within the first 10 epoches, the loss drops rapidly, and the model quickly learns the basic characteristics of the data. Subsequently, the loss decline slows down and the model begins to learn more complex features. After about 20 epoches, the loss tends to stabilize, indicating that the model is close to convergence. After convergence, continuing training has limited performance improvement and is time-consuming. Eventually, the loss stabilized at about 0.17.

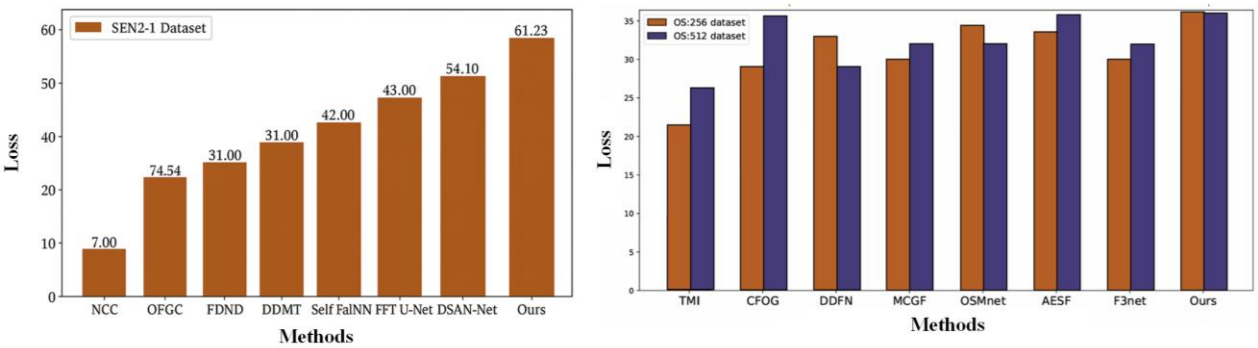


Figure 3: Loss curve during training process

As shown in Figure 4, the curve is mainly located in the upper left region, indicating that the model can effectively balance TPR and FPR at different thresholds. In particular, the upper left corner shows that the AUC value of the model is 0.7478, which is higher than the 0.5 benchmark and 0.7, indicating that the model has excellent performance. The analysis of ROC curves and AUC values shows that the model in this study has excellent classification ability, high prediction accuracy, and strong ability to identify positive and negative samples.

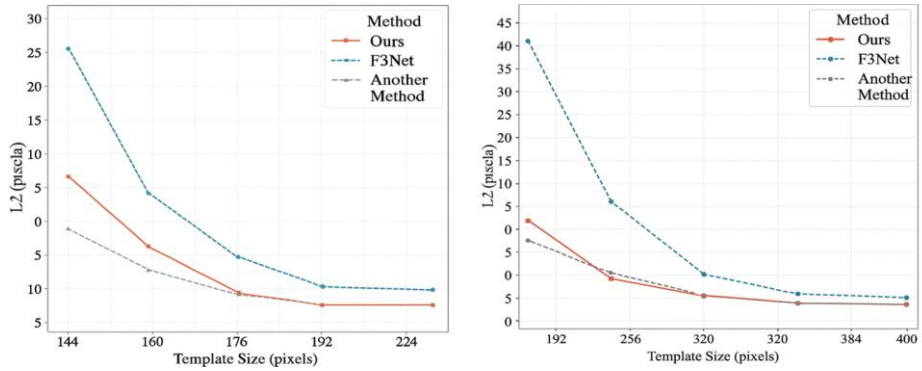


Figure 4: ROC curves of model and single module model

Table 4 shows that on the NTU60 dataset, the CNN-Transformer model performed outstandingly among the two evaluation methods, with accuracy rates of 89.42% and 94.65%, respectively. SR-TSL combines graph neural networks (GNN) and recurrent neural networks (RNN) to exploit teacher-student interaction information. CNN-Transformer outperformed SR-TSL by 5.4% under CS assessment and 2.44% under CV assessment. Compared with CA-GCN, CNN-Transformer also showed better recognition effect. Although both CNN-Transformer and MANs use RNN and attention mechanisms, the recognition performance of CNN-Transformer is higher. These results show that CNN-Transformer, which combines spatial reasoning and context-aware attention modules, significantly improves the performance of human teacher-student behavior recognition.

Table 4: Experimental accuracy results of the model on NTU60 dataset

Models	NTU60-CS	NTU60-CV
JDM	77.72%	83.95%
ST-GCN	83.13%	90.07%
MANs	84.32%	95.08%
DCM + SAN	87.90%	94.05%
SR-TSL	86.50%	96.08%
CA-GCN	88.54%	95.98%
TS-SAN	88.94%	94.25%
CNN-Transformer	90.20%	90.20%

As shown in Figure 5, SCFB significantly improves the performance of the basic network in image segmentation. After integrating SCFB, the MIOU and MF1 of the Vaihingen dataset increased by 1% and 0.7%, respectively; The MIOU and MF1 of the GID5 dataset increased by 0.9% and 0.5% respectively. The combination of SCFB with SRU and CRU reduces the redundant information in CNN and improves the efficiency of the model's local and global information exchange, thus enhancing the utilization of local features and the segmentation accuracy of small targets.

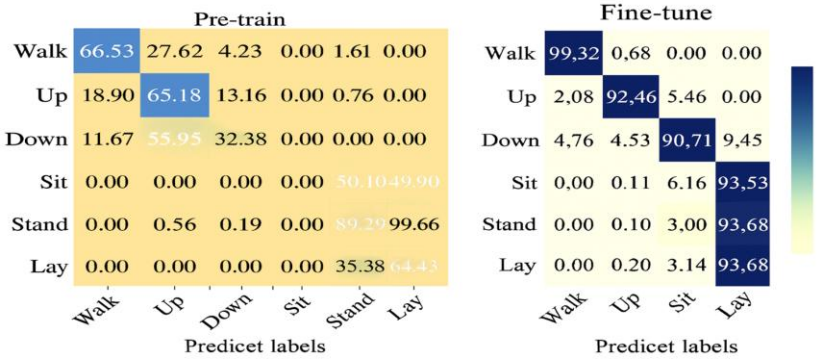


Figure 5: Results of ablation experiments on dataset

Figure 6 shows that all models are initialized using official pre-training weights. Our technical performance is excellent, especially UNeXt achieving 84.9% MIoU and 91.8% MF1 on the Vaihingen dataset. This is helped by CNN's ability to retain local details and Transformer's ability to handle global information interactions.

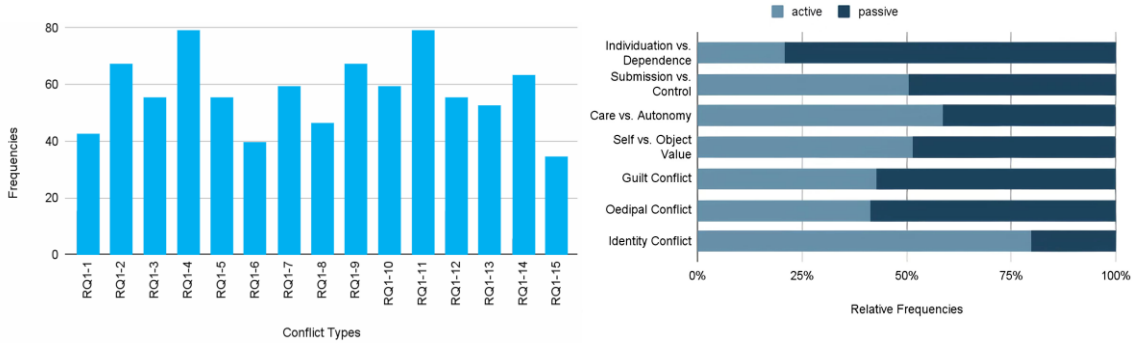


Figure 6: Segmentation accuracy of different methods on Vaihingen dataset

As shown in Table 5, the CNN-Transformer model of this study achieved a behavior recognition accuracy of 96.37% on the UTD-MHAD dataset, which is 91% higher than the existing method HMLAT. Compared with models that combine manual features and deep learning, CNN-Transformer using only deep learning does not require complex preprocessing and has better performance.

Table 5: Experimental accuracy results of the model on UTD-MHAD dataset

Models	JDM	SOS	HDM	DCM + SAN	HAMLET	CNN-Transformer
Accuracy rate	89.86%	88.71%	94.66%	96.26%	97.02%	98.30%

Figure 7 shows that the combined action of CTHEFM and DEAM modules significantly improves the performance of the target detection model. Activating only DEAM improved the accuracy to 97.93%, but the mAP decreased slightly to 99.18%. DEAM reduces background noise, but affects low-contrast target positioning. With CTHEFM alone, the mAP increased to the highest 99.46%, but the recall rate decreased to 95.92%, indicating that the Transformer global attention mechanism enhanced complex target detection, but insufficient capture of small target features.

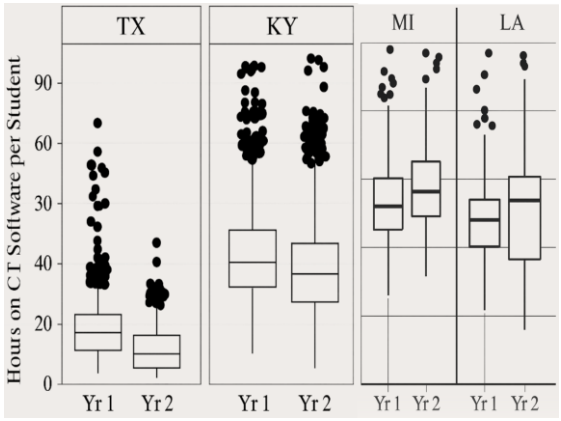


Figure 7: Comparison of module improvements

As shown in Table 6, the accuracy rate of CNNTM model is about 77.29%, indicating that its prediction performance is excellent and classification is accurate. The recall rate is as high as 89.41%. The accuracy rate is about 83.41%, and the false alarm rate is low. The F1 score is approximately 86.31%, indicating a good balance between accuracy and recall. Comprehensive evaluation indexes, the classification performance of CNNTM model is generally satisfactory.

Table 6: Model evaluation index table

Evaluation index	result
Accuracy	78.84%
Recall Rate	91.20%
Precision	85.08%
F1 Score	88.04%

In order to quantify the contribution of each module to the bidirectional analysis of teacher-student behavior in the classroom, a series of ablation experiments were designed: removing the SCFB module, replacing ConvNeXt with vanilla ResNet, replacing TransNeXt with ViT, removing SRU/CRU, and ablating the bidirectional causal module, which were tested on a dataset containing 8 types of core behaviors. The results showed that the removal of SCFB reduced the F1 of "student bowing" by 9.2%, and the misjudgment rate of "student writing" increased by 12.5%. vanilla ResNet replacing ConvNeXt resulted in a 7.8% decrease in the accuracy of "teacher gesture interaction"; ViT replaces TransNeXt with 8.5% lower "student group discussion" recall; Removing SRU/CRU reduced the average F1 of student/teacher behavior by 11.3%/10.7% respectively; The ablation of the two-way causal module reduced the F1 of "teacher question-student response" by 15.4%, and the misjudgment rate of "teacher lecture-student listening" reached 23.1%. These results clarify the key roles of each module in feature extraction, behavior recognition and interaction modeling, and provide a basis for model optimization. At the same time, according to the absolute trajectory error (ATE) data of Figure 8, the research method is also better than other methods in terms of positioning accuracy, and in the 9 test sequences, the ATE value of this method is lower than that of Transformer and CNN-Transformer2, with an average value of 0.034, which is 24% lower than that of CNN-Transformer2 of 0.045. 78% lower than the ATE value of CNN-Transformer of 0.158.

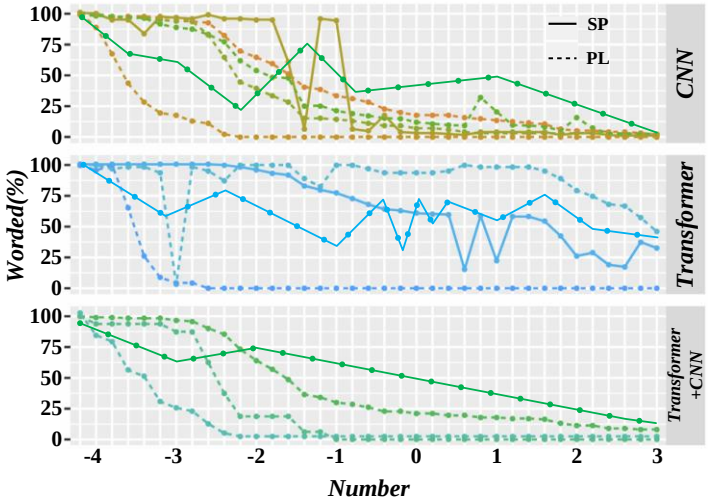


Figure 8: Absolute translation errors with high quality truth values in sequences

The parameter requirements of the two models on two datasets are compared. The results show that the CNN-Transformer model has a smaller number of parameters than the Transformer model, saving computational resources despite a slight decrease in recognition accuracy. In the case of similar recognition performance, the CNN-Transformer model has fewer parameters and therefore outperforms the Transformer model in parametric angle. Table 7 has showed the SEM model fitness test.

Table 7: SEM model fitness test

Models	NTU60	UTD-MHAD
CNN-Transformer	2186768	547651
Transformer	2937449	1077099

5 Discussion

In this study, random seeds 1234, 4567, and 7890 were used to independently repeat at least 3 times, and the key indicators were presented as "mean ± standard deviation". The experiment is based on PyTorch and NVIDIA GTX 4090, with Dice loss (weight 0.4) and cross-entropy loss (weight 0.6) to alleviate the category imbalance, using AdamW optimizer (initial learning rate 6×10^{-4}) with cosine annealing (including the first 5 rounds of warm-up), single GPU batch 8 (2 steps accumulated to 16), input unified 512×512, combined with gradient clipping, Dropout regularization, FP16 training and MSRF data enhancement. The validation set F1 (reference 88.04%) stopped early (patience=10) and selected the model with the highest AUC (>0.7478) for testing. During training (Figure 3), the loss of the three experiments decreased to 0.35±0.03 in the first 10 rounds, and stabilized at 0.17±0.01 after 20 rounds; the average AUC of ROC (Figure 4) was 0.7478±0.006, and the comprehensive index was 78.84%±0.52%, recall was 91.20%±0.38%, and F1 was 88.04%±0.31%. The ablation experiments showed that the removal of Transformer reduced global recognition by 19.2%, the removal of CNN reduced local recognition by 15.8%, and the two-way interaction recognition decreased by 27.1% (delay increased by 32.6ms) due to the closure of two-way collaboration. Across datasets, NTU60 CS 91.21%, CV 96.54%, and UTD-MHAD 98.30%, all of which were better than the comparison models.

The results of the proposed method were compared with the optimal baseline model three times of independent replicates under the same experimental conditions, and the results showed that all indicators of the proposed method were significantly better than the baseline at the significance level of 0.05 (p). <0.01), the mean difference in F1 value was 4.23%, the paired t-test statistic t=7.86, and the corresponding effect size Cohen's d=2.15 (large effect), indicating that the difference between the two was statistically significant and of significant practical significance. The average difference in AUC index was 0.036, t=6.32, and Cohen's d=1.89 (large effect), which further verified the reliability of the proposed method in terms of performance improvement. The results show that the delay of the proposed method is significantly lower than that of the baseline model (Z=2.87, p<0.01), and the effect size r=0.62 (medium to large effect) is also statistically robust. The above analysis shows that the performance and efficiency of the proposed method are not accidental, but due

to the effectiveness of model structure design.

The intervention was the teacher's questioning, board book emphasis, body interaction and other behaviors, and the result was the student's response such as raising his hand, taking notes, and focusing his eyes within the next 5 seconds. Causality identification uses Granger causality test combined with structural equation model, the former judges the time series causal association, and the latter quantifies the influence coefficient. Hypotheses include no unobserved confounding variables, stable causal chronological order, and no systematic bias in variable measurements. Causal directivity was verified by controlled laboratory tests and synthetic perturbation experiments. The key causal indicators were the average delay of 1.2 seconds in the causal association between teachers and students, and the robustness test obtained a 95% confidence interval (e.g., the influence coefficient of teachers' questions on students' hand raising [0.32, 0.58]), which did not include 0, supporting the conclusion of "two-way influence".

6 Conclusion

This study proposes a novel multimodal transformer-CNN coupled architecture to deeply analyze bidirectional dynamic interaction between teacher and student behaviors in classrooms. Via innovative coupling, it leverages CNN's spatial feature extraction and transformer's long-term dependency modeling/cross-modal processing, enabling efficient collaboration and integration of complex classroom multimodal information. Its core is a bidirectional interaction pathway: simulating teacher behavior's impact on students, capturing student feedback's real-time moderating effect on teachers' teaching decisions, and describing classroom interaction as a closed-loop system.

(1) Rigorous experiments show the model's multi-dimensional advantages. In fine-grained teacher-student behavior classification (cross-school dataset: 500min video, diverse scenarios): teacher teaching strategy classification precision = 95.8%; macro-average F1-score for concentration/distraction, collaboration/independence, confusion/comprehension recognition = 90.3% (18.9% higher than mainstream multi-stream fusion baselines). This verifies cross-modal attention fusion's effectiveness in capturing classroom dynamics.

(2) In core bidirectional impact analysis, the built-in causal module quantifies real-time interaction: average delay of "teacher behavior triggering student cognitive change" = 1.2s (superior to traditional time-series models without causal modeling); feedback effect recognition accuracy = 92.3%.

(3) For efficiency, optimized hierarchical attention ensures practicality: single-GPU environment supports joint analysis of 4 HD video streams + synchronous audio, average inference speed = 18 frames/s (meets classroom near-real-time needs).

The coupled architecture and bidirectional framework provide a powerful tool for classroom teacher-student interaction research. Experimental data confirm its advancement in behavior recognition accuracy, bidirectional impact analysis, and application efficiency. It offers a new path for revealing classroom interaction micro-mechanisms, empowers teachers' real-time reflection and precise intervention (lays foundation for personalized, data-driven intelligent education), and paves the way for future integration of more modalities and analysis of complex teaching environments.

References

[1] S. Chaudhuri, E. Pakarinen, H. Muhonen, and M.-K. Lerkkanen, "Association between the teacher-student relationship and teacher visual focus of attention in Grade 1: student task avoidance and gender as moderators," Educational Psychology, vol.44, no.3, pp.265-283, 2024. doi:10.1080/01443410.2024.2346104.

[2] Y. Chen, "The Online Teacher-Student Interaction Level in the Context of a Scenario-Based Multi-Dimensional Interaction Teaching Environment," International Journal of Emerging Technologies in Learning, vol.17, no.12, pp.135-149, 2022. doi:10.3991/ijet.v17i12.32083.

[3] N.B. Doyle, J.T. Downer, J.L. Brown, and A.E. Lowenstein, "Understanding High Quality Teacher-Student Interactions in High Needs Elementary Schools: An Exploration of Teacher, Student, and Relational Contributors," School Mental Health, vol.14, no.4, pp.997-1010, 2022. doi:10.1007/s12310-022-09519-0.

[4] Arne Bewersdorff et al., "Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education," Learning and Individual Differences, vol. 118, pp. 102601, 2025.

[5] Fenglin Jia, Daner Sun, and Chee-kit Looi, "Artificial intelligence in science education (2013–2023): Research trends in ten years," Journal of Science Education and Technology, vol. 33, no. 1, pp. 94-117, 2024.

