IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 November 2025, accepted 15 December 2025, date of publication 23 December 2025, date of current version 31 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3647686

SURVEY

Radio Signals Recognition With Unsupervised Deep Learning: A Survey

LJUPCHO MILOSHESKI[1,2], (Member, IEEE), BLAŽ BERTALANIČ[1], (Member, IEEE), CAROLINA FORTUNA[1], AND MIHAEL MOHORČIČ[1,2], (Senior Member, IEEE)

[1]Department of Communication Systems, Jožef Stefan Institute, 1000 Ljubljana, Slovenia
[2]Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

Corresponding author: Ljupcho Milosheski (ljupcho.milosheski@ijs.si)

This work was supported by Slovenian Research and Innovation Agency under Grant P2-0016.

ABSTRACT Optimization of wireless network parameters relies on the awareness of a dynamically changing radio environment, which depends on the presence of active devices characterized by various radio access technologies (RATs), modulation schemes, and overall spectrum usage patterns, and can be determined by advanced radio signal recognition methods. While various supervised machine learning (ML) models have been explored for signal recognition, their actual deployment has been limited so far due to challenges in acquiring labeled datasets. The emergence of Open Radio Access Network (O-RAN) architectures and open experimental testbed setups has enabled access to large-scale, unlabeled data through standardized interfaces, paving the way for unsupervised deep learning methods. These methods, unlike supervised approaches, require minimal labeled data and have shown promising results in domains such as computer vision and time-series processing. However, their application in wireless communications remains relatively unexplored. This survey aims to provide a comprehensive overview of unsupervised deep learning techniques for addressing key challenges for signal recognition in wireless communications, including automatic modulation classification (AMC), signal sensing, specific emitter identification (SEI), and anomaly detection. Specifically, we examine state-of-the-art approaches such as deep clustering, contrastive learning, autoencoder-based reconstruction, and generative models. Additionally, we discuss available open datasets and identify research opportunities to advance this field, leveraging the substantial successes of self-supervised learning in computer vision and natural language processing. By organizing the survey into two key complementary perspectives—wireless communication challenges and unsupervised deep learning solutions—this work provides a roadmap for researchers and practitioners seeking to develop innovative, data-efficient models for the next generation of AI-native wireless networks.

INDEX TERMS AMC, anomaly detection, radio signal recognition, SEI, signal sensing, spectrum sensing, unsupervised deep learning, wireless communications.

I. INTRODUCTION

Radio signal recognition concerns recognizing the presence, type, and characteristics of signals transmitted over the radio frequency (RF) spectrum. It is an essential task in wireless communications and spectrum management, supporting a variety of challenges. In this survey, we focus on four closely related and sometimes overlapping subdomains: modulation classification [1], signal sensing [2], specific emitter identification (SEI) [3], and anomaly detection [4]. These areas have been widely explored by the research community over the past decade, with a significant emphasis on deep learning approaches in the last five years. The awareness of the radio environment is important for optimizing various network parameters, such as selecting appropriate frequency channels, adjusting transmission power, choosing modulation types, and determining suitable radio access technologies. Timely information on the overall spectrum usage and activity from different aspects, such as the number of active devices, used radio access technologies (RATs), modulations, and

The associate editor coordinating the review of this manuscript and approving it for publication was Qiang Li.

anomalous signals, could enable more efficient and effective management of wireless networks, ensuring their stable performance and minimizing mutual interference.

Sensing capabilities offer significant benefits for both licensed and unlicensed spectrum networks, as direct feedback from the environment, for the purpose of network parameter control. In unlicensed spectrum, technologies like WiFi, LoRa, and Bluetooth use sensing to inform channel occupancy [5] to guide the selection of less congested channels, and detect interfering transmissions, which is increasingly common with the proliferation of wireless devices in everyday life.

Licensed wireless networks could benefit significantly from such environmental awareness, achieved through the different signal recognition capabilities. These proved the potential to further increase the efficiency of the spectrum usage, following the successfull applications in the past, such as the License Assisted Access (LAA) [6] at 2.4 GHz and 5 GHz providing coexistance of LTE with WiFi, as well as the 4G/5G sharing the 3.5 GHz Citizens Broadband Radio Service (CBRS) band with military radar [7]. This is highly relevant for the developments related to next-generation wireless networks such as 6G, considering that Integrated Sensing and Communication (ISAC) is envisioned as one of their key capabilities [8], [9]. Furthermore, signal recognition outcomes could be used for network parameters control and optimization, for objectives like interference detection (via spectrum sensing), improved security (through SEI), and traffic steering (based on RAT recognition). This is particularly the case in private network deployments of 5G networks, such as office buildings, and factories where the operator has no control of transmissions in similar neighbouring deployments. The utilization of such functionalities is enabled by Open Radio Access Network (O-RAN) [10], considering that they could easily be deployed as software applications within the non-real-time and near-real-time controllers (xApps and rApps) [11], leveraging open and standard interfaces [12] for data collection.

In addition to direct network parameter control, historical network monitoring data obtained through signal recognition functionalities can be used to build detailed environmental maps of network performance. They can form the basis for multimodal Digital Twins [13], i.e., virtual replicas of the network with its components and metrics such as latency and uplink/downlink speeds. These go beyond the existing concepts that provide only signal coverage [14], which is highly important for network planning and monitoring, by including multiple planes of data, such as mapping the uplink and downlink speed to the physical environment, thus paving the way for the emerging throughput intensive and low-latency [15] Extended Reality (XR) technologies [16].

All four subdomain tasks rely on processing radio signals, either in their raw format consisting of an in-phase and a quadrature (I/Q) component, usually utilized for SEI and AMC, or in transformed forms such as spectrograms based on FFT or wavelet transforms. These provide a

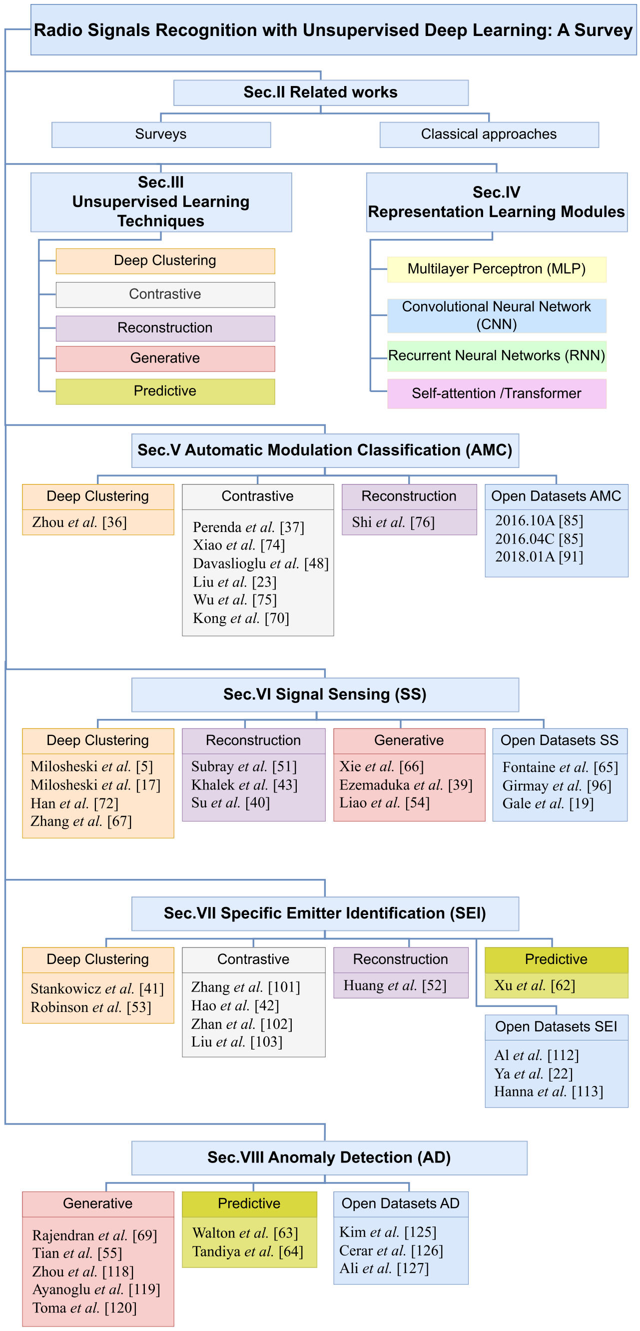


FIGURE 1. Overview of the survey structure, illustrating with color code the relationships between sections and the mapping of the surveyed works based on their employed approaches and addressed challenges.

pictorial format particularly suitable for existing image processing algorithms, thereby facilitating sensing capabilities. ML models are particularly suitable for these tasks [17] due to their capability to capture high-level correlations between temporal patterns, which in such use-cases of radio data are difficult or even impossible to detect with standard energy detection [18], [19] or cyclo-stationary momentum-based detectors [20], [21]. However, this comes at a price of increased computational complexity and lower explainability.

While the use of machine learning (ML), including Deep Learning (DL), for signal recognition-related tasks is not new, actual deployment in real-world systems is still in its early stages. Supervised learning approaches have been extensively validated, and offer promising results [22], however, they face several limitations. Application is limited by the need for large labeled datasets, which are costly and time-consuming to obtain [19]. Furthermore, they face generalization problems due to the constraints of the labelled dataset sizes. Unsupervised deep learning, which includes several approaches that learn based on pretext tasks derived from the data itself, emerges as a promising alternative. It addresses the generalization and labeling problem by achieving comparable results to supervised methods while requiring only a fraction of the labeled data. This is due to the capability of automatic learning by processing large amounts of unlabeled data. These are now widely accessible, either generated from many available testbeds operating in the unlicensed bands or through the standardized O-RAN interfaces. As a result, unsupervised learning serves as a practical solution for initial model development, with supervised learning reserved for targeted fine-tuning in specific deployment scenarios [23].

A. CONTRIBUTIONS

The unsupervised deep learning approaches have matured in fields such as machine vision [24], [25], and general time series processing [26]. However, it is still gaining traction in the wireless communications domain, with limited existing research, given that its adoption is not straightforward, both in its raw and transformed formats. It requires communications-specific alterations, such as special data augmentation functions or feature extraction kernels, e.g., convolutional neural networks (CNNs), with particular capabilities, which are tailored for the specific semantics contained in the radio signals.

To the best of our knowledge, no existing survey summarizes recent advances in unsupervised deep learning for wireless communications while systematically categorizing learning approaches and their relation to key signal recognition challenges. This work addresses that gap by providing a comprehensive overview of deep unsupervised learning approaches applied to wireless communications, along with open datasets to support and inspire further research. Given the remarkable progress in computer vision and natural language processing through self-supervised models, exploring their potential in this domain is both timely and promising. In the most general sense, our work focuses on a review of the existing work from two perspectives:

1) **Review of core wireless signal recognition-related challenges addressed with unsupervised deep learning**: These challenges include AMC, signal detection, SEI, and anomaly detection. Considering that these challenges are well-known and established research fields in the domain, each backed by a large corpus of publications, we adopt the existing categorization published in [28].
2) **Review of unsupervised deep learning solutions in communications**: Such approaches include deep clustering, contrastive learning, autoencoder-based reconstruction, and generative models in relation to the key signal recognition tasks within wireless communications. Specifically, we further build on the existing segmentation provided in [26]. For each specific research challenge, we examine which unsupervised learning approaches have been applied to improve performance.

These two analytical review perspectives ultimately lead to the final and most important contribution of our work as a survey, i.e., the **identification of open research problems and future directions** regarding the application of unsupervised deep learning to the wireless communications domain. More specifically, in addition to surveying current challenges and approaches, our work highlights existing gaps and unresolved issues where unsupervised learning methods could offer innovative solutions.

B. STRUCTURE OF THE PAPER

In Figure I, we summarize the structure of the entire survey, indicating the main parts and their relations, which will be followed through the discussions in the text. Section II reviews existing surveys and related work, highlighting their contributions and limitations while positioning our study within this research landscape. Section III introduces the fundamental unsupervised deep learning approaches applied in wireless communications, including deep clustering, contrastive learning, reconstruction-based methods, generative models, and predictive learning. Section IV describes the commonly used types of learning modules for those approaches. Sections V-VIII review the most recent works, report on open datasets, and discuss potential future work regarding signal recognition and classification for AMC, signal sensing, SEI, and anomaly detection. Section IX highlights the observed challenges and potential future work directions, while Section X presents the conclusions of this study.

II. RELATED WORK

Several surveys have addressed ML applications in radio signal recognition, focusing on various aspects. However, a closer look shows that the unsupervised deep learning is treated only marginally, either as a subcategory of deep learning, [27], [28], [30], [31], [32], or completely omitted due to the focus on particular application domains such as dynamic spectrum access [29], and attacks modeling [33]. This leaves a notable gap in the coverage of unsupervised deep learning methods for radio frequency signals, which in recent years have become increasingly relevant, considering their potential for automatic signal representation learning, as well as the availability of large open datasets.

A. GENERAL ML IN COMMUNICATIONS

Among the most comprehensive surveys in the field, [30] provides an extensive overview of deep learning algorithms applied to mobile and wireless networking, along with the enabling technologies supporting their deployment. The authors systematically review state-of-the-art models and offer a fine-grained categorization linking specific algorithms to the problems they address, thereby serving as a valuable reference for researchers and practitioners.

While slightly dated, [31] remains an influential and insightful contribution, offering a broad overview of deep learning applications in wireless networks from an application perspective. The survey organizes prior work according to the layers of the communication network stack, systematically analyzing how deep learning techniques have been applied to the physical, data link, network, and higher layers. This layered categorization provides a useful framework for understanding the scope and diversity of deep learning use cases in wireless systems. A related effort by [32] adopts a similar structure but focuses on applications within the Internet of Things (IoT) domain, addressing comparable challenges from a connectivity and device-integration standpoint. While both surveys introduce the main algorithmic paradigms, including supervised and unsupervised learning, their treatment of unsupervised deep learning remains high-level and mostly conceptual.

In [28], the authors provide a clear categorization of the existing challenges that are being addressed in the domain and a comprehensive overview of existing works from the perspective of the used dataset. They also provide insightful recommendations on the creation of future radio datasets. We take into consideration this categorization and complement it by reviewing the approaches that are used to address the discussed challenges, focusing on the unsupervised learning approaches, which we believe are highly relevant.

B. RADIO SIGNAL RECOGNITION

In the recent work [27], the authors provide a broad overview of machine learning (ML) approaches for spectrum sensing, focusing on supervised learning techniques, while paying limited attention to unsupervised deep learning, restricted to a brief mention of autoencoders. Such coverage overlooks other major unsupervised paradigms, such as contrastive learning and deep clustering, which have recently demonstrated strong potential for representation learning without labeled data [36], [37]. Furthermore, authors highlight the importance of large-scale radio frequency (RF) datasets for developing pre-trained models and enabling transfer learning; however, the discussion remains general and lacks concrete dataset references. As summarized in Figure I, we analyze these aspects in depth throughout the sections of this survey.

In [29], the authors provide a concise overview of the sensing-related challenges in the radio frequency (RF) domain addressed through machine learning (ML), focusing on cognitive radio functions such as spectrum sensing and dynamic spectrum access. The review includes supervised and reinforcement learning approaches, and, unlike our work, they offer no examination of unsupervised deep learning approaches from the perspective of their relevance for the pretraining phase of supervised approaches. This aspect is covered in our work in Sections V and VII.

In a similar manner, [35] provides systematic coverage of AI-based methods for the tasks of automatic modulation classification (AMC), signal detection, channel estimation, and MIMO beamforming. The work systematically reviews AI-based methods and highlights their potential for enabling intelligent and adaptive physical-layer processing. However, it gives limited attention to unsupervised deep learning approaches, which we systematically analyze in Section III and review in Sections V to VIII of this work as depicted in Figure I.

The most closely related work to ours is [34], which provides one of the first comprehensive surveys dedicated to unsupervised ML techniques in networking. Although dated (2019), it remains a valuable reference, as it captures fundamental challenges and methodological directions that are still relevant to current research. However, given the rapid evolution of deep learning in recent years, the survey is gradually becoming outdated, considering the recent emergence of transformer-based architectures and the contrastive learning paradigm, both of which have driven significant progress in self-supervised representation learning for radio signal processing.

C. OUR WORK

Considering the existing categorization of the domain-related challenges in prior reviews, the classification of unsupervised deep learning methods from closely related fields [24], [33], [38], and the identified gap in reviewing the recent advances in their use in radio signal recognition, we are focusing on these approaches from the perspective of wireless networks communications.

Summary of the related works, considering their key contents and shortcomings, as well as how we complement them in our work, is given in Table 1.

D. CLASSICAL UNSUPERVISED METHODS

While recent research in unsupervised learning increasingly emphasizes deep learning techniques due to their superior performance in feature extraction, classical unsupervised methods such as Energy Detector (ED), K-means clustering and its variants, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) remain highly relevant. These traditional approaches are frequently used as benchmarking baselines [5], [36] or as components within more advanced frameworks, such as deep clustering [17], where they are combined with neural network-based feature extractors. The following section provides a concise overview of recent works that deploy non-deep learning unsupervised methods.

TABLE 1. Summary of related works concerning unsupervised learning for wireless communication.

Ref.	Focus Area	Strengths	Limitations	Our Work
[27]	ML-based spectrum sensing with emphasis on supervised learning	Comprehensive overview of supervised learning and importance of RF datasets for pre-trained models	Brief coverage of unsupervised learning (only mentions Auto-encoders); lacks references to datasets/testbeds	Provides a detailed review of unsupervised learning techniques (e.g., contrastive learning, deep clustering) and their application to RF signal recognition challenges
[28]	Challenges in spectrum sensing and dataset usage	Clear categorization of challenges; recommendations for future dataset creation	Limited discussion on unsupervised deep learning	Builds on the dataset categorizations and challenges outlined, focusing on how unsupervised learning addresses these issues
[29]	Spectrum sharing in cognitive radio	Relevant to environmental adaptiveness and 6G ISAC concept	Short and shallow mention of unsupervised deep learning	Highlights the potential of unsupervised learning to enhance adaptability and efficiency in 6G spectrum sharing
[30]	Deep learning algorithms for wireless communication	Most comprehensive survey; fine-grain categorization of ML problems	Outdated coverage of unsupervised learning due to publication year	Expands on unsupervised learning advances (e.g., generative models, self-supervised techniques) not covered extensively in prior works
[31]	ML applications in network architecture	Covers relevant research topics from the application perspective	Outdated (6 years old); lacks specifics on unsupervised deep learning	Connects unsupervised learning to well-defined application challenges
[32]	IoT perspective on ML in wireless communication	Domain-related insights; discussion on algorithmic approaches	Fundamental coverage without details on unsupervised deep learning	Focuses on unsupervised deep learning approaches, considering the latest advances and existing datasets for further research relevant for radio signal recognition in a broader context
[33]	Adversarial Machine Learning in wireless communications	Insights into unsupervised ML for modeling security threats	Focuses on the specific use case of attacks modeling	Reviews unsupervised learning approaches for signal recognition in wireless communications
[34]	Domain challenges and unsupervised learning	Covers unsupervised approaches; highlights challenges and data-driven models	Slowly becoming outdated due to recent advancements concerning transformers and contrastive learning	Highlights gaps in most recent works; emphasizes open datasets crucial for development of unsupervised learning
[35]	Intelligent signal processing in wireless communications	Covers four topics for the physical layer, modulation classification, signal detection, beamforming, and channel estimation	Covers machine learning approaches in a wider sense, without focus on unsupervised deep learning	Focuses on unsupervised deep learning approaches for specific wireless communications signal recognition challenges

The Energy Detector is one of the simplest and most widely used methods for signal detection, especially in spectrum sensing. It operates by measuring the energy of a received signal over a predefined observation interval and comparing it to a threshold. If the measured energy exceeds this threshold, the presence of a signal is assumed; otherwise, the channel is considered idle. Its performance is highly sensitive to noise uncertainty and can degrade significantly in low Signal-to-Noise Ratio (SNR) conditions. In recent works, it is usually deployed as a baseline approach for benchmarking purposes [39], [40].