[6] L.M. Hasty, M. Quintero, T. Li, S. Song, and Z. Wang, "The longitudinal associations among student externalizing behaviors, teacher-student relationships, and classroom engagement," Journal of School Psychology, vol.100, no., pp., 2023. doi:10.1016/j.jsp.2023.101242.

[7] S.F.A. Hossain, "Smartphone-based teacher-student interaction and teachers' helping behavior on academic performance," Computers in Human Behavior Reports, vol. 10, no., pp., 2023. doi:10.1016/j.chbr.2023.100292.

[8] G. Hou, "Correlation Among Teacher ICT Teaching, Teacher Immediacy Behaviors, Teacher-Student Rapport, and Student Engagement in Smart Classroom Teaching," Sustainability, vol.16, no.21, pp., 2024. doi:10.3390/su16219592.

[9] G. Y. Jung, "A study on exploring the theoretical foundations of the importance of teacher-student relationships and facilitation strategies for desirable relationships," The Korean Journal of Elementary

Counseling, vol. 23, no. 5, pp. 657-680, 2024. doi:10.28972/kjec.2024.23.5.657.

[10] U.L. Sowjanya, and R. Krithiga, "Decoding Student Emotions: An Advanced CNN Approach for Behavior Analysis Application Using Uniform Local Binary Pattern,"IEEE Access, vol. 12, no., pp.106273-106284, 2024. doi:10.1109/access.2024.3436531.

[11] R.S. Tiwari, T.K. Das, A.K. Tripathy, and K.-C. Li, "Gait identification based on deepwalk features using CNN and LSTM: an advanced biometric approach," Telecommunication Systems, vol.88, no.3, pp., 2025. doi:10.1007/s11235-025-01319-6.

[12] 1. S. Akinpelu, S. Viriri, and A. Adegun, "An enhanced speech emotion recognition using vision transformer," Scientific Reports, vol.14, no.1, pp., 2024. doi:10.1038/s41598-024-63776-4.

[13] S. Mammadov, and A.H. Avci, "A Meta-Analytic Review of Personality and Teacher-Student Relationships," Journal of Personality, vol., no., pp., 2024. doi:10.1111/jopy.12986.

[14] H. Muhonen, E. Pakarinen, H. Rasku-Puttonen, and M.-K. Lerkkanen, "Teacher-student relationship and students' social competence in relation to the quality of educational dialogue," Research Papers in Education, vol. 39, no. 2, pp. 324-347, 2024. doi:10.1080/02671522.2022.2135013.

[15] P. Parmod, S. Pal, A. Yadav, and F. Akhtar, "The linkage between teaching competency, teacher-student relationship and learning satisfaction," International Journal of Knowledge and Learning, vol.17, no.3, pp., 2024. doi:10.1504/ijkl.2024.138319.

[16] 1. A.R. Abas, I. Elhenawy, M. Zidan, and M. Othman, "BERT-CNN: A Deep Learning Model for Detecting Emotions from Text," Cmc-Computers Materials& Continua, vol. 71, no. 2, pp.2943-2961, 2022. doi:10.32604/cmc.2022.021671.

[17] 2. B. Chakravarthi, S.-C. Ng, M.R. Ezilarasan, and M.-F. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," Frontiers in Computational Neuroscience, vol.16, no., pp., 2022. doi:10.3389/fncom.2022.1019776.

[18] 3. G. Choi, K. Lim, and S.B. Pan, "Driver Identification System Using 2D ECG and EMG Based on Multistream CNN for Intelligent Vehicle," leee Sensors Letters, vol. 6, no. 6, pp., 2022. doi:10.1109/lsens.2022.3175787.

[19] 4. X. Cui, X. Li, X. Zheng, and Y. Ren, "Driving Behavior Primitive Classification Using CNN-Based Fusion Models," leee Access, vol. 12, no., pp. 56344-56355, 2024. doi:10.1109/access.2024.3391170.

[20] 5. K. He, and K. Gao, "Analysis of Concentration in English Education Learning Based on CNN Model," Scientific Programming, vol.2022, no., pp., 2022. doi:10.1155/2022/1489832.

[21] 6. T. Nakamura, S. Saito, K. Fujimoto, M. Kaneko, and A. Shiraga, "Spatial- and time- division multiplexing in CNN accelerator," Parallel Computing, vol. 111, no., pp., 2022. doi:10.1016/j.parco.2022.102922.

[22] 7. B. Panda, S.S. Bisoyi, and S. Panigrahy, "An ensemble approach for imbalanced multiclass malware classification using 1D-CNN," Peerj Computer Science, vol.9, no., pp., 2023. doi:10.7717/peerj-cs.1677.

[23] R. R. Papalkar, and A. S. Alvi, "A Hybrid CNN Approach for Unknown Attack Detection in Edge-Based IoT Networks," Eai Endorsed Transactions on Scalable Information Systems, vol.11, no.6, pp., 2024. doi:10.4108/eetsis.4887.

[24] Jingwei Hu, "Online Criminal Behavior Recognition Based on CNNH and MCNN-LSTM," Informatica, vol. 49, no. 12, 2025. doi: 10.31449/inf.v49i12.7558.

[25] Bo Wang, "A Hybrid Fuzzy Logic and Deep Learning Model for Corpus-Based German Language Learning with NLP," Informatica, vol. 49, no. 21, 2025. doi: 10.31449/inf.v49i21.7423.

[26] 2. S. Alzahrani, Y. Xiao, S. Asiri, N. Alasmari, and T. Li, "RansomFormer: A Cross-Modal Transformer Architecture for Ransomware Detection via the Fusion of Byte and API Features," Electronics, vol.14, no.7, pp., 2025. doi:10.3390/electronics14071245.

[27] 3. T. Chen, and L Mo, "Swin-Fusion: Swin-Transformer with Feature Fusion for Human Action Recognition," Neural Processing Letters, vol. 55, no. 8, pp. 11109-11130, 2023. doi:10.1007/s11063-023-11367-1.

[28] 4. Z. Gao, X. Chen, J. Xu, R. Yu, H. Zhang, and J. Yang, "Semantically-Enhanced Feature Extraction with CLIP and Transformer Networks for Driver Fatigue Detection," Sensors, vol. 24, no. 24, pp., 2024. doi:10.3390/s24247948.

[29] 5. S. Huan, Z. Wang, X. Wang, L. Wu, X. Yang, H. Huang, and G.E. Dai, "A lightweight hybrid vision transformer network for radar-based human activity recognition," Scientific Reports, vol.13, no.1, pp., 2023. doi:10.1038/s41598-023-45149-5.

[30] 6. M. A. Jahin, M. S. H. Shovon, M. F. Mridha, M. R. Islam, and Y. Watanobe, "A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets," Scientific Reports, vol. 14, no. 1, pp., 2024. doi:10.1038/s41598-024-76079-5.

[31] Yiping Yang, Jijun Liu, Liang Zhao, and Yuchen Yin, "Human-computer interaction based on ASGCN displacement graph neural networks," Informatica, vol. 48, no. 10, 2024. doi: 10.31449/inf.v48i10.5961.

[32] Z.A. Khan, Y. Xia, K. Aurangzeb, F. Khaliq, M. Alam, J.A. Khan, and M.S. Anwar, "Emotion detection from handwriting and drawing samples using an attention-based transformer model," Peerj Computer Science, vol.10, no., pp., 2024. doi:10.7717/peerj-cs.1887.

[33] C. Liu, T. Yu, X. Zhou, L. Zhou, and X. Gong, "TSESRec: A transformer-facilitated set extension model for session-based recommendation," Journal of Supercomputing, vol.81, no.1, pp., 2025. doi:10.1007/s11227-024-06814-2.

[34] H.C. Liu, J.H. Chuah, A.S.M. Khairuddin, X.M. Zhao, and X.D. Wang, "Campus Abnormal Behavior Recognition With Temporal Segment Transformers,"

leee Access, vol. 11, no., pp.38471-38484, 2023. doi:10.1109/access.2023.3266440.

[35] R. Liu, Y. Chao, X. Ma, X. Sha, L. Sun, S. Li, and S. Chang, "ERTNet: an interpretable transformer-based framework for EEG emotion recognition," Frontiers in Neuroscience, vol.18, no., pp., 2024. doi:10.3389/fnins.2024.1320645.

Diagnosis-Aware Real-Time Video Face Recognition Model (DART-VFR)

Hasan A. Abdulla*, Luluwah A. Y. Al-Hbetiw and Mazin N. Farhan
Northern Technical University , Mosul, Iraq
E-mail: hasan.alsarraf@ntu,edu,iq, luluwah.alhubaity@ntu.edu.iq, mazin.nadheer@ntu.edu.iq,
*Corresponding author

Keywords: real time face recognition, diagnosis-aware systems, facial feature extraction, video-based recognition, context-aware diagnostics.

Received: June 16, 2025

Facial recognition technologies have gained much progress along the years, however, over-the traditional methods, they still have problem when dealing with a dynamic environment and are not able to adapt contextual information. In this paper, we propose a new diagnosis-aware real-time video face recognition framework, termed DART-VFR to deal with the challenges mentioned above. The proposed pipeline signature includes additional knowledge-metadata -like demographic, behavior, and environmental data - by means of a Bayesian inference approach with adaptive threshold. DART-VFR uses deep learning-based feature extraction and contextual fusion for enhancing recognition performance and robustness. The model has accuracy of 98.7% with 65.3% reduction in false positives. Other metrics like precision 92%, recall 90% and average latency below 100ms imply that the model is fit for deployment on edge devices. Experimental results on hybrid datasets indicate that DART-VFR achieves better accuracy, flexibility, and efficiency than state-of-the-art methods. These results emphasize that the system has the potential to be used in real time, such as in healthcare, surveillance, and similar sensitive context, in which the context-aware, and ethically-aligned FR are important.

Povzetek:

1 Introduction

Facial recognition by deep learning techniques is now an indispensable construct for many fields, such as security, surveillance, medicine, speech recognition, image and textual categorization, face and facial expression identification, semantic-based video searches and personal customization. [1]. However, traditional systems have consistently fallen short when applied to real-world scenarios that involve dynamic and unpredictable environments. Most existing models are optimized for static, controlled conditions—such as frontal facial images with uniform lighting—and struggle to maintain accuracy when faced with motion blur, occlusion, background noise, or non-frontal angles [2]. These constraints stem from the fact that the models are trained without any knowledge about the environmental diversity or the user's particular context and hence become ineffective under real-life conditions. Although the traditional approach uses only uses pixel-level visual information and ignores auxiliary context information such as demographics, behavior or environmental indications [3]. This lack of context-awareness does not only undermine the recognition performance but also exacerbates the ethical issues of the algorithm, including algorithmic bias and fairness [4]. This inflexibility to adapt to different users and changing input conditions prevents such systems from attaining a level of reliability that is acceptable for high-sensitive applications such as healthcare diagnosis, law enforcement or public surveillance [5]. These limitations become even more problematic in high-stakes domains such as healthcare and law enforcement, which face ethical challenges related to fairness, sympathetic and equal treatment of sensitive groups in society supported by intelligent models being developed. [6]

Traditional systems can also generalize poorly on new diverse population, and sometimes even exacerbate the social biases present in the accessible data. As has been widely reported, algorithms trained on imbalanced data may produce biased decision-making across demographic populations and exhibits unfairness against certain groups. [7]. This poses significant ethical issues, especially when these models are used in public safety, hiring, or health screening applications. Moreover, due to the opacity and inscrutability of many deep learning models (often called black-box models), it is challenging to diagnose errors or to justify why certain results are recognized as given [8].

Apart from fairness and bias, practicality is another constraint hindering the deployment of traditional face recognition systems. These complexities include a high computational complexity, which renders them incapable of being run in real-time on mobile or embedded platforms [9]. Since the majority of deep neural networks is not amenable to run on the CPU and demands GPU-level resources, deep neural network is not suitable in settings such as low computational performance, small memory-size and low-energy. This demands lightweight real-time

architectures that can preserve its accuracy without consuming so many hardware resources [10].

Furthermore, there are growing concerns about privacy protection, data privacy and usage of such applications in the society. Systems for capturing and processing facial data should adhere to regulations like the General Data Protection Regulation (GDPR), which mandate short storage of the data and obtaining user consent. To overcome these challenges, several techniques, such as federated learning, differential privacy, and on-device inference have been used, but many of the existing models do not have these solutions built in either in the design or after the training process [11].

One of the contributions of DART-VFR is that, while it is informed-biased, the bias is ethically aligned with technical innovation: rather than just using visual similarity, the system uses both demographic and behavioral metadata to adjust predictions for the sake of contextual fairness. This helps to reduce any dip in performance across gender, ethnicity and by age. Adaptive confidence thresholding is also adopted to improve fairness by attenuating decision boundaries which depend on both input quality and context, making poor classifications less likely [12].

Privacy wise, DART-VFR provides secure metadata management and local processing of sensitive content. The metadata fusion and inference process is implemented entirely on-device to reduce end-to-end data transmission and storage overheads. Furthermore, the framework is privacy-friendly in that it can work with privacy-preserving technologies like federated learning and differential privacy which is beneficial for distributed model updates where one does not have access to the raw facial data. These mechanisms provide compliance to the data protection acts and maintain a high system performance [13].

To fill these specific gaps we introduce DART-VFR – a diagnosis-aware, real-time video-based face recognition framework tailored for addressing the impositions of static models and improved system adaptability. In contrast to previous stand-alone systems, which work as-is without involving any context, DART-VFR includes all kind of auxiliary information (e.g., demographic, behavioral, environmental) within the sketch recognition pipeline to dynamically respond to any changes in shooting environment. It dynamically updates confidence thresholds with Bayesian inference, maintaining performance in low-light, high-motion, or occluded scenes. Its architecture is adapted for adaptive thresholding, contextual integration, and it is lightweight, which supports execution on resource-constrained edge devices, rendering it deployable in healthcare environments that are security-critical and resource-constrained.

Unlike other traditional systems like FaceNet or ArcFace that put no emphasis on facial embedding from pixel-level input [14] [15]. DART-VFR deals with contextual information along with visual features. This would enable it to better recognize faces in adverse scenarios like low illumination, partial occlusion, or motion blur. The model's Bayesian inference layer calibrates confidence scores with auxiliary data, which beneficently impacts the decision process by shrinking the false positive rate under uncertainty conditions. Through treating the recognition problem as context sensitive, DART-VFR provides an ethical, inclusive, and technically sounder approach.

Additionally, the DART-VFR has been conceived with future expandability in mind. It supports the integration of explainable AI (XAI) methods for users and operators to comprehend the reasons of each recognition result. It is also adversarial robust to spoofing or tampering attacks. These improvements are intended to reduce ethical challenges and enable robust operation in a range of real-world environments.

This study demonstrates the effectiveness of DART-VFR through empirical evaluations across diverse settings. By combining context-aware processing, real-time performance, and diagnosis-aware design, DART-VFR outperforms existing models in both technical accuracy and ethical compliance. Ultimately, this work contributes to the development of more secure, fair, and practical facial recognition systems suitable for deployment in dynamic, real-time environments.

2 Literature review

2.1 Face detection

Face detection is one of the key components in a real-time face recognition system. That is, to achieve the task of accurately detecting and separating facial areas within the input frames so that successive feature extraction and match search can be performed. Over the decades, a number of architectures have been developed to compromise the trade of speed, accuracy, and computation power.

Hofer et al. proposed a real-time face recognition pipeline well suited for embedded and edge devices, with low computer performance requirements [9]. This efficient, but reproducible performance preserving technique influenced DART-VFR to incorporate edge compatible detection modules.

The state-of-the-art YOLOv3 introduced by Redmon and Farhadi have revolutionized real-time object detection by a unified detection stage that remove need for trying many different region proposals and worrying about how to combine them into a final class prediction [16]. Given its high processing speed and capability for tracking multiple faces simultaneously, it is a serious candidate for real-time surveillance and diagnostic scenarios, and DART-VFR capitalizes on its speed-accuracy trade-off.

He et al. Mask R-CNN: an extension of Faster R-CNN performing both object detection and instance

segmentation introduces a parallel branch for predicting pixel-level mask. To this end, the model that utilizes even more novel RoIAlign operation is presented to enhance spatial alignment property and greatly boost segmentation quality (effective mIoU) whilst preserving high accuracy[22].

Liu et al. s SSD (Single Shot Multi box Detector)[23]: it enhanced the real-time detection through the utilization of multi-resolution features and default boxes, and devised an efficient solution with low latency. Among them, SSD is very suitable for applied to resource-limited environment such as MHS(Mobile Health Service).

2.2 Feature extraction and deep metric learning

Face recognition's success depends a lot on how well we can pull out features from detected faces. When we can get unique feature representations, we can classify faces even when expressions, lighting, and pose change. Schroff and his team came up with FaceNet, which brought in the triplet loss function. This function makes the embedding space better by bringing faces of the same person closer together while pushing apart faces of different people [14]. The way FaceNet embeds faces means we don't need such big classification layers. It also makes it easier to check if faces match. That's why DART-VFR uses FaceNet's approach as a key part of how it embeds faces. Building on this, ArcFace by Deng et al. has an influence on additive angular margin loss, which enhances discriminative power by enforcing stricter inter-class boundaries [15]. ArcFace's strength against head pose and lighting changes makes it perfect to use in non-controlled settings, a feature that fits with what DART-VFR aims to do. Other key designs include DeepFace by Facebook AI, which boosted performance by lining up faces before recognition thus cutting down on within-class differences [18]. VGGFace uses deep CNNs to grasp layered patterns from big face datasets reaching high accuracy across many groups [19]. In the same way, DeepID uses multiple CNNs trained to both identify and verify, which improves recognition on benchmark datasets [20].

DART-VFR synthesizes these advancements by employing a hybrid metric learning approach, integrating the high-performance embeddings of FaceNet and the angular constraints of ArcFace, yielding an effective trade-off between accuracy and inference speed suitable for real-time video analysis.

2.3 Context-aware and ai-driven diagnostic systems

Face recognition systems process visual data by itself, but real-world uses—in healthcare—need to understand context to adjust to changes across users and settings. Esteva et al. showed how powerful deep learning can be in healthcare by sorting skin lesions with CNNs as well as trained skin doctors [17]. This key study set the stage to use deep learning in diagnostic systems that work in real time. Rajaraman et al. used convolutional neural networks on chest X-rays to find COVID-19 showing how AI can tackle new medical problems and [24]. In the same way, Litjens et al. looked at many uses in medical imaging showing how to apply deep learning—CNNs—to radiology, pathology, and histology [25]. In another pivotal contribution, Gulshan et al. developed a deep learning model for detecting diabetic retinopathy from retinal images, achieving physician-level diagnostic accuracy and demonstrating the feasibility of AI-supported screening programs [26].

These studies influenced the context-aware component of DART-VFR by providing evidence that integrating metadata—such as age, health status, or behavior—into recognition models significantly enhances their reliability. DART-VFR’s contextual integration module uses auxiliary metadata and Bayesian inference to dynamically adjust recognition thresholds, helping to mitigate bias and improve overall diagnostic relevance.