K-means clustering is a partition-based clustering algorithm that aims to divide a dataset into K non-overlapping clusters. It is simple to use and requires only the number of clusters as a single parameter for initialization, which is independent of the feature space. Thus, it is the most frequently used non-deep learning approach, either as part of more complex deep-learning architectures [5], [17] such as deep clustering, or as a benchmarking baseline in its original form [5], [36], [41], [42]. Other clustering algorithms, for instance DBSCAN [41], [42], GMM [43], Agglomerative and Spectral clustering [36], are typically used as baselines due to their higher complexity initialization parameters that require feature-space specifics, such as distance between samples.

In Figure 2, we show the performance of the most commonly used classical unsupervised methods given as relative performance to the deep learning-based counterparts, proposed in the corresponding works. On the x-axis are the reference numbers of related works. On the y-axis is the relative performance given as percentages of the top-performing deep-learning counterpart (marked on top of each bar) evaluated in similar conditions, considering the reported metrics (also marked on top of each bar) in the corresponding works, for each of the three reviewed radio signal recognition tasks. No such data was available for the reviewed works in the Anomalies Detection task; thus, we consider only the AMC, SS, and SEI tasks. As can be seen from the plot, the classical approaches consistently show lower performance compared to the deep learning-based approaches across all three tasks and a variety of metrics. Only two cases ([39] and [43]) reach a significant 80% of the performance of the deep learning counterpart proposed in the corresponding work. However, such performance is observed only in the binary classification problems (note that the metric is Pd or Pd_{AUC}), while for the more complex tasks with multiple classes, the performance is significantly worse.

III. UNSUPERVISED DEEP LEARNING TECHNIQUES IN WIRELESS COMMUNICATIONS

Clear categorization of the Unsupervised Deep learning algorithms is still an open discussion, depending on the perspective from which the approaches are being observed. For example, in [44], authors make a distinction based on the way the pseudo-labels are generated, while in [26], authors

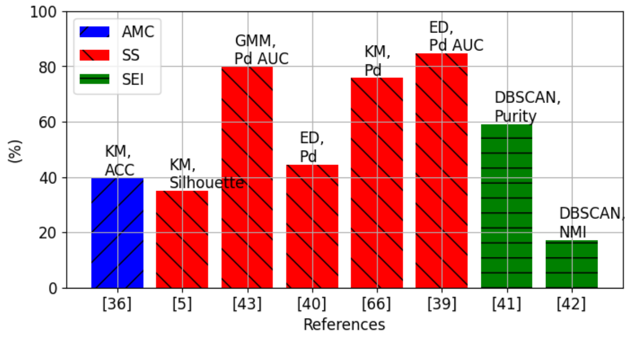


FIGURE 2. Relative performance of classical unsupervised algorithms compared to the deep learning approaches in the surveyed works as reported in the literature. Labels on bar charts: Method, Metric. Methods: KM=K-means, GMM=Gaussian Mixture Model, ED=Energy Detector, DBSCAN=Density-Based Spatial Clustering of Applications with Noise. Metrics: Pd=Probability of Detection, NMI=Normalized Mutual Information, ACC=Clustering Accuracy, Purity=Clustering Purity.

consider the final training phase only. Considering these two publications and the pool of relevant surveyed works, we distinguish five different unsupervised deep learning approaches utilized in signal recognition in the wireless communications domain: *deep clustering*, *reconstruction-based*, *generative*, *contrastive*, and *predictive*. In the following, we briefly introduce each of them, without going into fine-grained details, for which some of the early original works are referenced.

A. DEEP CLUSTERING METHODS

Deep clustering methods [45], illustrated in Figure 3, integrate deep learning with clustering techniques to extract semantically meaningful representations from high-dimensional RF data and to group similar signal patterns in an unsupervised fashion. In this context, neural networks are trained to map raw or preprocessed RF inputs, such as I/Q samples, spectrograms, or just averaged FFT amplitudes, into a latent space where clustering algorithms (e.g., K-means or Gaussian Mixture Models) can distinguish between signal classes [17]. The training process jointly optimizes two objectives: (1) representation learning, which structures the latent space to capture key features of RF signal variability (e.g., hardware impairments in the I/Q components for emitter identity, or amplitudes patterns at different frequencies for signal sensing), and (2) unsupervised clustering, which promotes the separation and compactness of signal clusters in the learned embedding space [17].

B. CONTRASTIVE METHODS

Self-supervised contrastive learning [46], [47] is a powerful representation learning paradigm that enables the extraction of discriminative features from unlabeled data (including RF data) by contrasting positive and negative signal instances. The core idea is to maximize the similarity between positive pairs, which are derived from the same RF signal sample through domain-specific augmentations (e.g., zero-masking, time-shifts, noise injection [48]), while minimizing similarity with negative pairs, which correspond to embeddings of different RF signals. Both augmented views of the same signal are passed through a shared neural encoder to generate feature embeddings, followed by a projection head that maps these into a latent space where a contrastive loss function (e.g., InfoNCE [49]) is applied, as depicted in Figure 4. This contrastive objective encourages embeddings of similar RF patterns (e.g., same emitter or modulation scheme) to cluster together while pushing dissimilar signals apart [42], thereby enabling robust, generalizable representations useful for downstream tasks such as AMC and SEI.

C. RECONSTRUCTION-BASED METHODS

Reconstruction-based methods [50] are unsupervised learning techniques particularly well-suited for RF signal processing. The goal is to learn compact and informative representations by reconstructing input RF signal data from compressed or latent representations. The basic model is the autoencoder (AE), a symmetric encoder-decoder architecture, visualized with violet in Figure 5a. An encoder transforms raw RF signals, such as raw I/Q samples or spectrograms, into a lower-dimensional latent space, and a decoder attempts to reconstruct the original signal from this compressed representation. The objective is to minimize the reconstruction error, ensuring the learned representations capture the most salient features of the RF data, such as modulation, temporal pattern, or energy burst shapes across the frequency domain. These approaches are widely used for signal recognition tasks, such as RAT classification [51], primary user (PU) recognition [43], and denoising [40]. By optimizing the reconstruction, these baseline autoencoders provide a robust framework for learning compact and informative representations suitable for downstream tasks, such as SEI [52], after further tuning of the encoder in a supervised manner.

However, relying solely on reconstruction loss, the autoencoder may not necessarily produce a compact and interpretable feature space of the input RF signals. This lack of structure can make it difficult to analyze or utilize the encoded representations in a completely unsupervised

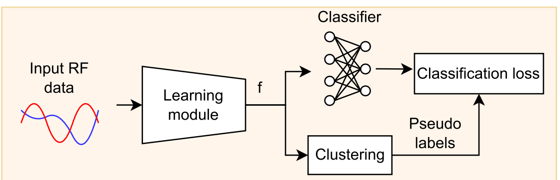


FIGURE 3. General structure of *Deep clustering* methods.

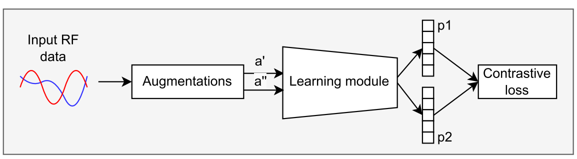
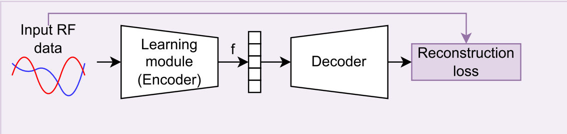
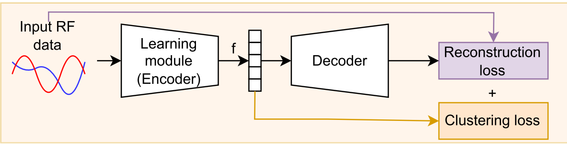


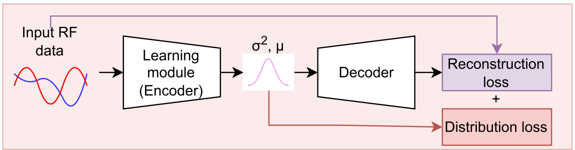
FIGURE 4. General structure of *Contrastive* learning methods.



(a) Autoencoder.



(b) Autoencoder-based Deep Clustering.



(c) Variational autoencoder.

FIGURE 5. General structure of ***Reconstruction-based*** learning methods.

setup [5], [19], where fine-tuning is not an option. To address this, introducing a clustering loss in the feature space f, as in [36], can guide the encoding process toward a more cluster-friendly distribution of samples. This adjustment makes the learned representations more suitable for further analysis, especially when labels are entirely absent. Such modifications in the learning process give rise to reconstruction-based deep clustering methods that are used for self-supervised RF signal feature learning as pretraining, for downstream tasks such as clustering for novel (unseen in training) device detection [41], [53]. Its functional structure is highlighted in orange in Figure 5b.

An important extension of the reconstruction-based approach for RF signal representation is the Variational Autoencoder (VAE), depicted in red in Figure 5c. Unlike standard autoencoders, which map each input RF signal to a fixed point in latent space, VAEs model the latent space as a multidimensional probability distribution, typically Gaussian. For each input, such as an I/Q sample stream [39] or spectrograms [54], [55], the encoder outputs parameters of a distribution (mean and variance) rather than a single deterministic code. The decoder then reconstructs the signal by sampling from this learned distribution. This probabilistic framework provides a continuous and structured latent space, which enables smooth interpolation between different RF signal samples and supports generative capabilities such as the synthesis of novel RF-like samples.

D. GENERATIVE METHODS

Generative methods [56] are a class of machine learning techniques designed to model the underlying data (including RF data) distribution and generate new, realistic samples that resemble the original dataset. These methods are widely used for tasks like image synthesis, data augmentation, and representation learning. Among the most prominent generative approaches besides VAEs discussed in the previous subsection are Generative Adversarial Networks (GANs). GANs generally consist of a generator that produces synthetic data and a discriminator that evaluates their authenticity, training them in an adversarial manner to improve the quality of generated samples as depicted in Figure 6.

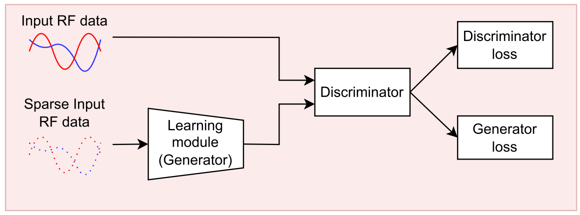


FIGURE 6. General structure of ***Conditional Generative Adversarial Networks.***

The GAN architecture has seen wide application in the physical layer of wireless communications [57], [58], usually in its conditional variant, cGAN, illustrated in Figure 6. In cGAN, instead of with random input, the generator is provided with sparse points of the actual input data. The use of GANs mainly focuses on pre-processing of radio signals, such as enhancing [59] and generating new data samples [60] for data-constrained use-cases. Such approaches could contribute to improving signal recognition [61]; however, many such solutions require labeled data for training, such as modified spectrograms in [54], thus positioning them as edge cases with regard to the scope of this survey.

E. PREDICTIVE METHODS

Self-supervised predictive learning is a representation learning method that leverages unlabeled data by training models to predict certain aspects or features of the data itself. As illustrated in Figure 7, they typically consist of a representation learning module followed by a classifier and a classification loss module.

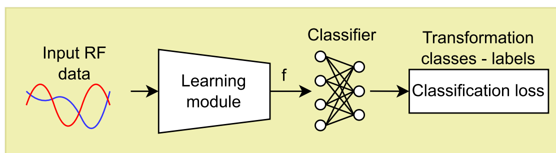


FIGURE 7. General structure ***Predictive*** methods.

In the RF domain, predictive self-supervised tasks are designed to exploit the temporal, spectral, and structural continuity inherent in RF signals. These tasks involve learning to predict transformations or future states of the signal, encouraging the model to internalize meaningful representations without labeled data. Typical examples include: predicting artificially applied transformations such as rotation classification [62]; forecasting the next I/Q symbol in a time series to model fine-grained temporal dependencies [63]; and predicting the next spectrogram frame to detect anomalies as deviations from expected spectral behavior learned during training [64]. These pretext tasks promote the learning of statistical regularities—such as modulation structure, spectral patterns, and temporal correlations in I/Q data—making the resulting representations highly informative for downstream applications like anomaly detection, AMC, or SEI.

F. OVERVIEW

We summarize the discussed unsupervised deep learning methods in Figure 8 using a Venn diagram to illustrate the relationship between them and to map the corresponding works in the wireless communications domain that will be surveyed through Sections V-VIII.

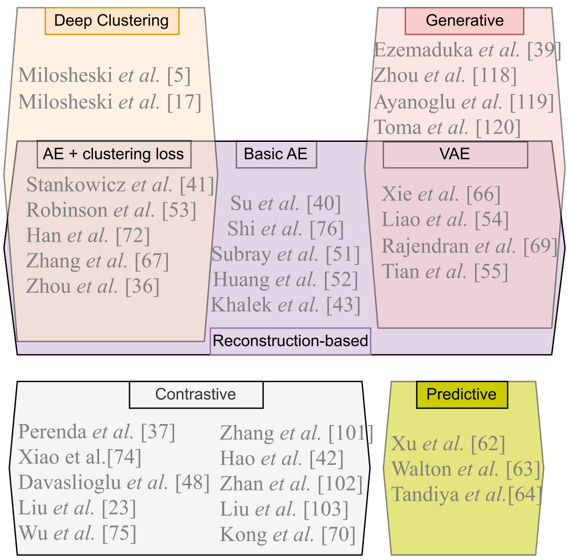


FIGURE 8. Unsupervised deep learning approaches. Color coding corresponds to Fig. I.

Each approach includes specific types of architectures, yet there are overlapping cases, particularly with autoencoders, which can be categorized differently depending on how their loss functions are modified. Standard autoencoders, as a reconstruction-based approach, consist of an encoder-decoder structure optimized using reconstruction loss. This provides a straightforward way to feature extraction in an unsupervised manner, capturing essential data representations in a low-dimensional latent space. Modifications to their loss functions improve different aspects of their functioning, thus placing them into different categories, even though they retain their fundamental structure and functionality of encoding input data into a lower-dimensional representation and reconstructing it at the output. Introducing a clustering loss in the latent space, in addition to the standard reconstruction loss (such as a distance-based loss), leads to the formation of more cluster-friendly feature spaces. This adaptation aligns autoencoders more closely with deep clustering methods. Alternatively, replacing the latent space loss with a statistical loss, while incorporating reparameterization to ensure gradient flow during backpropagation, transforms the model into a generative framework. This modification enables the generation of new samples by sampling from a smooth, continuous latent distribution, making it a VAE-based generative model.

As shown in Figure 8, contrastive and predictive learning approaches remain distinct categories that do not yet overlap with reconstruction-based, deep clustering, or generative models.

IV. NEURAL ARCHITECTURES FOR THE REPRESENTATION LEARNING MODULES

In this section, we shortly outline commonly used types of learning modules as part of previously described techniques in Section III.

A. MULTILAYER PERCEPTRON (MLP)

Multilayer Perceptrons are the most fundamental form of neural networks, consisting of fully connected layers where each neuron in one layer connects to all neurons in the next layer. They are composed of an input layer, one or more hidden layers, and an output layer, as visualized in Figure 9. Each layer performs a linear transformation followed by a nonlinear activation:

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}), \tag{1}$$

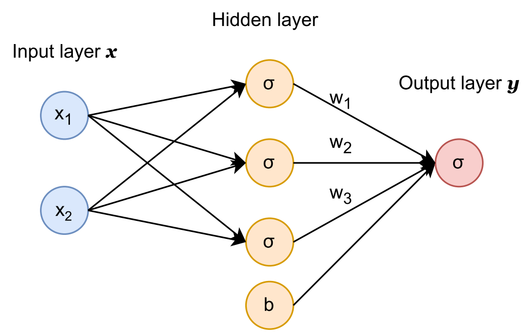


FIGURE 9. Multilayer Perceptron illustration.

where $\mathbf{x}$ is the input, $\mathbf{W}$ the weight matrix, $\mathbf{b}$ the bias, and $\sigma(\cdot)$ is the activation function (e.g., ReLU, Sigmoid).

In the wireless communications domain, MLPs are rarely used for direct processing of raw data, and usually only for low-density inputs, such as Received Signal Strength Indicator (RSSI) measurements. This is mainly because MLPs exhibit a high parameter count that increases rapidly with input dimensionality, resulting in substantial computational and memory demands. Furthermore, they are weak at capturing local dependencies, since they give equal weight to each input. However, they still offer decent performance [65], [66] if appropriate inputs, such as statistical moments, are provided. In recent works [5], [51], [67], the MLPs are mostly used as building blocks of more complex architectures, such as classification layers for the CNN-provided feature spaces, as will be discussed later in this section and visualized in Figure 13. Regarding the setup requirements, the MLPs are fairly easy to utilize since they require specification of a relatively small number of parameters. Besides the standard parameters which are common for most of the backpropagation-based learning architectures, such as optimizer, learning rate, batch size, etc., these include the activation function and the number of layers and neurons per layer.

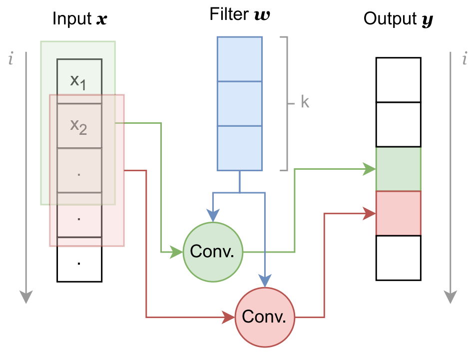


FIGURE 10. Single layer of 1D Convolutional Neural Network illustration.

B. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNNs are designed to process grid-like data, such as images or time series, by applying convolution operations to extract local features. A convolution layer applies filters (also referred to as kernels) that slide over the input, capturing spatial hierarchies. In the following, we formulate a one-dimensional (1D) CNN, also depicted in Figure 10, which is suitable for sequence processing. The extension to 2D, which is also commonly used when a sequence to image transformation is applied on the input data, is straightforward. Given an input sequence x and a filter (kernel) w of length k, the 1D convolution operation at position i is calculated as:

$$y(i) = \sum_{j=0}^{k-1} w(j) \cdot x(i+j), \tag{2}$$

where $y(i)$ is the output feature map at position i, $x(i+j)$ is the input value at position $i+j$ in the sequence and $w(j)$ is the weight of the kernel at position j. This operation is repeated across the entire sequence to form a complete output $\mathbf{y}$.

CNNs are the most widely used learning modules in the wireless communications domain, potentially as part of any of the discussed unsupervised deep learning approaches in Section III, due to their ability to capture and distill spatial correlations. Their hierarchical feature extraction provides strong generalization for structured inputs. However, their performance depends on more parameters, which have to be defined and are application dependent, such as filter sizes, stride, padding, and pooling strategy, which makes them more complex to set up compared to the MLPs. On the other side, they can be directly used on the raw data, avoiding manual feature engineering as it is usually needed with MLPs, which makes them more flexible and applicable.

There is a variety of existing implementations of CNN, which include different filter configurations and interconnection of layers (Visual Geometry Group (VGG), Residual Network (ResNet), GoogLeNet, etc.), enabling solutions of various complexity. While they exhibit great performance for time series or pictorial data processing and extraction of spatial correlations, they lack the ability to capture long-term dependencies, which are specific to I/Q signal components in wireless communications data. One way to address this is by increasing the receptive field by introducing dilated convolutions [68], or simply increasing the filter sizes.

C. COMMONLY USED TYPES OF RECURRENT NEURAL NETWORKS (RNN)

RNNs, depicted in Figure 11a are tailored for sequential data processing with variable length, where each neuron's output depends not only on the current input but also on the previous hidden state. Theoretically, they could be used in any of the unsupervised deep learning approaches discussed in Section III; however, there are no works adopting them in the Deep Clustering approach (Section III-A). There are works with contrastive (Section III-B) [23], generative (Section III-D) [69], predictive (Section III-E) [63] and reconstruction-based (Section III-C) [40] approaches. Their formal definition is given as:

$$h_t = \sigma\left(W_h h_{t-1} + W_x x_t + b\right), \tag{3}$$

where W_h is the weight matrix for the previous hidden state, W_x is the weight matrix for the current input, b is the bias vector and σ is an activation function. The output y_t is the same as the hidden state h_t. While they can capture temporal dependencies, their performance suffers from vanishing gradients, leading to forgetting longer-term dependencies. For the sake of simplicity, the bias vectors which are existing for each weight matrix are not visualized in Figure 11.

1) LONG SHORT-TERM MEMORY NETWORKS (LSTMS)

LSTMs, depicted in Figure 11b, address the limitations of standard RNNs by introducing gating mechanisms that control the flow of information through a cell state. The gates include input (i_t), forget (f_t), and output (o_t) gates.

The forget gate f_t controls which part of the information is discarded from the previous cell state c_{t-1}, modeled as:

$$f_t = \sigma\left(W_{fh} h_{t-1} + W_{fx} x_t + b_f\right), \tag{4}$$

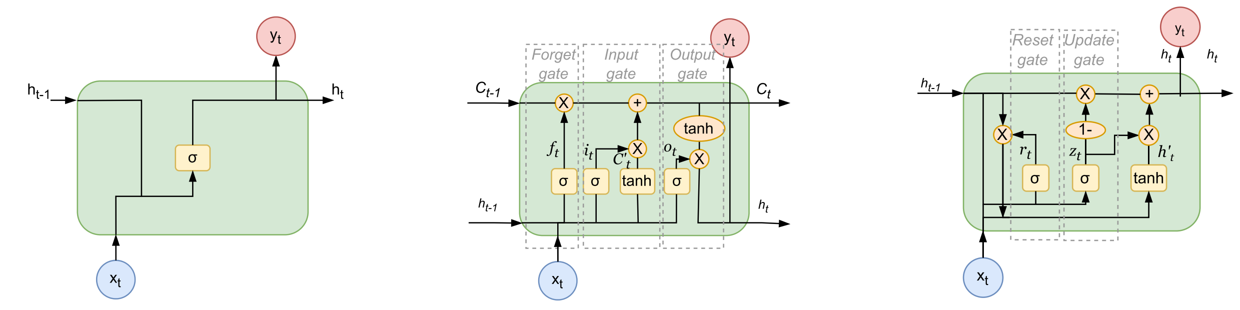
where σ is the sigmoid activation function, W_{fh} and W_{fx} are weight matrices for the hidden state and the input, and b_f is the bias term.

The input gate i_t controls which values will be updated in the cell state, i.e.

$$i_t = \sigma\left(W_{ih} h_{t-1} + W_{ix} x_t + b_i\right). \tag{5}$$

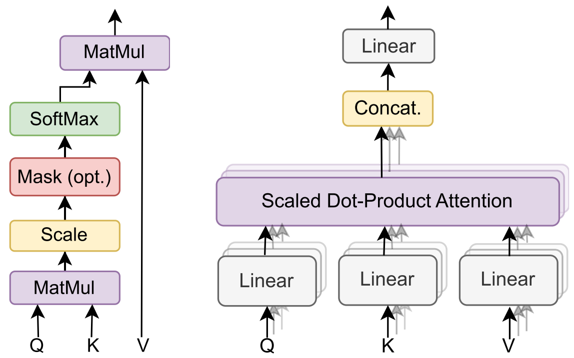
The candidate cell state $\tilde{c}_t$ generates the candidate values that can be added to the cell state, according to:

$$c_t' = \tanh\left(W_{ch} h_{t-1} + W_{cx} x_t + b_c\right). \tag{6}$$



(a) Basic RNN. (b) LSTM with forget gate. (c) GRU.

FIGURE 11. Types of recurrent neural networks.



(a) Self-attention module. (b) Multihead attention module.

FIGURE 12. Illustration inspired by [71].

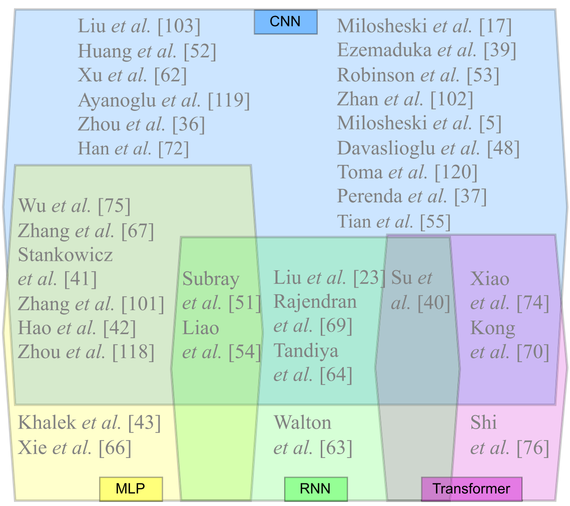


FIGURE 13. Venn diagram of the surveyed works with regard to the corresponding deep learning module used. Color coding corresponds to Fig. 1.

The new cell state c_t is a combination of the old cell state and the candidate values:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t. \tag{7}$$

The output gate o_t determines which parts of the cell state contribute to the hidden state:

$$o_t = \sigma\left(W_{oh}h_{t-1} + W_{ox}x_t + b_o\right). \tag{8}$$

And finally, the hidden State h_t is the actual output from the LSTM cell:

$$h_t = o_t \cdot \tanh\left(c_t\right). \tag{9}$$

The introduction of the three gates enables LSTMs to capture and model longer-term temporal dependencies. Thus, these models are suitable for time-series processing in general, including wireless communications signals. This approach is also one of the most common approaches for addressing classification problems in wireless communications. However, they are computationally more expensive to train compared to the basic RNNs.

2) GATED RECURRENT UNITS (GRUS)

GRUs, depicted in Figure 11c, are a simplified variant of LSTMs that combine the forget and input gates into a single update gate and use a reset gate.

The reset gate r_t determines how much of the past information to forget, according to:

$$r_t = \sigma\left(W_{rh}h_{t-1} + W_{rx}x_t + b_r\right), \tag{10}$$

where σ is a sigmoid activation function, W_{rh}, W_{rx} are the weight matrices for the previous hidden state and the current input, respectively, and b_r is the bias term for the reset gate.

The update gate z_t controls how much of the previous hidden state h_{t-1} is retained versus updated, according to:

$$z_t = \sigma\left(W_{zh}h_{t-1} + W_{zx}x_t + b_z\right). \tag{11}$$

The candidate hidden state $\tilde{h}_t$ incorporates the reset gate to decide which part of the past information to forget before applying a non-linear transformation, following:

$$h'_t = \tanh\left(W_{hh}\left(r_t \cdot h_{t-1}\right) + W_{hx}x_t + b_z\right). \tag{12}$$

The final hidden state h_t is a combination of the previous hidden state and the candidate hidden state, controlled by the update gate, calculated as:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h'_t. \tag{13}$$

GRUs are computationally more efficient than LSTMs while achieving comparable performance. Besides the basic training parameters, such as learning rate and activation function, these types of networks also require specifying the number of layers, the number of hidden units per layer, and the sequence length, which are the main drivers of the final model's complexity. Due to recurrent dependencies, the setup requirements are typically more complex but still comparable to those of CNNs. Furthermore, they are more time-consuming to train because they process data sequentially, which limits parallelization.

D. TRANSFORMERS (ATTENTION MECHANISMS)

The transformer architecture, originally designed for natural language processing, has proven highly effective for time series modeling due to its ability to capture both short- and long-term dependencies. While this learning module could be used in any of the approaches discussed in Section III, the existing works are only concerned with the contrastive (Section III-B) [70] and reconstruction-based (Section III-C) [40] learning approaches. The relatively low number of works utilizing transformers is due to the freshness of such architectures in the wireless communications domain, thus being still in the early adoption stage.