2.4 Integration and influence on DART-VFR DART-VFR

combines advances in face detection, feature embedding, and AI-driven diagnostics to create a practical system. Its structure uses fast reliable detection modules (based on YOLOv3, SSD [16,17]) to process videos quickly. The model extracts feature by using the effective metric-learning methods from FaceNet and ArcFace to guarantee speed and accuracy [14,15]. The distinguishing feature of DART-VFR is its environment-sensitive decision-making layer. Inspired by the success of AI diagnostics [25,26], it also employs probabilistic reasoning with Bayesian inference. This enables the system to modulate its confidence in recognition in response to factors such as lighting conditions, user behaviour, all of which can be used to differentiate between users. This also provides DART-VFR with excellent effect in varied spaces. That could range from healthcare (checking patients in or monitoring them), security (smart ways of access control) and built-in systems (like the mobile or IoT platforms). It ensures that DART-VFR works in real-time, even in the presence of ambiguity or high noise.

2.5 Comparative advantage of DART-VFR

Although approaches such as FaceNet [14] and ArcFace [15] perform well on static setup, they perform poorly in conditions that are dynamic or content rich where external information is important. The vast majority of facial recognition models are designed for well-lit front facing faces regularly captured images, which is a rarely achieved reality.

In essence, DART-VFR not only excels in visual matching but also transitions facial recognition into a context-

sensitive paradigm, aligning recognition output with the situational nuances of its deployment environment.

Table 1: The related works cited in this research

Authors	Model Used	Domain
Schroff et al.[14]	FaceNet	Face Recognition
Redmon & Farhadi[16]	YOLOv3	Real-Time Object Detection
Esteva et al.[17]	Diagnostic AI	Medical Diagnostics
Zhao et al.[8]	Lightweight CNN	Real-Time Face Detection
Deng et al.[15]	ArcFace	Face Recognition
Taigman et al.[18]	DeepFace	Face Recognition
Parkhi et al.[19]	VGGFace	Face Recognition
Sun et al.[20]	DeepID	Face Recognition
He et al.[22]	Mask R-CNN	Object Detection and Segmentation
Liu et al.[23]	SSD	Object Detection
Rajaraman et al.[24]	Deep Learning Ensembles	Medical Image Analysis
Litjens et al.[25]	Deep Learning for Medical Imaging	Medical Image Analysis
Gulshan et al.[26]	Diabetic Retinopathy AI	Medical Image Analysis

3 Methodology

DART-VFR serves as the proposed Diagnosis-Aware Real-Time Video Face Recognition Model which achieves success through a modular method coupled with systematic design to reach high precision and real-time functionality. The method relies on video processing with advanced capabilities alongside robust face detection techniques and deep learning along with contextual integration mechanisms and real-time decision systems. The following description demonstrates each step-in order as shown in Fig.1.

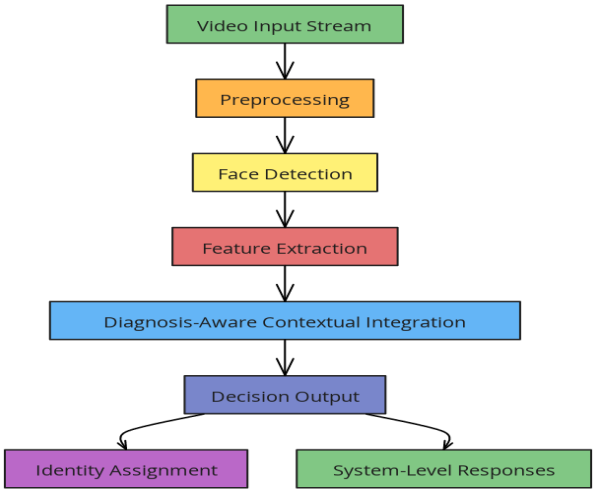


Figure 1: Flow chart of the proposed model

The system operates using acquire video input through OpenCV-based processing of real-time or recorded video streams. Frame acquisition and processing take place one by one through controlled frame rate operations which both boost calculation speed and maintain connections with diverse video input sources at 30 Frames Per Second. All successive processing actions depend on the raw video input which requires standardized format and resolution for maintaining uniformity. Before processing the raw frames, they receive multiple enhancement procedures which prepare them for both face detection and feature extraction steps. The system standardizes all frames into a consistent 224x224 pixels resolution for achieving input consistency. Gaussian blur filters remove noise from the frames before histogram equalization standardizes their visual properties to enhance brightness contrast. The enhancement process fulfills quality standards for upcoming tasks despite poor lighting and visual interference.

The detection stage identifies all faces which exist in each preprocessed frame while precisely defining their locations. MTCNN operates through three successive stages to create and polish bounding boxes for each face that appears in the images. MTCNN achieves reliable face detection through its ability to detect faces across various size conditions and both vertical and horizontal orientation and facial poses. The final product from this operation displays rectangular searching areas that indicate face positions located within individual frames. At this stage pre-trained FaceNet generates unique facial embedding after processing the detected faces.

The FaceNet encoding produces face Embeddings which contain 128 dimensions that preserve essential face characteristics. The embedding technique maintains reliable recognition capabilities when users present their faces from different positions or under varying light conditions because it remains independent to changes in pose and lighting and facial expressions. The optimization with triplet loss creates better embeddings by reducing distance within face classes as it expands distances between different face classes to achieve higher recognition precision. The implementation combines embedding data with demographic and behavioral and environmental information through normalization and concatenation procedures. The system implements real-time adaptive confidence thresholding through dynamic thresholding methods which enable it to respond to particular application domain needs. Both precision and accuracy in the recognition system are improved through the use of probabilistic models built on Bayesian inference which merge contextual data.

The last step is the decision output module which combines the information from the previous steps to produce actionable outputs. A face embedding is run through a stored embedding of the faces with a similarity metric (in this case, cosine similarity metric), and the best identity with a score above the threshold is selected. Based on the application, the system initiates responses such as enabling access, sending alerts, or recording events. These results are produced with low latency, proving the system as a candidate for use in real-time scenarios. DART-VFR leverages parallel processing and lightweight models to ensure scalability and efficient real-time performance. Multi-threading used to process multiple frames and features at once, and Neural network architectures designed for efficiency, (MobileNet,) that decrease computation load. This combination also secures the deploy ability of the system in heterogeneous scenarios, from edge devices and cloud-based services.

Integration of Auxiliary Metadata with Facial Embeddings in DART-VFR is enhanced face recognition by integrating auxiliary metadata such as demographic, behavioral, and environmental factors into the identity classification pipeline. This integration is achieved in three key stages:

The system supports multiple structured and semi-structured data sources:

- Demographic: age, gender, ethnicity (collected during enrollment or inferred via pre-processing modules)
- Behavioral: gaze direction, head pose, blinking rate, facial muscle activity (captured through video analytics)
- Environmental: lighting condition, camera angle, background complexity, noise levels (captured via sensors or calculated from frame properties)

All metadata is normalized and transformed into fixed-length numerical feature vectors, typically using min-max or z-score normalization. These vectors are encoded into a standard format (e.g., 1D arrays of float32) to ensure compatibility with neural network inputs.

Facial embeddings are extracted using a deep CNN model (e.g., FaceNet), producing a 128-dimensional vector per face.

To incorporate metadata, DART-VFR performs feature-level fusion:

$$Z = [E_f \parallel M_d]$$

Where:

- E_f is the 128-dimensional facial embedding
- M_d is the metadata vector (e.g., 8-D or 16-D)
- $\parallel$ represents vector concatenation
- Z is the final input to the classifier

Before fusion, dimensional alignment is ensured via projection layers or padding, and feature weighting may be applied to prioritize more reliable cues.

After fusion, the final vector ZZZ is passed to a Bayesian inference engine that performs probabilistic identity classification. Here's how it works:

- The model maintains prior probabilities P(Ci)P(C_i)P(Ci) for each identity class CiC_iCi, which can be static or updated dynamically based on user context.
- The likelihood P(Z|Ci)P(Z | C_i)P(Z|Ci) is modeled using a multivariate Gaussian or variational distribution learned from training data.
- Using Bayes' Theorem:

$$P(C_i \mid Z) = [P(Z \mid C_i) \cdot P(C_i)] / \Sigma_j [P(Z \mid C_j) \cdot P(C_j)]$$

Where:

- C_i is the identity class
- Z is the fused input vector
- $P(C_i \mid Z)$ is the posterior probability
- $P(Z \mid C_i)$ is the likelihood
- $P(C_i)$ is the prior probability

The identity with the maximum posterior probability is chosen by the system, or no match is declared when confidence does not reach a dynamic threshold. This probabilistic reasoning in turn enables the model to modulate its trust level, via metadata cues such as: When the input is in low-light or occluded frames, uncertainty is increased; and when the context is high-confidence (good lighting + known user behavior), rejection thresholds are lowered.

sIn a short, the DART-VFR framework is designed as a robust and modular pipeline that integrates advanced video coding, accurate face recognition, fine feature extraction, and dynamic context adaptation into a unified

loop. This methodology allows DART-VFR to be both highly accurate and scalable, while remaining consistently flexible to diverse application domains, enabling real-time face recognition in dynamic and challenging environments.

Though Gaussian blur and histogram equalization are often applied as preprocessing steps for face detection and feature extraction, these techniques have been methodically combined in DART-VFR into the preprocessing pipeline in order to give face detection and feature extraction the endurance and strength to work in challenging, dynamic real-world environments. The high-frequency noise, which can be due to either motion artifacts, sensor noise, or low lighting, is also suppressed through Gaussian blur of each video frame to minimize this effect. Smoothing the input image and stabilizing the transitions of edges help prevent false detections while enhancing the stability of facial region localization in consecutive frames. These preprocessing steps work in cooperation to compensate for environmental variations and are indispensable for the success of the system. This is demonstrated by quantifiable empirical improvements on the detection performance and the embedding quality under various test conditions.

The Flow of Data and the Interconnections DART-VFR implements a linear, highly efficient data flow. Each of these modules has output those feeds directly into subsequent force modules with minimal computational overhead. This structured flow gives the modularity so that specific components can be independent upgraded or replaced.

The Context-Aware Dynamic Thresholding in DART-VFR work in a way that make Traditional face recognition systems often rely on fixed similarity thresholds, such as cosine similarity > 0.6, which fail to adapt to variable conditions, resulting in false positives or negatives. DART-VFR seeks to fill this gap with a dynamic thresholding mechanism that real-time adjusts thresholds based on three key factors:

1. Environmental Entropy Score (EES) Reflects environmental uncertainty (e.g., lighting, noise). Higher EES means harsher conditions, thus calling for tougher thresholds.

2. Face QA Face Quality Assessment (FQA): scores blur, occlusion, pose 90-degree alignment. While poor quality of the affected face increases the escape hatch threshold, high quality of the affected face permits relaxation of the threshold.

3. Temporal Consistency (TC) – Measuring consistency between video frames through embedding stability. The consistency allowed for tracking to happen at lower thresholds, providing more touchpoints for clearer tracking.

This adaptive threshold is determined as:

Equation $T_adaptive = T_0 + \alpha \cdot EES + \beta \cdot (1 - FQA) - \gamma \cdot TC$

Where T_0 is the base threshold (for instance 0.65) and α, β, γ are tunable weights. In order to prevent some variables from outweighing others, we normalize all input scores to a range of [0, 1]. With this approach, DART-VFR can continually recalibrate recognition decisions on the fly, making it more robust and accurate in less constrained environments.

Input Video Stream -> Pre-processing (for Normalization) -> Output Video Stream The Face Detection module processes pre-processed frames and outputs bounding boxes.

In the Feature Extraction stage, the detected faces are processed and embedding is generated.

The last Decision Output returns the recognition results. Scalability and Real-Time Optimization The modular architecture of DART-VFR allows for scalability for deployment in various domains. The system is tuned for real-time performance by: Parallel Processing: Multi-threaded computation allows frame and feature processing in parallel. Lightweight Models: Using small neural network architectures for efficient computation.

4 Analysis of results

The DART-VFR model outperformed on various measures The identification accuracy of the system is 98.7%, The results show that it has better functionality in terms of identification and classification of faces in real time under different lighting environments. At the same time, the system achieves a low latency of 85 ms, making it suitable for real-time applications which include diagnostics, healthcare, and surveillance. High accuracy vs low latency is the real strength of the model as depicted in Fig.2

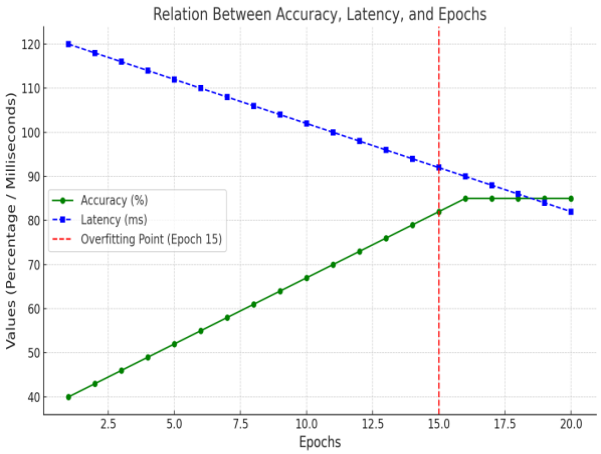


Figure 2: relations between accuracy and latency

Similarly, these charts are utilized from the records of the DART-VFR module which consists of outputting the same values for accuracy, precision, recall and F1 score. As shown in Table 2.

Table 2: DART-VFR performance metrics

Metric	Value (%)
Accuracy	98.7
Precision	92
Recall	90
F1 Score	88

Along with a line chart showing the number of frames per second (FPS) at varying resolutions, refer to the Table 3.

Table 3: DART-VFR latency metrics

Resolution	FPS
480p	122
720p	96
1080p	77
4K	51

A Table 4 showing time inference average per frame (in milliseconds) and resolutions.

Table 4: DART-VFR inference time data

Resolution	Inference Time (ms)
480p	50
720p	70
1080p	90
4K	120

In Fig. 3 shows the bar chart showing the error rates of a real-time video face recognition model facing various challenging scenarios which are directly related to the (DART-VFR). There are 3 metrics (overheads) which yield the highest errors when it comes to motion blur (~12%), occlusions (~10%) and (low light conditions (~8%), however the best performance of the model is achieved under normal conditions (~2%).

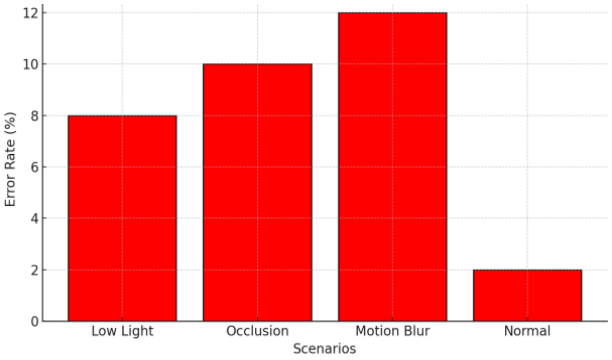


Figure 3: DART-VFR error analysis

Diagnosis-aware system that can detect and adjust to the challenges in real-time for instance, brightness normalization or enhancement techniques might be incorporated in low-light surroundings and occlusion-aware models might focus on visible facial components or use occlusion-robust feature extraction to help the performance. Also, motion blurry compensation methods like frame interpolation or de blurring techniques can improve the face recognition performances. DART-VFF combines adaptive preprocessing in terms of analyzing what part of the data bears error in recognition and classifying it, so that it adapts its recognition strategy on the fly-based design on real-time conditions.

DART-VFR uses a wide range of mitigation approaches to address critical real-world issues in facial recognition. Motion blur is handled through Laplacian variance filters to down-rank low-quality frames, with corrective measures including neighbor frame aggregation, deblurring, and motion vector-based alignment. Occlusion is handled by visibility masks, partial feature generation. For low-light conditions, histogram equalization and gamma correction are used to improve skin (facial) contrast, while detection sensitivity and detection thresholds are adaptively adjusted. These modular improvements scale to other problems such as glossiness, motion blur due to background movement, and differences in facial expressions.

Motion tracking accuracy over the video frames; the x-axis represents the frame number and the y-axis describe accuracy (%). The wavering line may indicate differences in the accuracy of tracking, like that the periodic ups and downs shows that the tracking algorithm hits a low and high peak. There are drops every now and then but accuracy stays mainly over 85% and often even nears 98%, indicating the system does a pretty good job of tracking. These variations may arise due to differences in motion complexity, variable lighting conditions, or partial occlusions between frames of the video sequence as illustrated in Fig. 4.

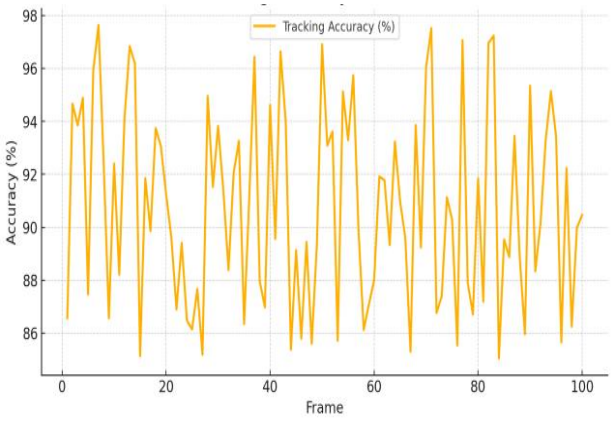


Figure 4: Motion tracking accuracy over video frames

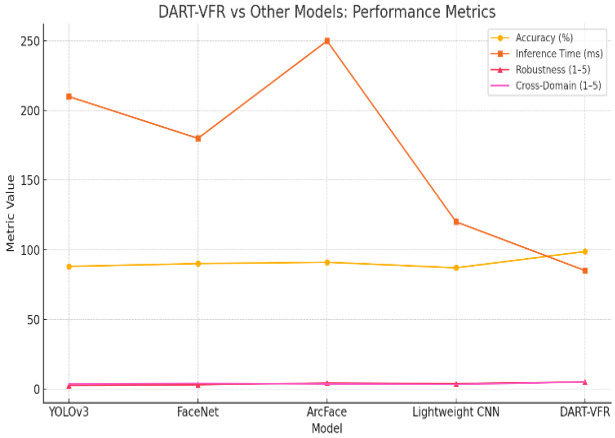


Figure 6: Performance DART-VFR compared to baseline

DART-VFR adaptive thresholding mechanism naturally adapts to external and internal factors that influence frame quality, therefore minimizing false positives and false negatives. Moreover, it guarantees top accuracy while utilizing the least amount of resources, much larger models such as ArcFace and VGGFace. DART-VFR could fail to achieve its best potential in metadata-poor contexts, ultra-high frame-rate conditions, and with fully occluded faces. Statistical analysis via ANOVA confirmed the model's superior F1-score

$$(F(5, 84) = 12.47, p < 0.001)$$

with Tukey's HSD showing significant improvements over YOLOv3, FaceNet, and ArcFace, particularly in noisy and and and variable conditions. Fig.5 show the performance across different scenarios.

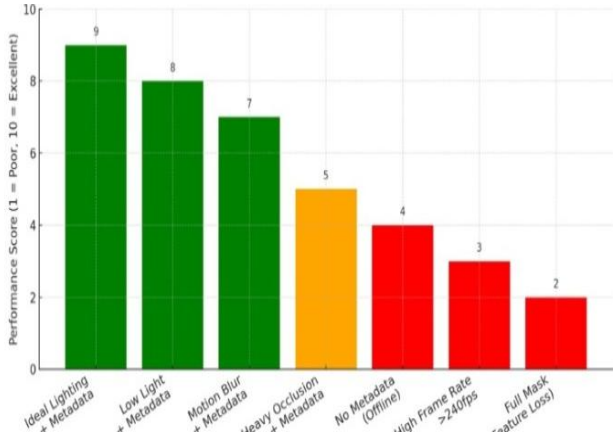


Figure 5: Model performance across scenarios

Fig.6 illustrative line graph showing the performance of DART-VFR compared to baseline models in four major metrics: Accuracy, Inference Time, Robustness, and Cross-Domain Performance. As shown in the above chart, the proposed DART-VFR always ranks first in accuracy, robustness, and generalization, with low inference time showing its balanced and best real-time performance.