Transformers rely entirely on attention mechanisms rather than recurrence or convolutions. The self-attention mechanism computes the importance of different parts of a sequence relative to each other, given as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (14)$$

As illustrated in Figure 12a, the core component is the scaled dot-product attention, where query (Q), key (K), and value (V) matrices are processed through a series of linear transformations and attention mechanisms. The attention scores are computed by taking the dot product of Q and K, scaled by the square root of the key dimension d_k, and passed through a softmax function to produce attention weights. These weights are applied to V to obtain the context-aware output.

To enhance the model's capacity, multi-head attention, illustrated in Figure 12b, runs several attention layers in parallel, allowing the model to focus on different representation subspaces simultaneously. This parallelism and flexibility make transformers particularly suitable for capturing complex temporal patterns in time series data without relying on recurrence. The main parameters defining the configuration of an attention mechanism are the embedding dimension, number of attention heads, number of layers, and feed-forward network size. Self-attention models are generally more complex than both LSTMs and CNNs, particularly in terms of computational and memory costs, due to a quadratic increase in complexity with sequence length. Unlike LSTMs, attention mechanisms are highly parallelizable, which makes them faster to train on GPUs despite their higher theoretical complexity.

E. MIXED LEARNING MODULES

Besides the standalone usage of the learning modules in the approaches discussed in Section III, their combinations proposed in multiple works [40], [62], [72] prove to be even more suitable for unsupervised deep learning. The use of different types of neural networks as learning modules in combination allows for capturing different types of dependencies that could exist in the input signals. For instance, consecutive CNN and LSTM [73] can be utilized for capturing fine-grain as well as temporal dependencies, while in a combination of CNN and a transformer [74], CNN provides the temporal dependencies encoded as input vectors for the transformer.

F. OVERVIEW

In Figure 13, we present a Venn diagram visualization of the surveyed works based on the learning models employed. As expected, the vast majority utilize CNNs, likely due to their widespread adoption in related fields, straightforward implementation, and competitive performance. The second most commonly used models are Multi-Layer Perceptrons (MLPs); however, they typically appear as components within more complex architectures that primarily rely on CNNs. This preference is largely driven by the high data density of radio signals (e.g., I/Q samples or spectrograms), which poses processing challenges for standard neural networks. Interestingly, only a small number of works employ Recurrent Neural Networks (RNNs), despite their suitability for sequential data. This may be attributed to their higher complexity and comparable performance to CNNs. Currently, transformer-based models (e.g., multi-head attention mechanisms) are the least frequently adopted. Although they demonstrate superior performance, their limited use may be due to their recent introduction to wireless communication signal processing and the associated computational complexity constraints. As it is clearly depicted, many works utilize multiple models to combine their individual benefits.

V. AUTOMATIC MODULATION CLASSIFICATION

AMC is a technique in wireless communications that enables automatic identification of a signal's modulation format without prior knowledge of the transmitted signal. Standard modulation formats, such as PSK, QAM, and OFDM, encode information onto carrier signals. AMC facilitates their recognition by leveraging signal processing methods, recently based on machine learning (ML), both supervised [77], [78], [79], [80], [81] and unsupervised [82], [83], [84]. Fast and accurate AMC on the receiver side could allow for adaptiveness on the transmitter side of the terminal, changing the modulation schemas based on the perceived channel conditions. This enables better spectrum efficiency and throughput, similar to traditional adaptive coding and modulation techniques.

A. OVERALL AMC COMPARATIVE OVERVIEW

In Table 2, we summarize the unsupervised AMC works. As can be seen from the second column and also highlighted with different colors, most of them rely on contrastive learning approaches (Section III-B), two rely on reconstruction methods (Section III-C), while one employs deep clustering (Section III-A). This distribution shows the growing dominance of contrastive paradigms for feature extraction and representation learning in AMC tasks.

Representation learning module: As per the third column, it is evident that the CNN-based (see Section IV-B) backbones are the most frequently adopted, appearing in eight studies. These are often combined with additional components, such as attention mechanisms (see Section IV-D), which are apparent in four references, including one standalone. The LSTM and MLP (see Section IV-C and IV-A) modules are integrated less frequently and appear only once in combination with CNN. The fourth column of the table reveals that all works rely on I/Q data, meaning the different modulations are more distinguishable in the raw data than in the derivations, such as spectrograms or higher-order statistics. As a consequence, CNNs remain the main tool for capturing spatial correlations in I/Q signal representations, due to their capability of processing raw data as well as application flexibility (see Section IV-B). Recurrent and attention layers are typically used to enhance temporal or contextual feature modeling as part of the hybrid approaches.

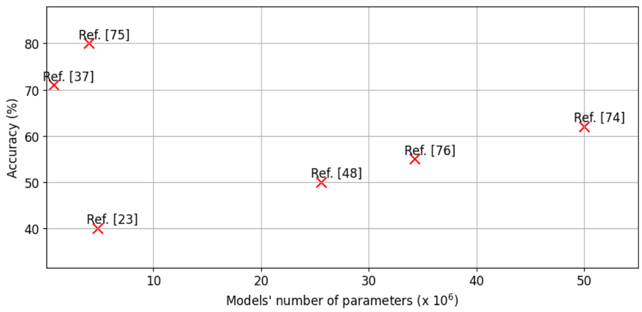
Experimental setup: The small-sample analysis column shows a consistent focus on label efficiency, with most approaches reporting results under extremely limited supervision, ranging from 0.5% to 10% labeled data or as few as 1–2 samples per class. Finally, the results and SNR/channel conditions columns reveal that evaluations are primarily conducted under Additive White Gaussian Noise (AWGN) conditions, with several studies also testing robustness under Rayleigh and Rician fading.

1) REPORTED PERFORMANCE ANALYSIS

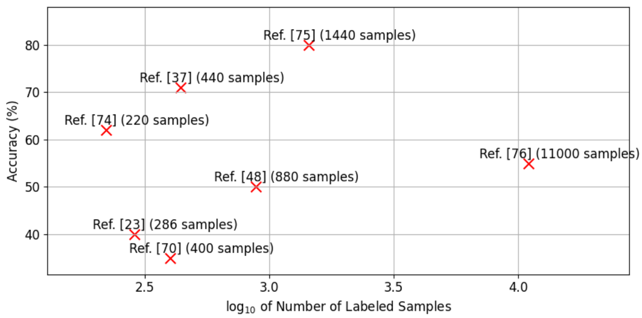
Reported accuracies in AMC works vary from around 35% to 80% depending on the proportion of labeled data and the network configuration, illustrating the performance–supervision trade-off typical in semi- and unsupervised AMC. Although difficult to compare, due to the different evaluation setups, considering different numbers of classes, labeled samples, and channel conditions, in Figure 14 we show approximate plots of *accuracy-complexity* and *accuracy-number of labeled samples*.

Figure 14a, which shows the dependency of accuracy on model size, indicates no clear monotonic relationship between the number of parameters and classification performance. Notably, [75] achieves the highest accuracy (80%) with a relatively compact CNN+MLP model; however, it uses the largest number of labeled samples. The work in [37] with a small architecture attains an accuracy of around 70% using only 440 samples. In contrast, larger models such as [74] (Attention+CNN with ≈50M parameters) and [76] (Attention-based AE with ≈35M parameters) reach moderate accuracies near 55-60%.

Figure 14b, which shows the dependency of accuracy on the number of labeled samples given in logarithmic scale, highlights the performance sensitivity to labeling effort. Here, [75] again leads with ≈80% accuracy using



(a) Accuracy vs. size comparison.



(b) Accuracy vs. number of labeled samples.

FIGURE 14. Reported performance in the reviewed works considering models' sizes and the number of labeled samples.

only 1% of labeled data (1440 samples), underscoring the effectiveness of its contrastive formulation in leveraging unlabeled data. In contrast, [37] and [74] maintain reasonable accuracies (≈62–70%) with substantially fewer labels (220–440 samples), reflecting strong data efficiency under minimal supervision. At the opposite end, [76] achieves lower accuracy (≈55%) despite employing the largest number of labeled samples (≈11,000), which may be attributed to the reconstruction-based learning objective and usage of GAF-transformed data being less discriminative.

Overall, these results reinforce that contrastive learning architectures, particularly those integrating CNNs with auxiliary modules (e.g., MLPs or attention layers), tend to outperform reconstruction-based or clustering counterparts on the AMC task, even with fewer labels and smaller model sizes.

Furthermore, we compare the performance of the surveyed models under low SNR conditions, specifically at –6dB, which represents the most common low-SNR evaluation point among the reviewed works. This comparison aims to highlight the robustness of the models in challenging noise environments. A direct and comprehensive comparison across all studies is difficult due to varying validation conditions, including the use of custom datasets, different modulation sets, and varying proportions of labeled samples. Therefore, we limit our analysis to works that report results on the common RML2018.01A dataset [85], and we compare their classification accuracy, as illustrated in Figure 15.

It is important to note that the evaluation protocols differ even within this subset of studies. Some works [23], [37], [70], [76] employ fully supervised training using

TABLE 2. Summary of unsupervised deep learning approaches and achievements in AMC.

Ref.	Approach	Module	Input	Small Sample Analysis	Evaluation type	Num. class.	Results (Accuracy / Labels)	Data	Code	SNR [dB]	Channel
[36]	Deep clustering (AE)	CNN	I/Q	–	unsupervised	6	0.796 ACC, 0.88 NMI at 10dB	✓	–	−18 to 20	AWGN
[37]	Contrastive + Reconstr. + Classif.	CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11, 20	70.54%, 55.84% / 2 labels per class (440, 800)	✓	–	−20 to 18	AWGN
[74]	Contrastive	Attention + CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11	62% / 1 label per class per SNR (220 samples)	✓	–	−20 to 18	AWGN / Rayleigh / Rician
[48]	MoCo-v3 Contrastive	CNN	I/Q	n=0.5%, 1%, 5%, 10%	supervised	11	50.4% / 0.5% labeled (880 samples)	–	–	−6 to 20	AWGN / Rayleigh / Rician
[75]	SimCLR Contrastive	CNN + LSTM	I/Q	n=1, 2, 5(1%), 10, 20, 30, 40, 50(10%)	supervised	11	40% / 2 label per class (286 samples)	–	–	−6 to 20	AWGN / Rayleigh / Rician
[76]	Contrastive	CNN + MLP	I/Q	n=1%	supervised	9	80% / 1% (1440 samples) labeled data	✓	–	−20 to 18	AWGN
[70]	Contrastive	Attention + CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11	35% / 2 per class per SNR (400)	✓	–	−20 to 18	AWGN
[77]	Reconstruction (AE)	Attention	GAF-I/Q	n = 5%, 10%, 20%	supervised	11	54.85% / 5% (11000) samples	✓	–	−20 to 18	AWGN

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction

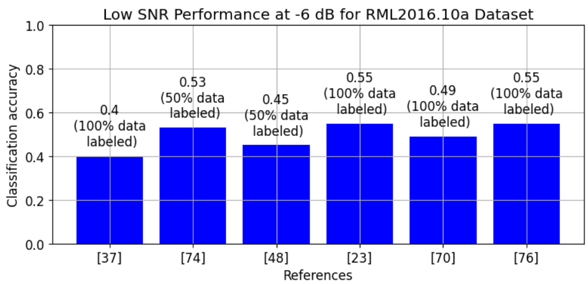


FIGURE 15. Reported accuracy of the surveyed models in low SNR conditions.

100% of the labeled data, whereas others [48], [74] use only 50% of the labels. Nevertheless, this comparison still provides meaningful insight into relatively low-SNR performance. Among the evaluated models, [74] demonstrates the best label-efficiency, achieving a classification accuracy of 0.53 fine-tuned on 50% labeled data. This strong performance is likely due to the model's large capacity, which allows it to effectively capture signal-specific features during pretraining. The top performing models at −6dB are reported in [23] and [76] while using 100% of the data labeled in the evaluation. Notably, all three of these leading models incorporate sequence-processing architectures, LSTM [23] and attention-based networks [74], [76], which suggests that such temporal or contextual modeling mechanisms exhibit superior robustness under low-SNR conditions.

B. OVERVIEW PER APPROACH TO AMC

1) DEEP CLUSTERING

In [36], the only study with a deep clustering approach in this section, the authors developed a custom clustering loss function for favoring better separation in the feature space provided by the encoding (middle) layer of the autoencoder. They combined it with the standard Mean Squared Error (MSE) reconstruction loss for training the models, thus enabling simultaneous feature learning and clusterability improvement in the feature space. The general constraint of such approaches is the combined loss function tuning. The separate components do not necessarily contribute to the general goal, considering that lowering the reconstruction loss does not always mean improving the clusterability. This could lead to complex shapes of clusters [17], which are impractical for post-processing and knowledge extraction.

2) CONTRASTIVE LEARNING

For the task of AMC, most of the recent works that rely on deep unsupervised learning [23], [37], [48], [70], [74] use contrastive learning architectures (see Section III-B), which are utilized in completely unsupervised setup or in semi-supervised [23] setup where the contrastive learning is only part of the pretraining phase. In [48], authors use the contrastive learning paradigm in one of its original setups, known as MoCo-v3 [86], using the ResNet50 CNN module (see Section IV-B), with special data augmentations relevant for the I/Q samples of radio data. They show that such a model is consistently better compared to other out-of-the-box contrastive learning frameworks, such as SimCLR [87].

In [37], the authors propose a weighted sum of three losses for the training architecture of three parts: reconstruction, contrastive, and classification, using weak and strong augmentations instead of the original signal and a single augmentation. They achieve 70.54% averaged accuracy over 11 classes and 55.84% averaged accuracy over 20 classes, trained with only 2 labelled samples per class, without unlabeled samples, under AWGN noise in the range between -6 and 20dB. Further work could consider performance validation with unsupervised pre-training so that the full potential of such an approach is explored.

The authors in [23] introduce contrastive learning to AMC by utilizing SimCLR [87] to train a feature extractor (encoder) in the pre-training phase and subsequent fine-tuning on labeled data. They show that pretraining of the encoder leads to better performance on the downstream task compared to a supervised model using the same amount of data.

In [75], contrastive learning is part of a more complex architecture that enables the model to learn robust and generalized representations of radar signals during the semi-supervised pre-training stage. They explore the effect on the overall performance of a small subset of iterative labeled data injection in the fine-tuning procedure, and achieve better performance compared to the baseline MoCo [88] approach. A similar approach is considered in [70], where the learning module is realized with a transformer (see Section IV-D) instead of a CNN. They also provide extensive evaluation regarding the influence of each augmentation technique, the feature embedding power of the self-attention module, and the influence of the number of labeled samples in the fine-tuning stage.

Considering that data augmentations take a central role in the performance of contrastive learning, authors in [74] propose more domain-specific, semantics-preserving augmentations that mimic the environmental influence that affects the signal's shape, including Time Warping, Carrier Frequency Offset, and Phase Offset. Additionally, they use strong and weak augmentations of the input signal instead of the signal itself and one augmentation for the contrastive setup, marked as a' and a'' in Figure 4. For the learning task, they employ a Temporal Convolutional Network (TCN) [89] with masking for capturing the short-term features in the raw signal and a consequent transformer for capturing the long-term features. Together, they build the learning module of the contrastive learning setup. While the basic contrastive setup, as depicted in Figure 4, rates equally all of the negative samples, it introduces a weighting function in the loss calculation. This function takes into consideration the distances between the encodings of the samples by favouring closer negatives in the feature space, which leads to improved performance on a downstream classification task. They evaluate the proposed architecture against multiple approaches, including supervised CNN and LSTM (Sections IV-B and IV-C), and prove that it outperforms the competing approaches.

3) RECONSTRUCTION

Autoencoders (see Section III-C) are yet another unsupervised deep learning approach utilized for AMC. Compared to the Contrastive Learning works (see Section III-B), the AE seems less capable, but they have the advantages of being simpler and do not require data-specific augmentations. The different AE solutions vary regarding the learning module that is used for the encoder and decoder functionalities [40] and modifications in the loss functions. In [40], authors combine GRU, self-attention, and residual CNN blocks in the encoder and decoder modules to capture fine-grained (CNN) and time-dependent (GRU) features with a low-dimensionality representation (self-attention). The resulting architecture provides a higher score than the regular off-the-shelf CNN-based autoencoders, such as ResNet-based, on the task of low-signal detection with various modulations (RadioML data) across different low levels of signal-to-noise ratio (SNR).

Another AE-based, innovative approach is proposed in [76], where I and Q data streams are transformed to Gramian Angular Field (GAF) images, and an autoencoder is trained to reconstruct the fused image (combination of both I and Q components) based on the masked input image. The encoder and decoder are realized with a transformer architecture. While the baseline concept is the same, the usage of image-like data instead of raw I/Q time series makes the approach more flexible regarding the selected encoder-decoder module, considering there are many proven image processing architectures in open-source libraries such as [90].

C. DATASETS

All of the reviewed works consider the well-known RadioML dataset with its three variations RADIOML 2016.04C [85], RADIOML 2018.01A [91], and RADIOML 2016.10A, summarized in Table 3, as an evaluation baseline.

TABLE 3. Summary of RadioML datasets and their characteristics.

Ref.	Characteristics	Specific Features
[86]	RADIOML 2016.10A Synthetic I/Q signal samples across 11 modulation types (3 analog, 8 digital)	Includes labeled SNR increments, (-20dB to +18dB) with samples containing 128 I/Q data points and balanced classes.
[86]	RADIOML 2016.04C Synthetic I/Q signal samples across 11 modulation types (3 analogs, 8 digital)	Includes labeled SNR increments, (-20dB to +18dB) with samples containing 128 I/Q data points and unbalanced classes.
[92]	RADIOML 2018.01A Synthetic I/Q signal samples across 24 modulation types (analog and digital)	Contains 1,024 I/Q data points per sample across 26 SNR levels (-20dB to +30dB), including AWGN and Rayleigh fading.

Both RADIOML 2016.10A and RADIOML 2016.04C data consist of synthetic I/Q signal samples across 11 modulation types, including three analog and eight digital schemes in varying SNR conditions. RADIOML 2016.04C has additional variation among the samples with labeled SNR increments, allowing for a more detailed analysis of the benchmarked models.

The RADIOML 2018.01A dataset is the most comprehensive one, containing synthetic I/Q signal samples from 24 digital and analog modulation types. Each sample consists of 1,024 I/Q data points across 26 SNR levels ranging from -20dB to $+30$dB, simulating realistic wireless channel conditions with additive white Gaussian noise (AWGN) and Rayleigh fading.

All of the related papers perform evaluations based on these three datasets. While these three datasets are a good base for direct model-to-model comparisons of the different approaches for AMC, real-world evaluation is lacking. The creation of real-world measurement datasets could provide great insights about the actual deployment capabilities of the developed models. As such, they could contribute to building models for actual existing challenges instead of models for the available data.