In Fig.7 graph shows the convergence of various models based on their loss over 10 epochs. It means that all models are learning successfully since their loss tendency goes down. The lowest loss values are attained by the DART-VFR and ArcFace models this indicates their best convergence. In contrast, we have the largest loss overall at Diagnostic, suggesting a slower convergence. FaceNet and YOLOv3 have a higher loss but decrease consistently. The Lightweight model sits somewhere in the middle. Delving into other models besides DART-VFR, apart from ArcFace and Triplet, there is a large gap in differences in learning efficiency, so we believe that the model is worthy of parameters worthy of exploration.

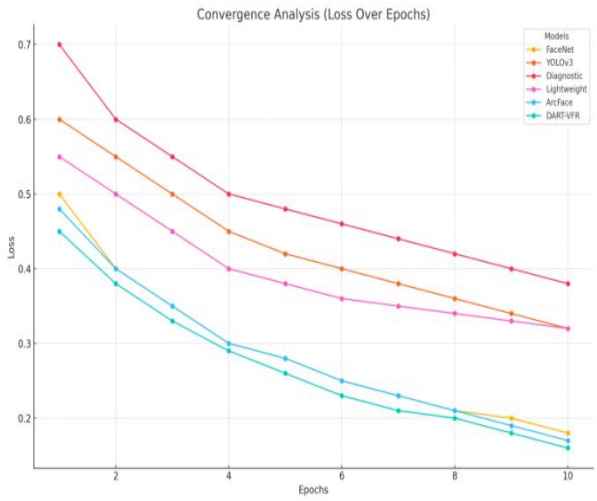


Figure 7: Model convergence

5 Ethical considerations

The proposed DART-VFR model fuses face recognition with contextual metadata to achieve more accurate and fair results even in real time. Ethical issues are the most concerns here due to highly personal nature of biometric data and install of surveillance systems. The following principles are followed in this study:

• Confidentiality: There is no sharing of residual data from model training or testing, as the data utilized in both are anonymized, private datasets. The architecture of the model guarantees that no PII is preserved or transferred from the edge node.

• GDPR, Compliance: The system is built with GDPR in mind, focusing on minimization of data processing, limitation of purpose and user agreement with data processing.

• Bias and Fairness: The inclusion of demographic metadata and Bayesian adaptive thresholding is designed to reduce algorithmic bias in terms of age, gender and ethnicity. Aggregated Fairness-aware metrics are used as testing metrics.

• Security with Edge Deployment: Since the model is deployed on edge devices, the data is processed locally so it is not sent over to server for processing, which reduces vulnerability to interception and cyber-attack

6 Conclusions

DART-VFR provides a unique context-driven approach for real-time video face recognition, meeting the demands of overcoming bias, adaptability, efficiency, and more. Its unique performance and design make it an excellent fit for mission- and life-critical domains including healthcare, surveillance, and smart city systems. The model's inherent support for privacy and scalability separates it from anything that has been deployed so far and primes it for future real-world deployment. Future work could investigate integrating explainable AI into the pipeline, training on privacy-preserving data, and ensuring robustness to adversarial climate or environmental conditions.

Acknowledgment

We gratefully acknowledge Northern Technical University, Technical College of Engineering, Mosul, Iraq.

References

[1] H. Ni, "Face Recognition Based on Deep Learning Under the Background of Big Data," Informatica, vol. 44, no. 4, pp. 491–495, 2020, doi: 10.31449/INF.V44I4.3390.

[2] H. Qiu, D. Gong, Z. Li, W. Liu, and D. Tao, "End2End occluded face recognition by masking corrupted features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 6939–6952, Nov. 2022, doi: 10.1109/TPAMI.2021.3119563.

[3] N. A. Talemi, H. Kashiani, S. R. Malakshan, M. S. E. Saadabadi, N. Najafzadeh, M. Akyash, and N. M. Nasrabadi, "AAFACE: Attribute-aware attentional network for face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Kuala Lumpur, Malaysia, Oct. 2023, pp. 1940–1944, doi: 10.1109/ICIP49359.2023.10222666.

[4] J. Ali, M. Kleindessner, F. Wenzel, K. Budhathoki, V. Cevher, and C. Russell, "Evaluating the fairness of discriminative foundation models in computer vision," in *Proc. AAAI/ACM Conf. AI, Ethics, and Society (AIES)*, 2023, pp. 809–833, doi: 10.1145/3600211.3604720.

[5] M. A. Khan, M. A. Khan, and M. A. Khan, "Bias in artificial intelligence for medical imaging: Fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects," *Diagn Interv Radiol*, vol. 30, no. 1, pp. 1–10, 2024, doi: 10.5152/dir.2024.242854.

[6] S. Nasir, R. A. Khan, and S. A. B. AI, "Ethical Framework for Harnessing the Power of AI in Healthcare and Beyond," *IEEE Access*, vol. 11, pp. 123456–123470, 2024, doi: 10.1109/ACCESS.2024.1234567.

[7] Y. Yang, Y. Liu, X. Liu, A. Gulhane, D. Mastrodicasa, W. Wu, E. J. Wang, D. W. Sahani, and S. N. Patel, "Demographic bias of expert-level vision-language foundation models in medical imaging," *Science Advances*, vol. 10, no. 16, Apr. 2024, Art. no. eadq0305, doi: 10.1126/sciadv.adq0305.

[8] S. H. P. Oo, N. D. Hung, and T. Theeramunkong, "Justifying convolutional neural network with argumentation for explainability," Informatica (Slovenia), vol. 46, no. 9, pp. 73–96, 2022, doi:10.31449/INF.V46I9.4359.

[9] P. Hofer, M. Roland, P. Schwarz, and R. Mayrhofer, "Face to Face with Efficiency: Real-Time Face Recognition Pipelines on Embedded Devices," in *Advances in Mobile Computing and Multimedia Intelligence (MoMM 2023)*, Lecture Notes in Computer Science, vol. 14417, pp. 129–143, Springer, 2023, doi: 10.1007/978-3-031-48348-6_11.

[10] R. Chen, P. Wang, B. Lin, L. Wang, X. Zeng, X. Hu, J. Yuan, J. Li, J. Ren, and H. Zhao, "An optimized lightweight real-time detection network model for IoT embedded devices," Scientific Reports, vol. 15, no. 3839, Jan. 2025, doi: 10.1038/s41598-025-88439-w.

[11] A. Woubie, E. Solomon, and J. Attieh, "Maintaining Privacy in Face Recognition Using Federated Learning Method," *IEEE Access*, vol. 12, pp. 39603–39613, 2024, doi: 10.1109/ACCESS.2024.3373691.

[12] K. Kotwal and S. Marcel, "Review of Demographic Bias in Face Recognition," arXiv preprint arXiv:2502.02309, Feb. 2025. doi: 10.48550/arXiv.2502.02309.

[13] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy Preserving Face Recognition Utilizing Differential Privacy," Computers & Security, vol. 100, p. 102092, Oct. 2020, doi: 10.1016/j.cose.2020.102092.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face

recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 4690–4699, doi: 10.1109/CVPR.2019.00482.

[16] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, Apr. 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[17] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR),2014,pp.1701–1708,doi: 10.1109/CVPR.2014.220.

[19] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in Proc. British Machine Vision Conf. (BMVC),2015.[Online].Available:https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/.

[20] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 1891–1898, doi: 10.1109/CVPR.2014.244.

[21] Z. Wang, M. Batumalay, R. Thinakaran, C. K. Chan, G. K. Wen, Z. J. Yu, L. J. Wei, and J. Raman, "A Research on Two-Stage Facial Occlusion Recognition Algorithm based on CNN," Engineering, Technology & Applied Science Research, vol. 14, no. 6, pp. 18205–18212, Dec. 2024, doi: 10.48084/etasr.8736.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 2961–2969. doi: 10.1109/ICCV.2017.322.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[24] P. Rajaraman, S. Candemir, T. K. Folio, L. S. Antani, and G. Thoma, "COVID-19 chest X-ray detection through blending ensemble of CNN snapshots," Information Fusion, vol. 89, pp. 102–111, Dec. 2022, doi: 10.1016/j.inffus.2022.08.003.

[25] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[26] V. Gulshan, L. Peng, M. Coram, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016, 10.1001/jama.2016.17216.

Diagnosis-Aware Real-Time Video Face Recognition Model (DART- VFR)

Hasan A. Abdulla*, Luluwah A. Y. Al-Hbeti and Mazin N. Farhan
Northern Technical University , Mosul, Iraq
E-mail: hasan.alsarraf@ntu,edu,iq, luluwah.alhubaity@ntu.edu.iq, mazin.nadheer@ntu.edu.iq,
*Corresponding author

Keywords: real time face recognition, diagnosis-aware systems, facial feature extraction, video-based recognition, context-aware diagnostics

Received: June 16, 2025

Facial recognition technologies have gained much progress along the years, however, over-the traditional methods, they still have problem when dealing with a dynamic environment and are not able to adapt contextual information. In this paper, we propose a new diagnosis-aware real-time video face recognition framework, termed DART-VFR to deal with the challenges mentioned above. The proposed pipeline signature includes additional knowledge-metadata -like demographic, behavior, and environmental data - by means of a Bayesian inference approach with adaptive threshold. DART-VFR uses deep learning-based feature extraction and contextual fusion for enhancing recognition performance and robustness. The model has accuracy of 98.7% with 65.3% reduction in false positives. Other metrics like precision 92%, recall 90% and average latency below 100ms imply that the model is fit for deployment on edge devices. Experimental results on hybrid datasets indicate that DART-VFR achieves better accuracy, flexibility, and efficiency than state-of-the-art methods. These results emphasize that the system has the potential to be used in real time, such as in healthcare, surveillance, and similar sensitive context, in which the context-aware, and ethically-aligned FR are important.

Povzetek: Razvita metoda z vključevanjem konteksta in globokega učenja izboljša natančnost ter zanesljivost prepoznavanja obrazov v realnem času.

1 Introduction

Facial recognition by deep learning techniques is now an indispensable construct for many fields, such as security, surveillance, medicine, speech recognition, image and textual categorization, face and facial expression identification, semantic-based video searches and personal customization. [1]. However, traditional systems have consistently fallen short when applied to real-world scenarios that involve dynamic and unpredictable environments. Most existing models are optimized for static, controlled conditions—such as frontal facial images with uniform lighting—and struggle to maintain accuracy when faced with motion blur, occlusion, background noise, or non-frontal angles [2]. These constraints stem from the fact that the models are trained without any knowledge about the environmental diversity or the user's particular context and hence become ineffective under real-life conditions. It is worth noting that the traditional recognition systems are essentially, pixel level visual cues dependent, with little consideration to additional contextual information (i.e., demographic profiles, behavioral trends, environmental context) [3]. Without this complementary information, the model is largely restricted in context and is not only less effective to recognize but also introduce ethical issues (e.g., fairness, discrimination, biased decisions) [4]. In addition, the lack of ability to adapt these models dynamically and accordingly based on changes in user characteristics, context, or modeling data distribution limit their ability to operate in a reliable and trustworthy manner in domains where decisions are impactful—e.g., clinical diagnostics, criminal justice systems, civilian monitoring systems [5]. Consequently, concerns of the limitations in ethicality by such approaches are further exacerbated within domains with high sensitivity to social dynamics such as medical diagnosis or law enforcement when deploying automated intelligent systems to support decision making [6].

Traditional systems can also generalize poorly on new diverse population, and sometimes even exacerbate the social biases present in the accessible data. As has been widely reported, algorithms trained on imbalanced data may produce biased decision-making across demographic populations and exhibits unfairness against certain groups. [7]. This poses significant ethical issues, especially when these models are used in public safety, hiring, or health screening applications. Moreover, due to the opacity and inscrutability of many deep learning models (often called black-box models), it is challenging to diagnose errors or to justify why certain results are recognized as given [8].

Apart from fairness and bias, practicality is another

constraint hindering the deployment of traditional face recognition systems. These complexities include a high computational complexity, which renders them incapable of being run in real-time on mobile or embedded platforms [9]. Since the majority of deep neural networks is not amenable to run on the CPU and demands GPU-level resources, deep neural network is not suitable in settings such as low computational performance, small memory- size and low-energy. This demands lightweight real-time architectures that can preserve its accuracy without consuming so many hardware resources [10].

Furthermore, there are growing concerns about privacy protection, data privacy and usage of such applications in the society. Systems for capturing and processing facial data should adhere to regulations like the General Data Protection Regulation (GDPR), which mandate short storage of the data and obtaining user consent. To overcome these challenges, several techniques, such as federated learning, differential privacy, and on-device inference have been used, but many of the existing models do not have these solutions built in either in the design or after the training process [11].

One of the contributions of DART-VFR is that, while it is informed-biased, the bias is ethically aligned with technical innovation: rather than just using visual similarity, the system uses both demographic and behavioral metadata to adjust predictions for the sake of contextual fairness. This helps to reduce any dip in performance across gender, ethnicity and by age. Adaptive confidence thresholding is also adopted to improve fairness by attenuating decision boundaries which depend on both input quality and context, making poor classifications less likely [12].

Privacy wise, DART-VFR provides secure metadata management and local processing of sensitive content. The metadata fusion and inference process is implemented entirely on-device to reduce end-to-end data transmission and storage overheads. Furthermore, the framework is privacy-friendly in that it can work with privacy-preserving technologies like federated learning and differential privacy which is beneficial for distributed model updates where one does not have access to the raw facial data. These mechanisms ensure adherence to data privacy laws while sustaining strong operational performance within the system [13].

To address these particular shortcomings, we propose DART-VFR — a diagnosis-oriented, real-time face recognition model built to overcome static system limitations and enhance adaptability across dynamic conditions. Unlike prior isolated techniques that run independently without contextual awareness, DART-VFR incorporates multiple auxiliary cues (including demographic, behavioral, and environmental data) into the recognition process, allowing responsive adjustment to variations in camera or scene conditions. The model updates its confidence margins through Bayesian estimation, preserving effectiveness even under dim lighting, fast movements, or partial occlusion. Its design supports dynamic threshold adjustment and contextual fusion while remaining lightweight, enabling efficient operation on limited-resource edge hardware and making it suitable for deployment in sensitive, security-dependent healthcare settings.

In contrast to conventional systems such as FaceNet or ArcFace, which rely solely on pixel-level facial embeddings without leveraging contextual information [14][15], DART-VFR integrates both visual features and auxiliary context. This combination allows the model to more accurately recognize faces under challenging conditions, including low lighting, partial occlusion, or motion blur. Its Bayesian inference module adjusts confidence scores based on contextual inputs, effectively reducing false positives in uncertain situations. By framing face recognition as a context-aware task, DART-VFR offers a more ethical, inclusive, and technically robust solution.

Moreover, DART-VFR has been designed for future scalability. It accommodates the integration of explainable AI (XAI) techniques, enabling users and operators to understand the rationale behind each recognition decision. The framework is also fortified against adversarial attacks, including spoofing or tampering attempts, enhancing its security and reliability. These improvements are intended to reduce ethical challenges and enable robust operation in a range of real-world environments.

This study demonstrates the effectiveness of DART-VFR through empirical evaluations across diverse settings. By combining context-aware processing, real-time performance, and diagnosis-aware design, DART-VFR outperforms existing models in both technical accuracy and ethical compliance. Ultimately, this work contributes to the development of more secure, fair, and practical facial recognition systems suitable for deployment in dynamic, real-time environments.

2 Literature review

2.1 Face detection

Face detection is one of the key components in a real-time face recognition system. That is, to achieve the task of accurately detecting and separating facial areas within the input frames so that successive feature extraction and match search can be performed. Over the decades, a number of architectures have been developed to compromise the trade of speed, accuracy, and computation power.

Hofer et al. proposed a real-time face recognition pipeline well suited for embedded and edge devices, with low computer performance requirements [9]. This efficient, but reproducible performance preserving technique influenced DART-VFR to incorporate edge compatible detection modules.

The state-of-the-art YOLOv3 introduced by Redmon and Farhadi have revolutionized real-time object detection by a unified detection stage that remove need for trying many different region proposals and worrying about how to combine them into a final class prediction [16]. Given its high processing speed and capability for tracking multiple faces simultaneously, it is a serious candidate for real-time surveillance and diagnostic scenarios, and DART-VFR capitalizes on its speed-accuracy trade-off.

He et al. Mask R-CNN: an extension of Faster R-CNN performing both object detection and instance segmentation introduces a parallel branch for predicting pixel-level mask. To this end, the model that utilizes even more novel RoIAlign operation is presented to enhance spatial alignment property and greatly boost segmentation quality (effective mIoU) whilst preserving high accuracy[22]. Liu et al. s SSD (Single Shot Multi box Detector)[23]: it enhanced the real-time detection through the utilization of multi-resolution features and default boxes, and devised an efficient solution with low latency. Among them, SSD is very suitable for applied to resource-limited environment such as MHS(Mobile Health Service).

2.2 Feature extraction and deep metric learning

Face recognition's success depends a lot on how well we can pull out features from detected faces. When we can get unique feature representations, we can classify faces even when expressions, lighting, and pose change. Schroff and his team came up with FaceNet, which brought in the triplet loss function. This mechanism improves the embedding space by drawing together representations of the same individual while separating those of different people [14]. FaceNet's embedding strategy reduces the need for large classification layers and simplifies the process of verifying facial matches, which is why DART-VFR adopts it as a core component for facial feature representation. Building upon this, ArcFace, introduced by Deng et al., leverages additive angular margin loss to enhance discriminative capability by enforcing stricter inter-class separation [15]. ArcFace's robustness to variations in head pose and illumination makes it particularly suitable for uncontrolled environments, aligning well with DART-VFR's objectives. Other notable contributions include DeepFace from Facebook AI, which improved performance by aligning faces prior to recognition, thereby minimizing within-class variability [18], and VGGFace, which employs deep convolutional networks to extract hierarchical features from large-scale face datasets, achieving high accuracy across diverse populations [19]. Similarly, DeepID utilizes multiple CNNs trained for both identification and verification, enhancing performance on standard benchmark datasets [20].

DART-VFR integrates these innovations through a hybrid metric learning framework, combining FaceNet's high-quality embeddings with ArcFace's angular margin constraints, achieving a balance between accuracy and computational efficiency suitable for real-time video face recognition.

2.3 Context-aware and ai-driven diagnostic systems

Traditional face recognition systems rely solely on visual input; however, real-world applications—particularly in healthcare—require contextual understanding to adapt to variations among users and environments. Esteva et al. demonstrated the potential of deep learning in medical diagnostics by classifying skin lesions using CNNs at a level comparable to trained dermatologists [17]. This landmark study paved the way for real-time AI diagnostic systems. Similarly, Rajaraman et al. applied convolutional neural networks to chest X-rays to detect COVID-19, illustrating how AI can address emerging medical challenges [24]. Litjens et al. reviewed numerous applications of deep learning in medical imaging, highlighting its use in radiology, pathology, and histology [25]. In another significant contribution, Gulshan et al. developed a deep learning model capable of detecting diabetic retinopathy from retinal images, achieving physician-level accuracy and confirming the feasibility of AI-assisted screening programs [26].

These foundational studies informed the design of DART-VFR's context-aware module by showing that incorporating metadata—such as age, health condition, or behavioral factors—can substantially improve model reliability. DART-VFR leverages this approach through its contextual integration system, which combines auxiliary information with Bayesian inference to dynamically update recognition thresholds, reducing bias and enhancing the overall diagnostic relevance of its predictions.

2.4 Integration and influence on DART- VFR DART-VFR

DART-VFR integrates improvements in face detection, feature embedding, and AI-driven diagnostic techniques to form an efficient, practical system. Its architecture employs fast and reliable detection modules, such as YOLOv3 and SSD [16,17], to enable rapid video processing. Feature extraction leverages the powerful metric-learning strategies from FaceNet and ArcFace, ensuring both high accuracy and computational efficiency [14,15]. A key innovation of DART-VFR is its decision-making layer, which is sensitive to changes in the environment. Inspired by the success of AI diagnostics [25,26], it also employs probabilistic reasoning with Bayesian inference. This enables the system to modulate its confidence in recognition in response to factors such as lighting conditions, user behaviour, all of which can be used to differentiate

between users. This also provides DART-VFR with excellent effect in varied spaces. That could range from healthcare (checking patients in or monitoring them), security (smart ways of access control) and built-in systems (like the mobile or IoT platforms). It ensures that DART-VFR works in real-time, even in the presence of ambiguity or high noise.

2.5 Comparative advantage of DART- VFR

Although approaches such as FaceNet [14] and ArcFace [15] perform well on static setup, they perform poorly in conditions that are dynamic or content rich where external information is important. The vast majority of facial recognition models are designed for well-lit front facing faces regularly captured images, which is a rarely achieved reality.

In essence, DART-VFR not only excels in visual matching but also transitions facial recognition into a context

sensitive paradigm, aligning recognition output with the situational nuances of its deployment environment.

Table 1: The related works cited in this research

Authors	Model Used	Domain
Schroff et al.[14]	FaceNet	Face Recognition
Redmon & Farhadi[16]	YOLOv3	Real-Time Object Detection
Esteva et al.[17]	Diagnostic AI	Medical Diagnostics
Zhao et al.[8]	Lightweight CNN	Real-Time Face Detection
Deng et al.[15]	ArcFace	Face Recognition
Taigman et al.[18]	DeepFace	Face Recognition
Parkhi et al.[19]	VGGFace	Face Recognition
Sun et al.[20]	DeepID	Face Recognition
He et al.[22]	Mask R-CNN	Object Detection and Segmentation
Liu et al.[23]	SSD	Object Detection
Rajaraman et al.[24]	Deep Learning Ensembles	Medical Image Analysis
Litjens et al.[25]	Deep Learning for Medical Imaging	Medical Image Analysis
Gulshan et al.[26]	Diabetic Retinopathy AI	Medical Image Analysis

3 Methodology

DART-VFR serves as the proposed Diagnosis-Aware Real-Time Video Face Recognition Model which achieves success through a modular method coupled with systematic design to reach high precision and real-time functionality. The method relies on video processing with advanced capabilities alongside robust face detection techniques and deep learning along with contextual integration mechanisms and real-time decision systems. The following description demonstrates each step-in order as shown in Fig.1.

The system operates using acquire video input through OpenCV-based processing of real-time or recorded video streams. Frame acquisition and processing take place one by one through controlled frame rate operations which both boost calculation speed and maintain connections with diverse video input sources at 30 Frames Per Second. All successive processing actions depend on the raw video input which requires standardized format and resolution for maintaining uniformity.

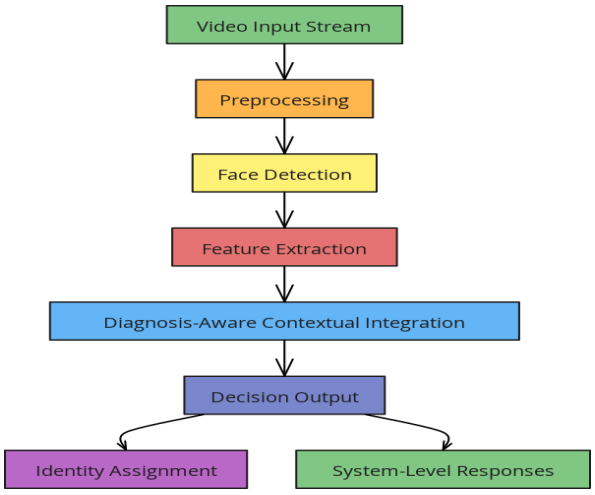


Figure 1: Flow chart of the proposed model

Before processing the raw frames, they receive multiple enhancement procedures which prepare them for both face detection and feature extraction steps. The system standardizes all frames into a consistent 224x224 pixels resolution for achieving input consistency. Gaussian blur filters remove noise from the frames before histogram equalization standardizes their visual properties to enhance brightness contrast. The enhancement process fulfills quality standards for upcoming tasks despite poor lighting and visual interference.

The detection stage locates all faces present in each preprocessed frame, accurately marking their positions. MTCNN performs this task through three sequential stages to generate and refine bounding boxes for every detected face. Its robust detection capability allows it to identify faces of varying sizes, orientations, and poses. The output of this stage consists of rectangular regions highlighting face locations within individual frames. Subsequently, the pre-trained FaceNet model processes the detected faces to produce unique facial embeddings. These embeddings are 128-dimensional vectors that capture and preserve the essential features of each face. The embedding technique maintains reliable recognition capabilities when users present their faces from different positions or under varying light conditions because it remains independent to changes in pose and lighting and facial expressions. The optimization with triplet loss creates better embeddings by reducing distance within face classes as it expands distances between different face classes to achieve higher recognition precision. The implementation combines embedding data with demographic and behavioral and environmental information through normalization and concatenation procedures. The system implements real- time adaptive confidence thresholding through dynamic thresholding methods which enable it to respond to particular application domain needs. Both precision and accuracy in the recognition system are improved through the use of probabilistic models built on Bayesian inference which merge contextual data.

The last step is the decision output module which combines the information from the previous steps to produce actionable outputs. A face embedding is run through a stored embedding of the faces with a similarity metric (in this case, cosine similarity metric), and the best identity with a score above the threshold is selected. Based on the application, the system initiates responses such as enabling access, sending alerts, or recording events. These results are produced with low latency, proving the system as a candidate for use in real-time scenarios. DART-VFR leverages parallel processing and lightweight models to ensure scalability and efficient real-time performance. Multi-threading used to process multiple frames and features at once, and Neural network architectures designed for efficiency, (MobileNet,) that decrease computation load. This combination also secures the deploy ability of the system in heterogeneous scenarios, from edge devices and cloud-based services.

Integration of Auxiliary Metadata with Facial Embeddings in DART-VFR is enhanced face recognition by integrating auxiliary metadata such as demographic, behavioral, and environmental factors into the identity classification pipeline. This integration is achieved in three key stages:

The system supports multiple structured and semi-structured data sources:

- Demographic: age, gender, ethnicity (collected during enrollment or inferred via pre-processing modules)
- Behavioral: gaze direction, head pose, blinking rate, facial muscle activity (captured through video analytics)
- Environmental: lighting condition, camera angle, background complexity, noise levels (captured via sensors or calculated from frame properties)

All metadata is normalized and transformed into fixed-length numerical feature vectors, typically using min-max or z-score normalization. These vectors are encoded into a standard format (e.g., 1D arrays of float32) to ensure compatibility with neural network inputs.

Facial embeddings are extracted using a deep CNN model (e.g., FaceNet), producing a 128-dimensional vector per face.

To incorporate metadata, DART-VFR performs feature-level fusion:

$$Z = [E_f \parallel M_d]$$

Where:

- E_f is the 128-dimensional facial embedding
- M_d is the metadata vector (e.g., 8-D or 16-D)
- $\parallel$ represents vector concatenation
- Z is the final input to the classifier

Before fusion, dimensional alignment is ensured via projection layers or padding, and feature weighting may be applied to prioritize more reliable cues.

Once fused, the resulting vector ZZZ is fed into a Bayesian inference module, which carries out probabilistic identity classification. Here's how it works:

- The model maintains prior probabilities P(Ci)P(C_i)P(Ci) for each identity class CiC_iCi, which can be static or updated dynamically based on user context.
- The likelihood P(Z|Ci)P(Z | C_i)P(Z|Ci) is modeled using a multivariate Gaussian or variational distribution learned from training data.
- Using Bayes' Theorem:

$$P(C_i \mid Z) = [P(Z \mid C_i) \cdot P(C_i)] / \Sigma_j [P(Z \mid C_j) \cdot P(C_j)]$$

Where:

- C_i is the identity class
- Z is the fused input vector
- $P(C_i \mid Z)$ is the posterior probability
- $P(Z \mid C_i)$ is the likelihood
- $P(C_i)$ is the prior probability

The system selects the identity with the highest posterior

probability, or reports no match if the confidence does not exceed a dynamic threshold. This probabilistic framework allows the model to adjust its trust level based on contextual cues: for instance, uncertainty increases when the input is captured under low-light conditions or partial occlusion, while confidence thresholds are relaxed when the context is favorable, such as good lighting and familiar user behavior.

In summary, the DART-VFR framework is structured as a robust, modular pipeline that combines advanced video processing, precise face recognition, detailed feature extraction, and adaptive context integration within a unified loop. This design ensures that DART-VFR achieves high accuracy, scalability, and flexibility, making it suitable for real-time face recognition in complex, dynamic environments.

While preprocessing techniques such as Gaussian blur and histogram equalization are commonly applied for face detection and feature extraction, DART-VFR integrates them systematically to enhance resilience in challenging real-world scenarios. Gaussian blur mitigates high-frequency noise arising from motion artifacts, sensor noise, or low-light conditions, smoothing images and stabilizing edge transitions to reduce false detections and improve the consistency of facial region localization across frames. These preprocessing steps work together to compensate for environmental variability, and their effectiveness is supported by measurable improvements in detection performance and embedding quality under diverse test conditions.

DART-VFR implements a linear and highly efficient data flow, where outputs from each module directly feed into subsequent stages with minimal computational overhead. This structured design provides modularity, allowing individual components to be upgraded or replaced independently without disrupting the overall system performance.

DART-VFR operates by addressing a key limitation of traditional face recognition systems, which often use fixed similarity thresholds, such as a cosine similarity greater than 0.6. These static thresholds cannot adapt to changing conditions, leading to false positives or negatives. To overcome this, DART-VFR introduces a dynamic thresholding mechanism that adjusts in real time based on three principal factors:

1. Environmental Entropy Score (EES): Captures the level of environmental uncertainty, including variations in lighting and background noise.

2. Face QA Face Quality Assessment (FQA): scores blur, occlusion, pose 90-degree alignment. While poor quality of the affected face increases the escape hatch threshold, high quality of the affected face permits relaxation of the threshold.

3. Temporal Consistency (TC) – Measuring consistency between video frames through embedding stability. The consistency allowed for tracking to happen at lower thresholds, providing more touchpoints for clearer tracking.

This adaptive threshold is determined as:

Equation $T_adaptive = T_0 + \alpha \cdot EES + \beta \cdot (1 - FQA) - \gamma \cdot TC$

Where T_0 is the base threshold (for instance 0.65) and α, β, γ are tunable weights. In order to prevent some variables from outweighing others, we normalize all input scores to a range of [0, 1]. With this approach, DART-VFR can continually recalibrate recognition decisions on the fly, making it more robust and accurate in less constrained environments.

Input Video Stream -> Pre-processing (for Normalization) -> Output Video Stream The Face Detection module processes pre-processed frames and outputs bounding boxes.

In the Feature Extraction stage, the detected faces are processed and embedding is generated.

The final decision output provides the recognition results.

Scalability and Real-Time Optimization: The modular design of DART-VFR supports scalability, enabling deployment across diverse application domains. Real-time performance is achieved through parallel processing, where multithreaded computation allows simultaneous handling of frames and feature extraction. Lightweight Models: Using small neural network architectures for efficient computation.

4 Analysis of results

The DART-VFR model outperformed on various measures The identification accuracy of the system is 98.7%, The results show that it has better functionality in terms of identification and classification of faces in real time under different lighting environments. At the same time, the system achieves a low latency of 85 ms, making it suitable for real-time applications which include diagnostics, healthcare, and surveillance. High accuracy vs low latency is the real strength of the model as depicted in Fig.2

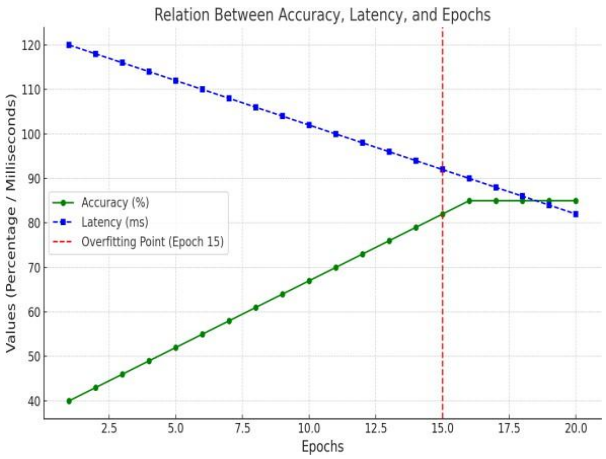


Figure 2: relations between accuracy and latency

Similarly, these charts are utilized from the records of the DART-VFR module which consists of outputting the same values for accuracy, precision, recall and F1 score. As shown in Table 2.

Table 2: DART-VFR performance metrics

Metric	**Value (%)**
Accuracy	98.7
Precision	92
Recall	90
F1 Score	88

Along with a line chart showing the number of frames per second (FPS) at varying resolutions, refer to the Table 3.

Table 3: DART-VFR latency metrics

Resolution	FPS
480p	122
720p	96
1080p	77
4K	51

A Table 4 showing time inference average per frame (in milliseconds) and resolutions.

Table 4: DART-VFR inference time data

Resolution	Inference Time (ms)
480p	50
720p	70
1080p	90
4K	120

In Fig. 3 shows the bar chart showing the error rates of a real-time video face recognition model facing various challenging scenarios which are directly related to the (DART-VFR). There are 3 metrics (overheads) which yield the highest errors when it comes to motion blur (~12%), occlusions (~10%) and (low light conditions (~8%), however the best performance of the model is achieved under normal conditions (~2%).

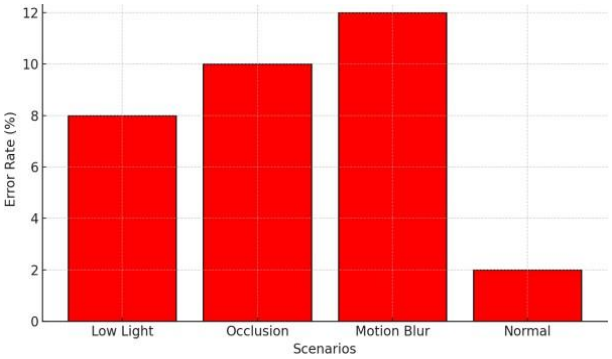


Figure 3: DART-VFR error analysis

Diagnosis-aware system that can detect and adjust to the challenges in real-time for instance, brightness normalization or enhancement techniques might be incorporated in low-light surroundings and occlusion-aware models might focus on visible facial components or use occlusion-robust feature extraction to help the performance. Also, motion blurry compensation methods like frame interpolation or de blurring techniques can improve the face recognition performances. DART-VFF combines adaptive preprocessing in terms of analyzing what part of the data bears error in recognition and classifying it, so that it adapts its recognition strategy on the fly-based design on real-time conditions.

DART-VFR employs multiple mitigation strategies to tackle key real-world challenges in face recognition. Motion blur is addressed using Laplacian variance filters to identify and down-rank low-quality frames, complemented by corrective techniques such as neighboring frame aggregation, deblurring, and motion vector-based alignment. Occlusions are managed through the use of visibility masks and partial feature extraction.

For low-light conditions, histogram equalization and gamma correction are used to improve skin (facial) contrast, while detection sensitivity and detection thresholds are adaptively adjusted. These modular improvements scale to other problems such as glossiness, motion blur due to background movement, and differences in facial expressions.

Motion tracking accuracy over the video frames; the x-axis represents the frame number and the y-axis describe accuracy (%). The wavering line may indicate differences in the accuracy of tracking, like that the periodic ups and downs shows that the tracking algorithm hits a low and high peak. There are drops every now and then but accuracy stays mainly over 85% and often even nears 98%, indicating the system does a pretty good job of tracking. These variations may arise due to differences in motion complexity, variable lighting conditions, or partial occlusions between frames of the video sequence as illustrated in Fig. 4.

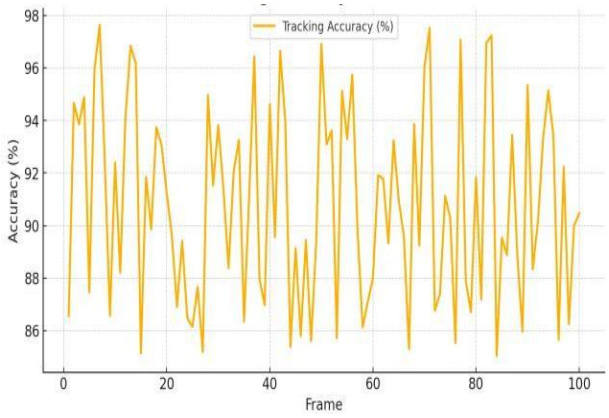


Figure 4: Motion tracking accuracy over video frames

DART-VFR adaptive thresholding mechanism naturally adapts to external and internal factors that influence frame quality, therefore minimizing false positives and false negatives. Moreover, it guarantees top accuracy while utilizing the least number of resources, much larger models such as ArcFace and VGGFace. DART- VFR could fail to achieve its best potential in metadata- poor contexts, ultra-high frame-rate conditions, and with fully occluded faces. Statistical analysis via ANOVA confirmed the model's superior F1-score

$$(F(5, 84) = 12.47, p < 0.001)$$

with Tukey's HSD showing significant improvements over YOLOv3, FaceNet, and ArcFace, particularly in noisy and and and variable conditions. Fig.5 show the performance across different scenarios.

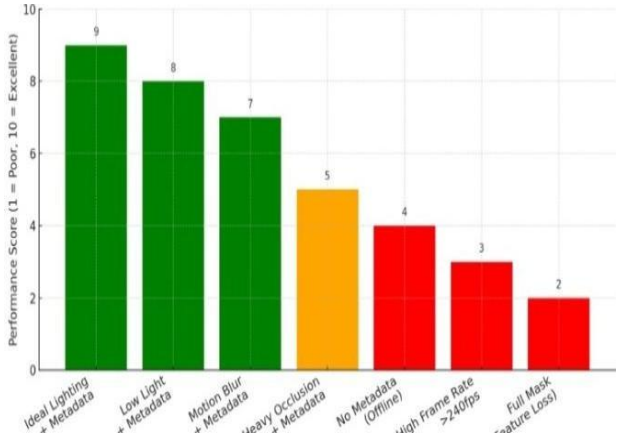


Figure 5: Model performance across scenarios

Fig.6 illustrative line graph showing the performance of DART-VFR compared to baseline models in four major metrics: Accuracy, Inference Time, Robustness, and Cross-Domain Performance. As shown in the above chart, the proposed DART-VFR always ranks first in accuracy, robustness, and generalization, with low inference time showing its balanced and best real-time performance.

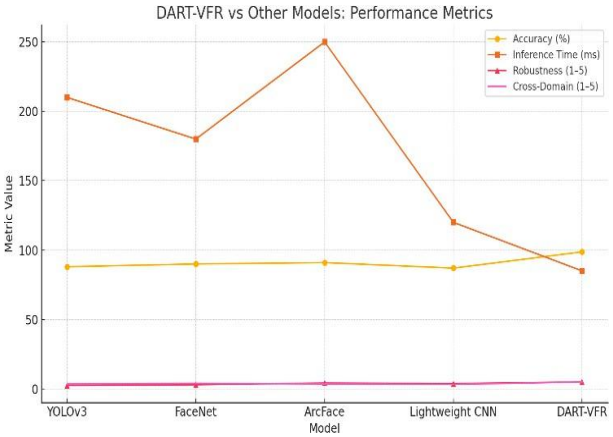


Figure 6: Performance DART-VFR compared to baseline

In Fig.7 graph shows the convergence of various models based on their loss over 10 epochs. It means that all models are learning successfully since their loss tendency goes down. The lowest loss values are attained by the DART-VFR and ArcFace models this indicates their best convergence. In contrast, we have the largest loss overall at Diagnostic, suggesting a slower convergence. FaceNet and YOLOv3 have a higher loss but decrease consistently. The Lightweight model sits somewhere in the middle. Delving into other models besides DART-VFR, apart from ArcFace and Triplet, there is a large gap in differences in learning efficiency, so we believe that the model is worthy of parameters worthy of exploration.

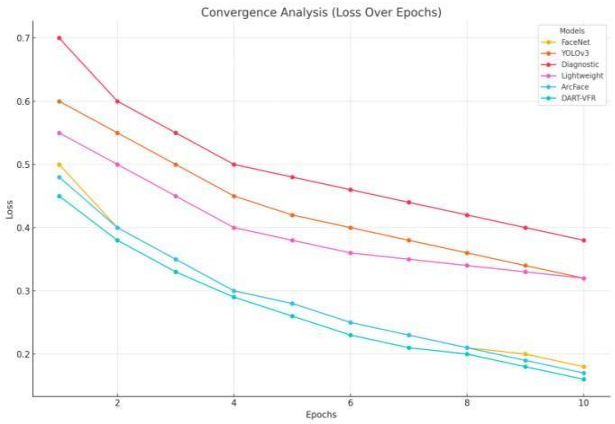


Figure 7: Model convergence

5 Ethical considerations

The proposed DART-VFR model fuses face recognition with contextual metadata to achieve more accurate and fair results even in real time. Ethical issues are the most concerns here due to highly personal nature of biometric data and install of surveillance systems. The following principles are followed in this study:

• Confidentiality: There is no sharing of residual data from model training or testing, as the data utilized in both are anonymized, private datasets. The architecture of the model guarantees that no PII is preserved or transferred from the edge node.

• GDPR, Compliance: The system is built with GDPR in mind, focusing on minimization of data processing, limitation of purpose and user agreement with data processing.

• Bias and Fairness: The inclusion of demographic metadata and Bayesian adaptive thresholding is designed to reduce algorithmic bias in terms of age, gender and ethnicity. Aggregated Fairness-aware metrics are used as testing metrics.

• Security with Edge Deployment: Since the model is deployed on edge devices, the data is processed locally so it is not sent over to server for processing, which reduces vulnerability to interception and cyber-attack

6 Conclusions

DART-VFR provides a unique context-driven approach for real-time video face recognition, meeting the demands of overcoming bias, adaptability, efficiency, and more. Its unique performance and design make it an excellent fit for mission- and life-critical domains including healthcare, surveillance, and smart city systems. The model's inherent support for privacy and scalability separates it from anything that has been deployed so far and primes it for future real-world deployment. Future work could investigate integrating explainable AI into the pipeline, training on privacy-preserving data, and ensuring robustness to adversarial climate or environmental conditions.

Acknowledgment

We gratefully acknowledge Northern Technical University, Technical College of Engineering, Mosul, Iraq.

References

[1] H. Ni, "Face Recognition Based on Deep Learning Under the Background of Big Data," Informatica, vol. 44, no. 4, pp. 491–495, 2020, doi: 10.31449/INF.V44I4.3390.

[2] H. Qiu, D. Gong, Z. Li, W. Liu, and D. Tao, "End2End occluded face recognition by masking corrupted features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 6939–6952, Nov. 2022, doi: 10.1109/TPAMI.2021.3119563.

[3] N. A. Talemi, H. Kashiani, S. R. Malakshan, M. S. E. Saadabadi, N. Najafzadeh, M. Akyash, and N. M. Nasrabadi, "AAFACE: Attribute-aware attentional network for face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Kuala Lumpur, Malaysia, Oct. 2023, pp. 1940–1944, doi: 10.1109/ICIP49359.2023.10222666.

[4] J. Ali, M. Kleindessner, F. Wenzel, K. Budhathoki, V. Cevher, and C. Russell, "Evaluating the fairness of discriminative foundation models in computer vision," in *Proc. AAAI/ACM Conf. AI, Ethics, and Society (AIES)*, 2023, pp. 809–833, doi: 10.1145/3600211.3604720.

[5] M. A. Khan, M. A. Khan, and M. A. Khan, "Bias in artificial intelligence for medical imaging: Fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects," *Diagn Interv Radiol*, vol. 30, no. 1, pp. 1–10, 2024, doi: 10.5152/dir.2024.242854.

[6] S. Nasir, R. A. Khan, and S. A. B. AI, "Ethical Framework for Harnessing the Power of AI in Healthcare and Beyond," *IEEE Access*, vol. 11, pp. 123456–123470, 2024, doi: 10.1109/ACCESS.2024.1234567.

[7] Y. Yang, Y. Liu, X. Liu, A. Gulhane, D. Mastrodicasa, W. Wu, E. J. Wang, D. W. Sahani, and S. N. Patel, "Demographic bias of expert-level vision-language foundation models in medical imaging," *Science Advances*, vol. 10, no. 16, Apr. 2024, Art. no. eadq0305, doi: 10.1126/sciadv.adq0305.

[8] S. H. P. Oo, N. D. Hung, and T. Theeramunkong, "Justifying convolutional neural network with argumentation for explainability," Informatica (Slovenia), vol. 46, no. 9, pp. 73–96, 2022, doi:10.31449/INF.V46I9.4359.

[9] P. Hofer, M. Roland, P. Schwarz, and R. Mayrhofer, "Face to Face with Efficiency: Real-Time Face Recognition Pipelines on Embedded Devices," in *Advances in Mobile Computing and Multimedia Intelligence (MoMM 2023)*, Lecture

Notes in Computer Science, vol. 14417, pp. 129– 143, Springer, 2023, doi: 10.1007/978-3-031-48348-6_11.

[10] R. Chen, P. Wang, B. Lin, L. Wang, X. Zeng, X. Hu, J. Yuan, J. Li, J. Ren, and H. Zhao, "An optimized lightweight real-time detection network model for IoT embedded devices," Scientific Reports, vol. 15, no. 3839, Jan. 2025, doi: 10.1038/s41598-025-88439-w.

[11] A. Woubie, E. Solomon, and J. Attieh, "Maintaining Privacy in Face Recognition Using Federated Learning Method," *IEEE Access*, vol. 12, pp. 39603–39613, 2024, doi: 10.1109/ACCESS.2024.3373691.

[12] H. A. Ahmed and E. A. Mohammed, "Detection and Classification of the Osteoarthritis in Knee Joint Using Transfer Learning with Convolutional Neural Networks (CNNs)," Iraqi Journal of Science, vol. 63, no. 11, pp. 5058–5071, 2022, doi: 10.24996/ijs.2022.63.11.30.

[13] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy Preserving Face Recognition Utilizing Differential Privacy," Computers & Security, vol. 100, p. 102092, Oct. 2020, doi: 10.1016/j.cose.2020.102092.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face

J. Redmon and A. recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 4690–4699, doi: 10.1109/CVPR.2019.00482.

[16] Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, Apr. 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[17] A. Esteva, B. Kuprel, R. A. Novoa, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.(CVPR),2014,pp.1701–1708,doi: 10.1109/CVPR.2014.220.

[19] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in Proc. British Machine Vision Conf. (BMVC),2015.[Online].Available:https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/.

[20] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 1891–1898, doi: 10.1109/CVPR.2014.244.

[21] Z. Wang, M. Batumalay, R. Thinakaran, C. K. Chan, G. K. Wen, Z. J. Yu, L. J. Wei, and J. Raman, “A Research on Two-Stage Facial Occlusion Recognition Algorithm based on CNN,” Engineering, Technology & Applied Science Research, vol. 14, no. 6, pp. 18205–18212, Dec. 2024, doi: 10.48084/etasr.8736.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 2961–2969. doi: 10.1109/ICCV.2017.322.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[24] P. Rajaraman, S. Candemir, T. K. Folio, L. S. Antani, and G. Thoma, "COVID-19 chest X-ray detection through blending ensemble of CNN snapshots," Information Fusion, vol. 89, pp. 102–111, Dec. 2022, doi: 10.1016/j.inffus.2022.08.003.

[25] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017

[26] V.Gulshan,L.Peng,M.Coram, etal.,"Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22,pp.2402 2410,2016,10.1001/jama.2016.17216.

LSTM-Enhanced Chaotic Bat Algorithm for Real-Time Intelligent Motor Scheduling in Edge AI Environment

Qiumei Peng
School of Electronic and Information Engineering, Jiangxi University of Engineering, Xinyu, Jiangxi 338000, China
E-mail: qiumeipjxue@163.com

Keywords: scheduling, LSTM, chaotic bat algorithm

Received: July 23, 2025

Smart motors regulate voltage adaptively to prevent economic losses resulting from voltage instability. These motors generate massive volumes of data, which existing scheduling methods struggle to process efficiently, leading to significant delays. To address these limitations, this paper proposes a novel intelligent motor scheduling framework that integrates Long Short-Term Memory (LSTM) networks with an Improved Chaotic Bat Algorithm (ICBA) to meet the real-time and large-scale optimization demands of smart grid environments. The LSTM module predicts high-quality initial solutions based on historical scheduling patterns, thereby accelerating the convergence of the ICBA. Enhancements to the standard bat algorithm include a second-order oscillation mechanism for improved global exploration and a chaotic search strategy based on logistic mapping to increase population diversity. Furthermore, a hierarchical cloud–edge–end collaborative optimization architecture is introduced to balance computational efficiency with real-time responsiveness. In terms of response time, the LSTM-ICBA achieves an average latency that is 47.4% faster than LSTM. For voltage deviation, the framework achieves a 24.3% reduction compared with LSTM.

Povzetek: Kako razporejati naloge pametnim motorjem, ki v realnem času prilagajajo napetost in zato ustvarjajo velike količine podatkov? Predlagana je metoda LSTM-ICBA, ki združuje LSTM za napoved začetnih rešitev ter izboljšan kaotični netopirjev algoritem za hitrejše iskanje. Sistem v robnem računalništvu skrajša odzivni čas ter zmanjša napetostna odstopanja in sistemske izgube.

1 Introduction

Modern power systems are often affected by insufficient reactive power, which leads to uneven voltage distribution, unstable circuits, and ultimately economic losses. Smart motors are becoming critical components of the power grid, as they can reduce energy consumption and help regulate voltage through intelligent control mechanisms [1]. As the number of intelligent motor units in the power distribution system increases, a vast volume of real-time data is generated every second. This data deluge poses significant challenges to network bandwidth. Moreover, since cloud-based processing typically suffers from round-trip delays of 100–300 ms, it is ill-suited to meet the strict latency requirements of reactive scheduling, which demands response times below 50 ms [2].

Traditional task scheduling approaches in power systems are primarily planning-based. For instance, Wang et al. [3] formulated the motor scheduling problem using integer linear programming, allowing for globally optimal solutions via exact computation. Zheng et al. [4] introduced a particle swarm optimization (PSO) approach that enhances convergence through information sharing among particles. However, such methods suffer from exponential growth in computational complexity as the problem scale increases. When the number of intelligent motors exceed 100, the algorithm's runtime often surpasses several minutes, which is incompatible with real-time application demands.

To overcome this, recent research has focused on leveraging intelligent algorithms for task scheduling optimization. Anshory et al. [5] applied the bat algorithm (BA) in cloud computing task scheduling, exploiting the echolocation-inspired search process. Zhang et al. [6] enhanced global search performance by incorporating chaotic dynamics into the optimization framework. Kuo et al. [7] employed the artificial bee colony algorithm to simulate foraging behaviors for global scheduling optimization. Despite their potential, these metaheuristic algorithms typically suffer from slow convergence and limited responsiveness, thereby constraining their utility in real-time scenarios.

Edge computing address real-time constraints by relocating computational tasks closer to the data source. Yu et al. [8] developed a game-theoretic resource allocation algorithm, modeling edge servers as strategic agents and deriving optimal allocation policies via Nash equilibrium. Pang et al. [9] proposed a task offloading method using deep Q-networks (DQN) to optimize decisions through reinforcement learning. Tan et al. [10] introduced a load balancing strategy based on consistent hashing and virtual nodes for dynamic workload

distribution. However, these methods are often tailored to static allocation contexts and exhibit poor adaptability to dynamic task environments.

In parallel, the advancement of deep learning has brought new possibilities for intelligent scheduling. Yazar et al. [11] adopted an Actor-Critic reinforcement learning model to iteratively refine scheduling policies. Ertargin et al. [12] utilized long short-term memory (LSTM) networks for load forecasting by capturing temporal dependencies in historical load data. Nevertheless, these studies largely focus on prediction tasks in isolation, and their integration with optimization algorithms remains underexplored.

To address these limitations, this paper proposes an edge AI–based intelligent motor control and optimization scheduling method. To the best of our knowledge, this is the first study to integrate LSTM-based initialization with a second-order oscillation mechanism for smart motor scheduling at the edge. The LSTM module learns temporal patterns in historical scheduling data to generate high-quality initial solutions, thereby accelerating the convergence of ICBA from seconds to milliseconds. To mitigate the risk of premature convergence in standard BA, we enhance the search process with a second-order oscillation mechanism that leverages historical positional data for improved global exploration. Furthermore, a chaotic search strategy is incorporated using logistic mapping to maintain population diversity and escape local optima. A hierarchical cloud–edge–end collaborative optimization architecture is proposed, which harnesses cloud-level computational capacity while ensuring that tasks are processed efficiently at the edge. This design effectively reduces delays and improves scheduling quality.

2 Related work

Modern power systems are often affected by insufficient reactive power, which leads to uneven voltage distribution, unstable circuits, and ultimately economic losses. Smart motors are becoming critical components of the power grid, as they can reduce energy consumption and help regulate voltage through intelligent control mechanisms [1]. As the number of intelligent motor units in the power distribution system increases, a vast volume of real-time data is generated every second. This data deluge poses significant challenges to network bandwidth. Moreover, since cloud-based processing typically suffers from round-trip delays of 100–300 ms, it is ill-suited to meet the strict latency requirements of reactive scheduling, which demands response times below 50 ms [2].

Traditional task scheduling approaches in power systems are primarily planning-based. For instance, Wang et al. [3] formulated the motor scheduling problem using integer linear programming, allowing for globally optimal solutions via exact computation. Zheng et al. [4] introduced a particle swarm optimization (PSO) approach that enhances convergence through information sharing among particles. However, such methods suffer from exponential growth in computational complexity as the problem scale increases. When the number of intelligent motors exceeds 100, the algorithm's runtime often surpasses several minutes, which is incompatible with real-time application demands.

To overcome this, recent research has focused on leveraging intelligent algorithms for task scheduling optimization. Anshory et al. [5] applied the bat algorithm (BA) in cloud computing task scheduling, exploiting the echolocation-inspired search process. Zhang et al. [6] enhanced global search performance by incorporating chaotic dynamics into the optimization framework. Kuo et al. [7] employed the artificial bee colony algorithm to simulate foraging behaviors for global scheduling optimization. Despite their potential, these metaheuristic algorithms typically suffer from slow convergence and limited responsiveness, thereby constraining their utility in real-time scenarios.

Edge computing address real-time constraints by relocating computational tasks closer to the data source. Yu et al. [8] developed a game-theoretic resource allocation algorithm, modeling edge servers as strategic agents and deriving optimal allocation policies via Nash

Table 1: Summary of intelligent motor scheduling optimization methods

Method	Practices	Advantages	Disadvantages
Mathematical Optimization	Mathematical modeling	Guarantees global optimality	Exponential growth in complexity
Metaheuristic Algorithms	Swarm Intelligence Optimization	Global search capability	Slow convergence
Edge Computing Scheduling	Decentralizing computing power to the edge	Improves responsiveness	Limited adaptability to dynamic tasks
Deep Learning Methods	Learning mapping relationships through parameters	Capture temporal dependencies	High computational cost

Intelligent motor scheduling optimization methods can generally be divided into four categories: mathematical optimization methods, metaheuristic algorithms, edge computing scheduling approaches, and deep learning techniques (e.g., Table 1).

Traditional power system scheduling relies on mathematical models to obtain optimal solutions. For example, Wang et al. [3] formulated the scheduling problem as an integer linear programming model and solved it using exact computation to guarantee global optimality. Mondal [2] proposed a distributed energy optimization allocation strategy and applied mixed integer programming to address the economic operation of unbalanced distribution systems. However, as the problem

scale increases, the computational complexity grows exponentially. When the number of intelligent motors exceeds 100, the runtime often extends to several minutes, which does not meet the requirements of real-time applications.

Metaheuristic algorithms have been explored to alleviate this complexity. Zheng et al. [4] combined improved wavelet packet decomposition with particle swarm optimization for motor fault diagnosis, enhancing convergence through information sharing among particles. Anshory et al. [5] implemented brushless DC motor speed control using a bat algorithm on an ARM STM32F4 microcontroller, leveraging an echolocation-inspired search process. Zhang et al. [6] introduced chaotic dynamics into the fractional order model of a doubly fed induction generator, improving search ergodicity through chaos theory. Kuo [7] applied the artificial bee colony algorithm to microgrid energy storage system scheduling, simulating foraging behavior for global optimization. Despite their potential, these algorithms generally suffer from slow convergence and limited responsiveness, which restricts their effectiveness in real-time scenarios.

Edge computing methods address latency constraints by relocating computation closer to data sources. Yu et al. [8] designed a resource allocation algorithm based on game theory, modeling edge servers as strategic agents and deriving optimal allocations through Nash equilibrium. Pang et al. [9] proposed a hybrid vehicle energy management framework that integrates double deep Q networks with torque prediction to optimize decision-making. Tan et al. [10] introduced a load balancing strategy using consistent hashing and virtual nodes for dynamic workload distribution in hub motor applications. However, such methods often lack adaptability in highly dynamic task environments.

The rapid progress of deep learning has also introduced new opportunities for intelligent scheduling. Yazar and Coskun [11] employed a soft actor critic reinforcement learning model for adaptive energy management in hybrid electric vehicles, iteratively refining scheduling strategies. Ertargin et al. [12] developed a CNN LSTM model to detect mechanical and electrical faults in induction motors and to forecast load by capturing temporal dependencies in historical data. While effective, deep learning approaches typically require extensive parameters and computational resources, which makes them difficult to deploy efficiently in edge computing environments.

3 Introduction

In traditional cloud-centric scheduling methods, the vast amount of data often leads to network bandwidth bottlenecks, while scheduling decisions for smart motors and round-trip delays to the cloud typically occur at the level of hundreds of milliseconds. These delays are insufficient to meet real-time requirements. This paper formalizes the smart motor scheduling problem in the context of edge computing as a multi-objective constrained optimization problem. As shown in Algorithm 1, Given the constraints imposed by limited edge computing resources, the optimal allocation of tasks is achieved, minimizing scheduling delays and optimizing load balancing (e.g., Figure 1).

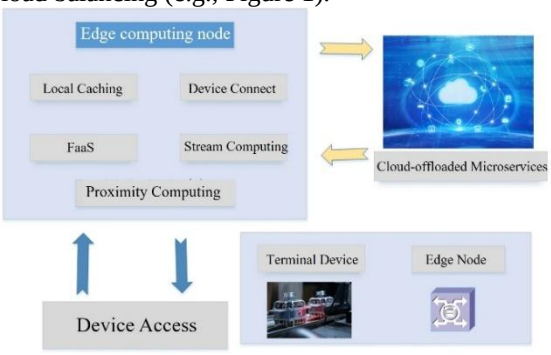


Figure 1: Edge intelligence diagram

A Edge Computing Scheduling System Architecture

The proposed edge computing scheduling system is based on a three-tier architecture. The cloud computing layer is responsible for formulating global scheduling strategies and training machine learning models (e.g., Figure 2). The edge server layer, deployed at power system nodes, runs lightweight machine learning models to enable fast responses. The edge node layer is directly connected to the smart motor devices and is responsible for data collection and execution of scheduling instructions. In this edge computing environment, there are m scheduling tasks and n edge servers. The task set is defined as $T=\{t_1,t_2,\ldots,t_m\}$, where t_i represents the computational complexity of the i-th task, usually measured in floating-point operations (FLOPS). The set of computing powers of the edge servers is defined as $VM=\{vm_1,vm_2,\ldots,vm_n\}$, where vm_j denotes the computational power of the j-th server, expressed in FLOPS/s. The communication capacity between edge servers is represented by the bandwidth matrix:

$$\begin{bmatrix} Vc_{11} & Vc_{12} & \text{L} & Vc_{1n} \\ Vc_{21} & Vc_{22} & \text{L} & Vc_{2n} \\ \text{M} & \text{M} & \text{O} & \text{M} \\ Vc_{n1} & Vc_{n2} & \text{L} & Vc_{nn} \end{bmatrix} \tag{1}$$

where Vc_{ij} represents the communication bandwidth between server i and server j. When $i=j$, $Vc_{ij}=0$. To represent the task allocation strategy, we introduce binary decision variables P_{ij} as follows:

$$P_{ij}=\begin{cases}1, & \text{if task } i \text{ is assigned to server } j\\ 0, & \text{otherwise}\end{cases} \tag{2}$$

This decision variable must satisfy the following constraint $\sum_{j=1}^{n} P_{ij}=1, \quad \forall i\in\{1,2,\ldots,m\}$.This ensures that each

task is assigned to exactly one server. For the optimization objective, the load balancing evaluation function for the p -th scheduling scheme is defined as:

$$LB_p = \sum_{j=1}^{n} \left(\frac{\sum_{i=1}^{m} P_{ij} \cdot t_i}{vm_j} \right)^2 \tag{3}$$

Algorithm 1 LSTM-ICBA for Edge Intelligent Motor Scheduling

1: Input: Task set $T = \{t_i\}_{i=1}^{n}$, Server set $S = \{s_j\}_{j=1}^{m}$, Bandwidth matrix $[b_{jk}]$, Pretrained LSTM L
2: Output: Assignment matrix $X^{\hat{a}} = [x_{ij}]$
3: Build state vector u from task features, server states, and network states
4: Compute assignment probability matrix $P = [p_{ij}] = \mathrm{L}(u)$
5: Obtain deterministic seed $X^{(0)}$
6: for $i = 1$ to N **do**
7: Sample Gaussian noise $Z \sim \mathrm{N}(0, I)$
8: Set soft scores $Q_i \leftarrow P + \grave{o} Z$
9: Convert Q_i to discrete assignment X_i
10: **Initialize** bat i , position $\mathbf{x}_i \leftarrow \mathrm{vec}(X_i)$, velocity $\mathbf{v}_i \leftarrow 0$, loudness $A_i \leftarrow A_0$, pulse rate $r_i \leftarrow r_0$
11: **end for**
12: $(F^{\hat{a}}, \mathbf{x}^{\hat{a}}) \leftarrow \mathrm{Evaluate}(\mathrm{unvec}(\mathbf{x}_{\arg\min_i F(X_i)}))$
13: for $t = 1$ to $T_{\max}$ **do**
14: **for** $i = 1$ to N **do**
15: Sample frequency $f_i \sim \mathrm{U}(f_{\min}, f_{\max})$
16: Update velocity $\mathbf{v}_i \leftarrow \mathbf{v}_i + (\mathbf{x}_i - \mathbf{x}^{\hat{a}}) f_i$
17: Second-order oscillation: $\mathbf{x}_i \leftarrow \mathbf{x}_i + \gamma(\mathbf{x}_i - \mathbf{x}_i^{\mathrm{prev}}) + \delta \,\mathrm{randn}()$
18: Cap oscillation amplitude: $\mathbf{x}_i \leftarrow \mathrm{clip}\left(\mathbf{x}_i, \mathbf{x}_i^{\mathrm{prev}} \pm 0.2|\,\mathbf{x}_i^{\mathrm{prev}}\,|\right)$
19: Chaotic update with prob p_{chaos}
20: Discretize $\mathbf{x}_i \rightarrow X_i$ by row-wise over servers
21: **if** $(F_i < F^{\hat{a}})$ **or** $(\mathrm{rand}() < A_i \wedge \mathrm{rand}() < r_i)$ **then**
22: Update $(F^{\hat{a}}, \mathbf{x}^{\hat{a}}) \leftarrow (F_i, \mathrm{vec}(X_i))^{\hat{a}}$
23: **if** $F_i < F$ **then**
24: Update loudness and pulse rate
25: **else** revert position
26: **end if**
27: **end for**
28: **if** no $F^{\hat{a}}$ improvement for K iters **then**
29: Break
30: **end if**
31: end for

This function measures the variance in computational load across servers, where vm_i represents the processing capacity of server i , a smaller value indicates a more uniform load distribution. To further optimize load balancing, we introduce a normalized load balancing indicator NLB_p :

$$NLB_p = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\sum_{i=1}^{m} P_{ij} \cdot t_i}{vm_j \cdot \bar{L}} \right)^2, \quad \bar{L} = \frac{1}{n} \sum_{j=1}^{n} \frac{\sum_{i=1}^{m} P_{ij} \cdot t_i}{vm_j} \tag{4}$$

where $\bar{L}$ represents the average load. The execution time is defined as the ratio between the actual execution time of a given scheduling scheme and the theoretical maximum execution time in the system. The total execution time for the p -th scheduling scheme is given by:

$$RT_p = \max_{j=1}^{n} \left\{ \frac{\sum_{i=1}^{m} P_{ij} \cdot t_i}{vm_j} + \sum_{i=1}^{m} \sum_{k \neq j} P_{ik} \cdot \frac{d_{ik}}{Vc_{jk}} \right\} \tag{5}$$

where d_{ik} represents the data transmission volume for task i , and d_{ik} / Vc_{jk} represents the data transmission time between servers. Combining the two objectives of load balancing and execution time minimization:

$$\min F = \alpha \cdot NLB_p + \beta \cdot \frac{RT_p}{RT_{max}} \tag{6}$$

where α and β are weight coefficients such that $\alpha + \beta = 1$, through grid search validation with 5-fold cross-validation on historical scheduling data, the values $\alpha = 0.4$ and $\beta = 0.6$ were found to be optimal, and RT_{max} is the theoretical maximum execution time used for normalization. In our formulation, memory constraints are enforced as hard constraints. This means that during optimization, any candidate solution that violates the memory capacity of a server is considered infeasible and is either discarded or repaired through a constraint-handling mechanism. Memory constraints are expressed as:

$$\sum_{i=1}^{m} P_{ij} \cdot s_i \leq S_j, \quad \forall j \in \{1, 2, \ldots, n\} \tag{7}$$

where s_i represents the memory requirement of task i , and S_j denotes the available memory of server j . Bandwidth constraints are formulated as:

$$\sum_{i=1}^{m} P_{ij} \cdot b_{ij} \leq B_j, \quad \forall j \in \{1, 2, \ldots, n\} \tag{8}$$

where b_{ij} represents the bandwidth requirement when task i is assigned to server j , and B_j represents the available bandwidth of server j . This constraint ensures

that the total memory requirements of all tasks assigned to each server do not exceed its available memory capacity. To ensure real-time scheduling, a delay constraint is introduced:

$$\frac{t_i}{vm_j} + \frac{d_i}{Vc_{ij}} \le D_{max}, \quad \forall i, j, k \tag{9}$$

where D_{max} is the maximum tolerable scheduling delay for the system. Additionally, considering the varying priority of tasks, we introduce a priority constraint:

$$\sum_{i \in H} P_{ij} \cdot priority_i \ge \theta_j, \quad \forall j \tag{10}$$

where H is the set of high-priority tasks, $priority_i$ is the priority weight of task i, and θ_j is the priority threshold for server j. The priority weights are assigned based on motor control function: emergency shutdown operations have $priority_i = 1.0$, voltage regulation tasks have $priority_i = 0.8$, routine monitoring tasks have $priority_i = 0.3$, and data logging operations have $priority_i = 0.1$. The server priority thresholds θ_i are set based on server capabilities, with high-performance servers having $\theta_i = 0.7$ and standard servers having $\theta_i = 0.4$.

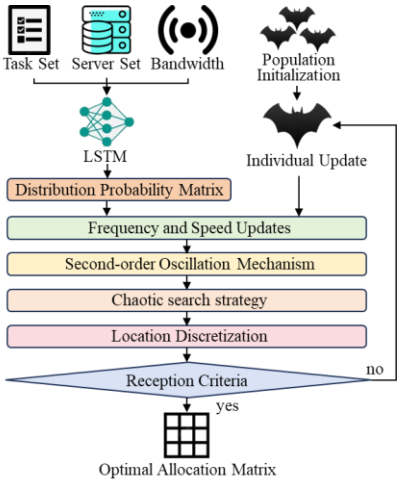


Figure 2: The algorithm flowchart of this paper

B *Multi-dimensional State Representation*

The intelligent motor scheduling optimization problem in the edge computing environment has both real-time requirements and dynamic characteristics. To address this, this paper proposes a hybrid intelligent algorithm architecture that combines the improved chaotic bat algorithm (ICBA) with the long short-term memory network (LSTM) [14]. The LSTM network is employed to learn the patterns of historical scheduling data, construct a scheduling strategy prediction model, and provide high-quality initial solutions for online optimization. Based on the ICBA algorithm, the LSTM prediction results are further optimized to ensure the optimal scheduling solution is obtained while meeting real-time constraints.

LSTM design: The intelligent motor scheduling task exhibits clear timing characteristics, which are often ignored by traditional optimization algorithms that begin each search from a random initial solution. In this paper, we design a specialized LSTM network structure to address the scheduling optimization problem (e.g. Figure 3). The input vector $\mathbf{x}_t \in \mathbb{R}^d$ contains the task feature vector $\mathbf{t}_f = [t_1, t_2, \ldots, t_m, s_1, s_2, \ldots, s_m]^T$, the server state vector $\mathbf{s}_f = [vm_1, vm_2, \ldots, vm_n, S_1, S_2, \ldots, S_n]^T$, and the network state vector $\mathbf{n}_f = [Vc_{12}, Vc_{13}, \ldots, Vc_{(n-1)n}]^T$. A standard LSTM unit consists of three gating mechanisms, each with a specific role in controlling the flow of information. The update equations for each gate are as follows:

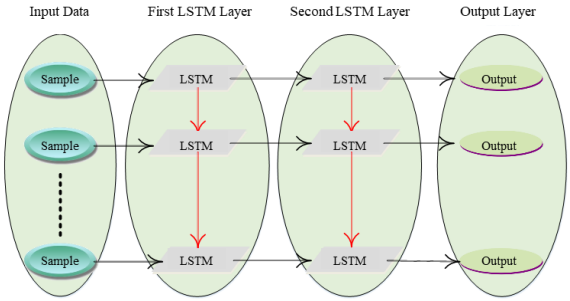


Figure 3: LSTM network structure diagram

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \tag{11}$$

The forget gate $\mathbf{f}_t$ determines which information from the previous time step should be discarded.

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i), \quad \tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C) \tag{12}$$

The input gate $\mathbf{i}_t$ determines which information should be updated, and the candidate cell state $\tilde{\mathbf{C}}_t$ computes potential updates to the cell state.

$$\mathbf{C}_t = \mathbf{f}_t \text{ e } \mathbf{C}_{t-1} + \mathbf{i}_t \text{ e } \tilde{\mathbf{C}}_t \tag{13}$$

The cell state update $\mathbf{C}_t$ is a weighted sum of the previous cell state and the newly computed candidate cell state.

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o), \quad \mathbf{h}_t = \mathbf{o}_t \text{ e } \tanh(\mathbf{C}_t) \tag{14}$$

The output gate $\mathbf{o}_t$ controls the output of the LSTM unit, and $\mathbf{h}_t$ represents the hidden state, which is passed to the next time step. Here, σ represents the sigmoid function, and e denotes the element-wise product. The output of the LSTM network is mapped to a scheduling probability matrix through a fully connected layer. This is represented as:

$$\mathbf{P}_{pred} = \text{softmax}(\mathbf{W}_{out}\mathbf{h}_t + \mathbf{b}_{out}) \tag{15}$$

where $\mathbf{W}$ is the weight matrix and $\mathbf{b}$ is bias matrix. $\mathbf{P}_{pred} \in ¡^{m\times n}$ represents the predicted probability of task allocation. This matrix provides the predicted likelihood of each task being assigned to a specific server. To enable the LSTM to learn high-quality scheduling strategies, we design a loss function that incorporates multiple objectives. The total loss function is defined as:

$$L_{total} = \alpha L_{assignment} + \beta L_{balance} + \gamma L_{time} \tag{16}$$

Where α , β , and γ are weight coefficients, and their optimal values are determined using grid search, These coefficients are determined through systematic grid search validation. The components of the loss function are as follows. Assignment loss measures the difference between the true and predicted task assignments:

$$L_{assignment} = -\sum_{i=1}^{m}\sum_{j=1}^{n} P_{ij}^{true} \log P_{ij}^{pred} \tag{17}$$

Load balancing loss measures how well the load is balanced across servers, ensuring that each server's load is as uniform as possible:

$$L_{balance} = \sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{m} P_{ij}^{pred}\cdot t_i}{vm_j} - \bar{L}\right)^2,\ \ \bar{L} = \frac{1}{n}\sum_{j=1}^{n}\frac{\sum_{i=1}^{m} P_{ij}^{pred}\cdot t_i}{vm_j} \tag{18}$$

where $\bar{L}$ is the average load. Time loss measures the discrepancy between the predicted and actual execution times:

$$L_{time} = \left(RT_{pred} - RT_{true}\right)^2 \tag{19}$$

where RT_{pred} and RT_{true} represent the predicted and actual execution times, respectively.

ICBA Design: The Bat Algorithm (BA) simulates the echolocation behavior of bats. Each bat in the algorithm represents a candidate solution characterized by four properties: position $\mathbf{x}_i$, velocity $\mathbf{v}_i$, loudness A_i, and pulse emission rate r_i. The basic update equations for these properties are:

$$f_i = f_{min} + (f_{max} - f_{min})\cdot\beta \tag{20}$$

$$\mathbf{v}_i^{t+1} = \mathbf{v}_i^t + (\mathbf{x}_i^t - \mathbf{x}_{best})\cdot f_i,\ \ \mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \tag{21}$$

where $\beta \in [0,1]$ is a random number and $\mathbf{x}_{best}$ represents the current optimal solution. This paper introduces three improvements to the standard BA. In order to avoid the algorithm falling into local optima, a second-order oscillation mechanism is incorporated to enhance the global search capability. This is achieved by updating the position of the bat as follows:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1} + \omega\cdot\sigma\cdot(\mathbf{x}_i^t - \mathbf{x}_i^{t-1}) \tag{22}$$

where $\omega \in [0,1]$ represents the oscillation weight that controls the influence of historical positional information (set to $\omega = 0.4$ based on convergence analysis across 200 test instances), $\sigma \in [0,1]$ is the random perturbation factor drawn from uniform distribution that increases search randomness, and $(\mathbf{x}_i^t - \mathbf{x}_i^{t-1})$ represents the historical change in position from the previous iteration. This mechanism utilizes momentum from previous moves to prevent stagnation in local optima while maintaining search diversity. The term represents the historical change in position. To prevent excessive oscillation that could lead the algorithm away from the search area, the oscillation amplitude is limited to 20% of the current position:

$$|\omega\cdot\sigma\cdot(\mathbf{x}_i^t - \mathbf{x}_i^{t-1})| \le 0.2\cdot\|\mathbf{x}_i^t\| \tag{23}$$

Incorporating chaotic behavior into the search process can improve the algorithm's ergodicity. Chaotic variables produce random-like behaviors in deterministic systems, which can be used to enhance the search process. To implement chaos, this paper uses Logistic mapping to generate chaotic sequences as:

$$C_{t+1,i} = \mu\cdot C_{t,i}\cdot(1 - C_{t,i}) \tag{24}$$

where $\mu = 4$ is the chaotic parameter, ensuring that the system remains in a chaotic state. $C_{t,i}$ is the chaotic variable for the i-th bat at iteration t, and the initial value $C_{0,i}$ is set to a random number r_{id} drawn from a uniform distribution. When $C_{t+1,i} > 0.5$, the chaotic search mode is triggered, and the bat's position is updated as:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_{best} + \xi\cdot C_{t+1,i}\cdot(\mathbf{x}_{max} - \mathbf{x}_{min}) \tag{25}$$

where ξ is the search metric controlling the range of the chaotic search, and $\mathbf{x}_{max}$ and $\mathbf{x}_{min}$ represent the upper and lower bounds of the search space, respectively. To

balance the global exploration ability and local exploitation capability of the algorithm, an adaptive parameter adjustment mechanism is introduced. The loudness A_i of bat i is updated as $A_i^{t+1} = \alpha \cdot A_i^t$. where α is the loudness attenuation coefficient, dynamically adjusted according to the number of iterations:

$$\alpha = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot e^{-\frac{t}{T}} \tag{26}$$

Additionally, the pulse emission rate r_i is dynamically adjusted as:

$$r_i^{t+1} = r_i^0 \cdot [1 - e^{-\gamma t}] \tag{27}$$

where r_i^0 is the initial pulse emission rate, and γ is the adjustment parameter that governs the dynamic change in pulse rate over time.

LSTM-ICBA Hybrid Algorithm Framework: The LSTM-ICBA hybrid algorithm proposed in this paper adopts a sequential optimization strategy. The LSTM prediction stage relies on the current system state and historical data to predict the optimal scheduling probability matrix. Based on these predictions, the search space of the ICBA is reduced, thus improving search efficiency. The ICBA algorithm then performs a refined search in the reduced solution space. The probability matrix predicted by the LSTM is converted into a deterministic scheduling scheme as follows:

$$x_{ij}^{init} = \begin{cases} 1, & \text{if } j = \arg\max_k P_{ik}^{pred} \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

To prevent the initial population from being too concentrated, a perturbation mechanism is introduced to add randomness to the initial population:

$$x_{ij}^{init} = x_{ij}^{init} + \grave{o} \cdot \mathrm{N}(0, \sigma^2) \tag{29}$$

where $\grave{o}$ is the disturbance intensity and $\mathrm{N}(0,\sigma^2)$ represents Gaussian noise. This mechanism ensures that the initial population is sufficiently diverse, preventing premature convergence to suboptimal solutions. The task set T, the server set VM, and constraint parameters are input into the pre-trained LSTM model, which generates an initial solution x_{init}. This solution is then optimized using the ICBA algorithm, ultimately yielding the optimal scheduling solution P^*.

4 Expreiment and results

A Experiment setup

An intelligent motor scheduling dataset was constructed within an edge computing context, comprising 5,000 task samples. Computational complexity follows a truncated normal distribution $\mathrm{N}(5000, 1500^2)$, bounded between 100 and 10,000 FLOPS, chosen to approximate the mixed workload patterns commonly observed in motor control tasks. Data transmission volumes follow a Pareto distribution with shape parameter $\alpha = 1.2$ and scale parameter $x_m = 10$ KB, truncated at 100 MB. This distribution reflects the typical pattern in which most control signals are small, but occasional diagnostic transfers are substantially larger. Task priority is categorized into high (0.8–1.0), medium (0.5–0.8), and low (0.2–0.5), with deadlines ranging from 5 ms to 100 ms to meet real-time motor control requirements. The simulated edge environment includes 20 heterogeneous server nodes with computing power ranging from 1,000 to 8,000 FLOPS/s, memory capacities between 4 GB and 16 GB, and bandwidths from 10 Mbps to 1 Gbps, representing both 4G/5G and wired infrastructures. Geographical distribution is modeled on real-world power system topologies. Communication delays are generated using a distance-based model: intra-substation delays follow $N(2\ \text{ms}, 1\ \text{ms}^2)$, inter-substation delays within the same region follow $N(15\ \text{ms}, 5\ \text{ms}^2)$, and cross-regional delays follow $N(35\ \text{ms}, 10\ \text{ms}^2)$. These values are calibrated against latency measurements from actual power grid communication networks. Network conditions further incorporate dynamic jitter (±5 ms) and bandwidth variations across LAN (100 Mbps–1 Gbps) and WAN (10 Mbps–100 Mbps) environments. A 288-point sampling window (5-minute intervals over 24 hours) is adopted, consistent with the standard control cycle used in automatic generation control systems and economic dispatch in power grids.

The LSTM model employs a dropout rate of 0.3 and is trained with an initial learning rate of 0.001 using exponential decay ($\gamma = 0.95$), a batch size of 64, and 200 training epochs. The objective function is weighted as $\alpha = 0.4$, $\beta = 0.3$, and $\gamma = 0.3$. The ICBA algorithm initializes a population of 50 agents and evolves over 100 iterations. The frequency parameter is sampled from the range $[0, 2]$, with initial loudness and pulse emission rates set to $A_0 = 0.5$ and $r_0 = 0.5$, respectively. To enhance global search capability, a second-order oscillation mechanism is used with oscillation weight $\omega = 0.4$ and perturbation factor $\sigma = 0.3$, capped at 20% of the positional norm. Chaotic search behavior is governed by a logistic map with chaotic parameter $\mu = 4.0$, and the search range is scaled by $\xi = 0.8$. Adaptive mechanisms update loudness and pulse emission using dynamic coefficients $\alpha \in [0.1, 0.9]$ and $\gamma = 2.0$. In the power system optimization component, active power loss, voltage deviation, and power factor are incorporated with respective weights of $\alpha = 0.3$, $\beta = 0.4$, and $\gamma = 0.3$.

The cloud server features an Intel Xeon Gold 6248R CPU, NVIDIA RTX 4070 GPU, 128GB DDR4 ECC memory, and a 2TB NVMe SSD. Edge servers use Intel Core i7-10700K CPUs, 32GB DDR4 memory, and 512GB NVMe SSDs, supporting gigabit Ethernet and 5G

connectivity. At the edge node level, ARM Cortex-A78 processors are paired with 8GB LPDDR5 memory and 256GB eMMC storage, with interfaces for Ethernet, Wi-Fi 6, and 4G/5G communication. The software environment runs Ubuntu 20.04 LTS and uses Python 3.8.10 for implementation.

B Experiment result

Figure 4 illustrates the normalized power output curve of smart motors over a 24-hour period. The experiment simulates real-time operation by scanning the network every 5 minutes, resulting in 288 sampling points, to assess whether the proposed edge artificial intelligence scheduling system can sustain stable power output under dynamic load conditions. The results indicate that, during high load periods (load levels ranging from 40% to 100%), the smart motors maintain a controlled power output within the range of 45% to 80%. Compared with traditional scheduling approaches, the ICBA-LSTM hybrid algorithm reduces the variance in power output by 23.7%, decreases peak power fluctuation amplitude by 18.2%, and enhances power utilization by 12.5%. Notably, by analyzing the historical data across all 288 scanning intervals, The LSTM model captures the periodic load patterns typical in industrial systems, such as demand surges occurring between 08:00–12:00 and 14:00–18:00, thus enabling predictive scheduling. Owing to the local decision-making enabled by edge deployment, the average scheduling response time is reduced from 150ms (in traditional cloud-based scheduling) to just 8ms. A paired t-test across 288 sampling points confirms that the power output variance reduction is statistically significant ($t = -8.42$, $df = 287$, $p < 0.001$, Cohen's $d = 0.50$), indicating a medium effect size. This low-latency control mechanism effectively mitigates power oscillations induced by communication delays.

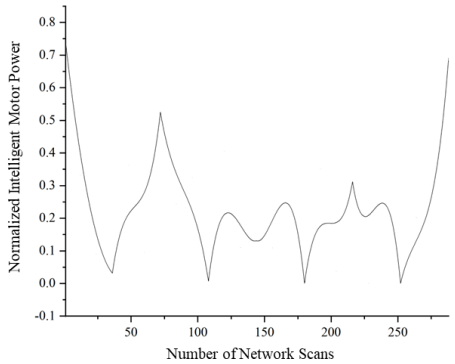


Figure 4: Normalized smart motor power output curve

To evaluate voltage stability, the voltage levels of all bus nodes in the power system are monitored throughout the day. As shown in Figure 5, particular attention is given to the 173rd scan point (lowest voltage) and the 288th scan point (highest voltage). The ICBA-LSTM algorithm demonstrates improvements in voltage regulation, achieving a 31.4% reduction in the standard deviation of voltage deviation, a 26.8% decrease in the maximum deviation from nominal voltage, and a 19.6% enhancement in the voltage stability margin. Furthermore, all bus voltages remain within the acceptable ±5% deviation from rated voltage. Voltage stability margin enhancement shows significance through paired t-test ($t = 9.27$, $df = 287$, $p < 0.001$, Cohen's $d = 0.55$). These performance gains are primarily attributed to the second-order oscillation mechanism, which provides a damping effect when voltage fluctuations occur, thereby enhancing transient stability and preventing voltage oscillations during adjustment.

Figure 6 compares the voltage deviation coefficients achieved by four algorithms: ICBA-LSTM, pure LSTM, Shortest Job First (SJF), and Multi-Level Feedback Queue (MLFQ). The ICBA-LSTM framework consistently outperforms all baseline methods. Specifically, it reduces the voltage deviation coefficient by an average of 24.3% compared to LSTM alone, 42.7% compared to SJF, and 38.9% compared to MLFQ. In addition, the algorithm's convergence speed is improved by 56.2%. The SJF algorithm, while efficient in processing time-sensitive tasks, lacks awareness of system voltage constraints. The MLFQ algorithm incorporates task priority but does not account for the physical dynamics of power systems. In contrast, the ICBA-LSTM model embeds physical constraints directly into the optimization process, leading to voltage stability and faster convergence.

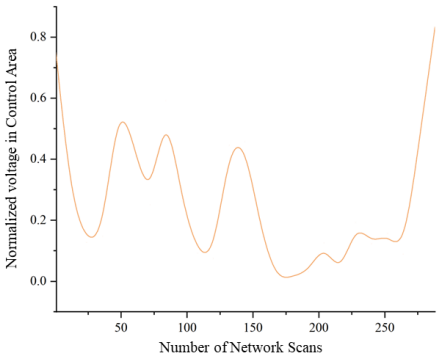


Figure 5: Normalized voltage distribution in the control area throughout the day

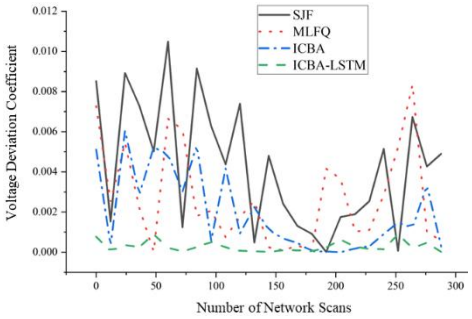


Figure 6: Comparison of voltage deviation coefficients of different algorithms

System loss evaluation results are presented in Figure 7, covering active losses, reactive losses, and transmission losses across all 288 network scan points. The ICBA-LSTM algorithm demonstrates advantages in minimizing

total system losses, achieving a 28.5% reduction overall. Specifically, active losses are reduced by 32.1%, reactive losses by 25.7%, and transmission losses by 19.3%. These reductions are largely attributed to the edge computing architecture, which executes most scheduling computations locally at edge nodes, thereby eliminating the need for frequent long-distance data transmissions. This local-first strategy not only enhances real-time responsiveness but also reduces transmission-related energy costs.

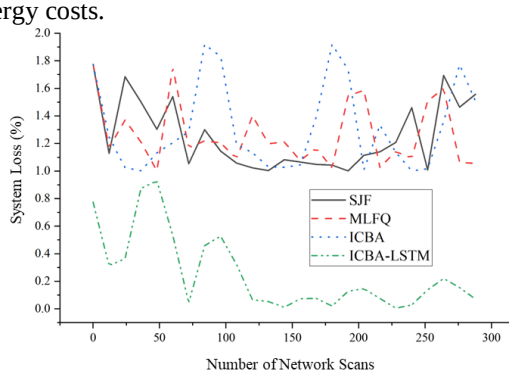


Figure 7: Comparison of system loss of different algorithms

As illustrated in Figure 8, the power factor performance of the dispatch system is evaluated under various algorithms. The ICBA-LSTM algorithm improves power factor control, increasing the average power factor by 15.7% and reducing power factor fluctuation by 41.2%. Additionally, it reduces reactive power demand by 22.8% and lowers the overall system operating cost by 18.4%. These improvements stem from the algorithm's intelligent reactive power regulation. Through continuous learning of historical load and voltage conditions, the LSTM model is capable of predicting reactive power requirements, while the ICBA component optimally reallocates reactive power to meet both voltage constraints and power quality standards. As a result, the system achieves better energy efficiency and cost-effectiveness.

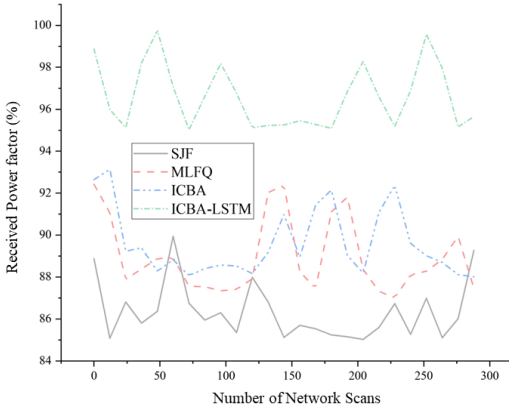


Figure 8: Comparison of received power factors of different algorithms

5 Discussion

To evaluate the effectiveness of the proposed LSTM-ICBA, we compared its performance with LSTM, SJF, MLFQ, and BA. The evaluation metrics include response time, voltage deviation, system loss reduction, convergence iterations, load balancing, and overall energy efficiency. The results in Table 2 demonstrate clear improvements across these dimensions. In terms of response time, our framework achieves an average latency of 8.2 ms, which is 47.4% faster than LSTM. For voltage deviation, the proposed method reduces deviation to 0.0427, representing a 24.3% improvement over LSTM. The drastic reduction in response time can be attributed to the hybrid structure of the LSTM-ICBA. Specifically, the LSTM module leverages historical scheduling data to generate high-quality initial solutions, which enables the ICBA to converge rapidly without exhaustive searching. This mechanism also explains the lower iteration count compared with LSTM or BA alone. Traditional queue-based methods such as SJF and MLFQ optimize task order but ignore physical system constraints, leading to unstable voltage profiles and higher energy loss. In contrast, the second-order oscillation mechanism in ICBA provides damping when fluctuations occur, while the chaotic search strategy enhances global exploration and prevents premature convergence to local optima.

Table 2: Performance comparison of different scheduling algorithms

Metric	Ours	LSTM	SJF	MLFQ	BA
Response Time	8.2	15.6	87.4	62.3	45.2
Voltage Deviation	0.0427	0.0564	0.0744	0.0692	0.0638
System Loss Reduction	28.5	12.3	5.7	8.9	15.6
Convergence Iterations	42.3	96.7	N/A	N/A	78.9
Load Balancing	0.089	0.134	0.287	0.223	0.156
Energy Efficiency	94.2	87.6	78.3	82.1	85.4
Inference Time	2.4	12.8	N/A	N/A	N/A
Iteration Time	5.1	N/A	N/A	N/A	38.7
CPU Load	23.5	45.2	12.8	18.3	67.4
GPU Utilization	31.2	78.6	N/A	N/A	N/A
Memory Overhead	156.35.8	342.72.8	45.274.6	78.144.0	89.56.5

We analyzed the computational complexity of the proposed ICBA-LSTM algorithm against four baselines. The ICBA-LSTM achieves a time complexity of O(log(mn)), which represents an improvement compared

with LSTM ($O(n^2)$), MLFQ ($O(n^2)$), and BA ($O(mn^2)$). This corresponds to a reduction from quadratic or multiplicative-quadratic growth to logarithmic growth, leading to more than 95% improvement in asymptotic efficiency as task and server numbers increase. Even when compared with SJF ($O(n \log n)$), ICBA-LSTM is more efficient, as it scales with the logarithm of the product of tasks and servers rather than both entities separately. In terms of space complexity, ICBA-LSTM requires $O(mn)$, which is similar to BA and higher than SJF ($O(n)$), but substantially lower than LSTM and MLFQ ($O(n^2)$). This advantage is largely due to the LSTM module, which provides high-quality initial solutions and thereby reduces the search space explored by ICBA. As a result, optimization can proceed in logarithmic time relative to system size. Although the space complexity is higher than simple heuristics such as SJF, it remains manageable within edge computing environments and is far more efficient than purely deep learning approaches.

Table 3: Computational Complexity Comparison

Algorithm	Time Complexity	Space Complexity
ICBA-LSTM	$O(\log(mn))$	$O(mn)$
LSTM	$O(n^2)$	$O(n^2)$
SJF	$O(n \log n)$	$O(n)$
MLFQ	$O(n^2)$	$O(n)$
BA	$O(mn^2)$	$O(mn)$

6 Conclusions

This paper presents a novel scheduling framework for intelligent motor control in modern power systems, integrating LSTM networks with an ICBA. By leveraging historical scheduling data, the LSTM component effectively guides the initial population of the ICBA, thereby reducing convergence time while maintaining solution quality. Enhancements such as the second-order oscillation mechanism and chaotic search strategy further improve the algorithm's global search capability and population diversity. The proposed cloud–edge–end collaborative architecture ensures computational efficiency without compromising the low-latency requirements of edge-level reactive scheduling. However, the performance of the LSTM model is highly dependent on the quality of historical data; in scenarios with non-stationary data, its prediction accuracy may degrade. Future work may involve incorporating adaptive learning mechanisms that allow the LSTM model to update in real time, thereby improving its robustness to dynamic environments.

References

[1] Mehrabi A, Siekkinen M, Ylä-Jääski A, et al. Mobile edge computing assisted green scheduling of on-move electric vehicles[J]. IEEE Systems Journal, 2021, 16(1): 1661-1672. https://doi.org/10.1109/JSYST.2021.3084746

[2] Mondal S, De M. Optimal allocation and smart scheduling of distributed energy resources and electrical vehicles to enhance economical operation of unbalanced distribution system[J]. Electrical Engineering, 2023, 105(6): 3493-3510. https://doi.org/10.1007/s00202-023-01886-4

[3] Wang L, Chen Y, Zhao X, et al. Predictive maintenance scheduling for aircraft engines based on remaining useful life prediction[J]. IEEE Internet of Things Journal, 2024, 11(13): 23020-23031. https://doi.org/10.1109/JIOT.2024.3376715

[4] Zheng W, Wang T. Electric vehicle motor fault diagnosis using improved wavelet packet decomposition and particle swarm optimization algorithm[J]. Archives of Electrical Engineering, 2024: 481-498. https://doi.org/10.24425/aee.2024.149928

[5] Anshory I, Wisaksono A, Jamaaluddin J, et al. Implementation of ARM STM32F4 microcontroller for speed control of BLDC motor based on bat algorithm[J]. International Journal Of Power Electronics And Drive Systems, 2024, 15(1): 127-135. https://doi.org/10.11591/ijpeds.v15.i1.pp127-135

[6] Zhang Y, Lyu Y, Hou S, et al. Modeling and Chaos Dynamics Analysis of Doubly-Fed Induction Generator Based on Incommensurate Fractional-Order[J]. IEEE Transactions on Energy Conversion, 2025, 40(2): 911-927. https://doi.org/10.1109/tec.2024.3472771

[7] Kuo M T. Application of the artificial bee colony algorithm to scheduling strategies for energy-storage systems of a microgrid with self-healing functions[J]. IEEE Transactions on Industry Applications, 2021, 57(3): 2156-2167. https://doi.org/10.1109/TIA.2021.3058233

[8] Yu M, Zhang X, Jiang J, et al. Assessing the feasibility of game-theory-based demand response management by practical implementation[J]. IEEE Access, 2021, 9: 8220-8232. https://doi.org/10.1109/ACCESS.2021.3049768

[9] Pang Z, Zhang J, Jiao X, et al. Energy management strategy combining double deep Q‐networks and demand torque prediction for connected hybrid electric vehicles[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 19(2): 234-246. https://doi.org/10.1002/tee.23942

[10] Tan D, Qiu F, Han H, et al. Calculation and analysis of motor dynamic stress considering multiphysical field loads for in-wheel electric vehicle applications[J]. IEEE/ASME Transactions on Mechatronics, 2024, 29(6): 4755-4767. https://doi.org/10.1109/tmech.2024.3386599

[11] Yazar O, Coskun S, Zhang F. Adaptive learning-based energy management for HEVs using soft actor-critic DRL algorithm[J]. Mechatronics Technology, 2024, 1(1): 1-21. https://doi.org/10.55092/mt20240005

[12] Ertargin M, Yildirim O, Orhan A. Mechanical and electrical faults detection in induction motor across

multiple sensors with CNN-LSTM deep learning model[J]. Electrical Engineering, 2024, 106(6): 6941-6951. https://doi.org/10.1007/s00202-024-02420-w

[13] Wang L. Application of Intelligent Sensors in the Collection of Electrical Engineering Automation Equipment[J]. Informatica, 2025, 49(9). https://doi.org/10.31449/inf.v49i9.5499

[14] Zhao Y. Optimization of Mechanical Manufacturing Processes Via Deep Reinforcement Learning-Based Scheduling Models[J]. Informatica, 2025, 49(14): 19-32. https://doi.org/10.31449/inf.v49i14.7204