D. FINDINGS

Based on the reviewed literature in this domain (Table 2), contrastive learning appears as the most promising approach, given that more than half of the recent research works propose it as the superior solution for AMC. Thus, future efforts should explore more domain-specific data augmentations, such as those mimicking environmental influences, to improve model robustness in diverse operating scenarios. The combination of multiple methods in sequential or parallel learning phases may help train more capable models regarding their generalization to unseen environmental conditions.

1) EVALUATION PRACTICES

Nearly all surveyed studies rely on simulated datasets, with the RadioML corpus and its variants being the most frequently used benchmarks. While these datasets facilitate controlled experimentation, they often fail to capture the full complexity of real-world environments. Thus, an important direction for future work is the development and open dissemination of real-world RF benchmark datasets. These would reveal challenges beyond the signal-to-noise ratio (SNR) variability, such as channel fading models and environmental dependencies, which could be obtained by spatial Radio Environment Maps (REMs) [92]. In such a way they would support validation of the proposed models in operational conditions that are more representative of the actual deployment.

Moreover, there is a pressing need for a standardized evaluation framework, tailored to domain-specific characteristics such as viable I/Q sample length for actual deployment, considering the time delays of the decision-making and the system itself, SNR levels, labeling ratios, and augmentation techniques. Although many studies assess model performance under limited labeled data conditions, there is significant inconsistency in the selection of sample percentages or absolute quantities, making direct comparisons between works problematic. Currently, comparisons often center around performance with only one or two labeled samples, but the diversity in evaluation setups obscures the relative progress across methods. A consistent benchmarking methodology, such as enforcing a fixed number of labeled samples, using standardized activation functions in the final fully connected layers, or defining common augmentation pipelines, would greatly enhance transparency, reproducibility, and comparability across studies.

2) ROAD AHEAD

In standardized wireless systems where a fixed set of modulation schemes is used, supervised learning models have already demonstrated high classification accuracy [80]. However, recent advances suggest a shift towards machine-learned physical layers, where deep learning models replace traditional components of the communication chain [93]. These approaches enable the design of custom modulations, dynamically adapted to specific channel conditions. Such a paradigm shift will demand self-adaptive, label-free learning models capable of generalizing to new, previously unseen signal formats and operating conditions. This challenge points to unsupervised deep learning (see Section III) as a key enabler, allowing models to extract general signal features from large-scale unlabelled data and to adaptively refine their internal representations based on deployment-specific observations.

VI. SIGNAL SENSING

While AMC, discussed in the previous section, focuses on the categorization of signal-related features for a given signal, this section addresses signal sensing, which involves detecting, analyzing, and interpreting signals within a specific radio spectrum or environment to determine their presence, characteristics, and properties. Signal sensing provides general awareness in the time and frequency domains regarding the operational patterns of various technologies and devices. It processes sensed data that may or may not contain signals, making the problem more general and closely tied to the physical layer of communication systems. The main purpose of such awareness is the smart utilization of the radio spectrum as a scarce resource by multiple devices that rely on different RATs, thus avoiding interference. This is especially important in the unlicensed radio spectrum bands such as the Industry, Scientific and Medical (ISM) [94] and 5.9 GHz Intelligent Transportation Systems (ITS) [95], [96] in which multiple technologies share similar frequency bands.

Signal sensing in the following is focused on the general monitoring of radio spectrum occupancy only by analyzing the patterns of spectral activity in the concerned frequency bands. This involves extracting knowledge about activity patterns in frequency and time domains with the main goal of using the free spectrum and time resources in a dynamic manner based on historical knowledge, complementary to other sensing techniques, such as AMC and SEI.

A. TYPES OF SIGNAL SENSING

For clarity and consistency, we group the reviewed works into three complementary research directions aimed at understanding and managing spectrum usage, namely Primary User (PU) detection, Radio Access Technology (RAT) recognition, and Jamming Signal Recognition (JSR). Each category addresses a specific aspect of spectrum awareness. More specifically, they identify active licensed transmissions, characterize the operating technology and detect intentional or unintentional interference.

1) PRIMARY USER DETECTION (PUD)

PUD focuses on determining whether a licensed or prioritized user is transmitting within a given frequency band. Although related to Specific Emitter Identification (SEI), PUD operates at a broader level, as it does not aim to identify individual transmitters but rather to detect the presence of active users or user groups. While direct processing of raw in-phase and quadrature (I/Q) signal samples could achieve this task, such methods raise security and privacy concerns with respect to potentially revealing user-specific information. As a result, existing approaches often rely on short I/Q sequences, typically from the signal preamble [96], to infer PU activity without accessing sensitive data, or some derived data representation such as spectrograms [54] or energy levels [66].

2) RADIO ACCESS TECHNOLOGY (RAT) RECOGNITION

RAT recognition represents a key component of signal sensing, aimed at identifying the type of radio access technology active in a monitored frequency band. This task involves analyzing spectral activity patterns across time and frequency to distinguish among different communication standards. The extracted knowledge could enable information for dynamic spectrum access by predicting and exploiting unused spectrum resources based on historical activity. RAT recognition, therefore, complements techniques such as AMC and SEI, providing a higher-level insight into spectrum utilization and technology coexistence.

3) JAMMING SIGNAL RECOGNITION (JSR)

We also categorize the *jamming signal detection* as part of the general signal recognition since it requires distinguishing between the wanted and unwanted signals operating in the same portion of the radio spectrum. For this challenge, the unsupervised deep learning models are particularly suitable for two reasons. Firstly, there is usually a lot of sensed data that is difficult to label since the jamming signals are typically unknown. Secondly, the radio spectrum environment may be very dynamic, where there could be many unknown sources of radiation in the monitored band, especially in the unlicensed bands.

B. OVERALL SS COMPARATIVE OVERVIEW

In Table 4, we summarize the works employing unsupervised deep learning for signal sensing. Three types of methods are equally used: deep clustering (Section III-A), reconstruction (Section III-C), and generative (Section III-D), as can be seen from the second column of the table and also highlighted with different colors. As can be seen from the third column, seven architectures contain CNN modules (see Section IV-B) at least as part of the complete solution, four contain MLPs (see Section IV-A), while LSTM, MLP, and GRU (see Section IV-C) appear twice each.

Regarding the data type, I/Q samples are most frequently utilized, in six of the surveyed works, which is expected considering that it is the highest density data and thus provides the richest features. However, for the signal sensing challenges, low data intensity solutions also exist, such as energy levels of the monitored channel for PU detection [43], [66] and FFT-derived data such as spectrograms for RAT classification [5], [17].

PU detection is the most frequently explored task according to the reviewed works. Most approaches rely on AE or VAE architectures applied to I/Q data [39], [40], [43]; however, simpler energy detection solutions based on MLP are also viable [66]. AE and VAE are the most common approaches for the PU task, which means that reconstruction and generative models provide robust detection capabilities under low supervision and variable noise conditions. Interestingly, the PU detection is the only task that considers the channel model.

RAT recognition tasks are addressed in [5], [17] and [51], mainly through deep clustering and AE-based approaches. This confirms that clustering-based feature learning is well-suited for RAT differentiation, as the RAT-specific patterns are distinct and sufficiently captured through frequency-domain representations.

Jamming recognition appears in two of the reviewed works [67], [72], both relying on AE architectures and considering the Jamming to Signal ratio in the experimentation. This shows that AEs, although being simple models, are well-suited for the task of JSR.

In terms of evaluation and reproducibility, most studies conduct unsupervised validation, with a few complementing it by supervised fine-tuning for benchmarking. Reported class numbers range from two to eight, depending on task specifics, which makes potential direct comparison of the methods difficult.

1) REPORTED PERFORMANCE ANALYSIS

Figure 16 compares the model sizes across the reviewed signal sensing works, categorized by task type. The results show that RAT (in blue) and jamming (in red) recognition models tend to be the largest, with parameters ranging from approximately 0.13 M to 11.2 M, reflecting the need for richer spectral feature extraction in multi-signal environments. In contrast, PUD models (in green) are considerably smaller, mostly below 1 K parameters, except for the cGAN-based approach [54] which reaches 3 M parameters due to its generative nature. This trend indicates that PU detection tasks can be effectively addressed with compact architectures, while more complex sensing problems, such as RAT classification, benefit from deeper or wider networks.

C. OVERVIEW PER APPROACH TO SS

1) DEEP CLUSTERING

a: RAT

The performance of deep clustering (see Section III-A) with different 2D CNNs and transformer deep learning modules is explored in [5] for wireless technology recognition in the license-free 868 MHz frequency band by using spectrogram

TABLE 4. Summary of works concerning unsupervised deep learning approaches and achievements in signal sensing.

Ref.	Approach	Module	Input	Problem type	Evaluation type	Num. class.	Performance	Data	Code	Noise conditions	Channel
[5]	DeepCluster	CNN	Spectrogram	RAT clustering	unsupervised, supervised	unkn.	0.6 Silhouette score / 0.84 F1 score	✓	✓	–	Real-world
[17]	DeepCluster	CNN	FFT Amp.	RAT clustering	unsupervised, supervised	unkn. /3 /6	0.6 Silhouette score / 0.99 F1 score avg.	✓	✓	–	Real-world
[72]	Autoencoder + Classifier	CNN	I/Q	Jamming recognition	unsupervised	7	>0.95 F1 score	–	–	JSR $(-30$ to $20)$dB	–
[67]	Autoencoder	CNN + MLP	I/Q	Jamming recognition	unsupervised	8	0.95 Accuracy	–	–	JSR (10 to 15)dB	–
[51]	Autoencoder	CNN, MLP, LSTM	I/Q	RAT classification	supervised	2	0.93 F1 score	–	–	–	–
[43]	Sparse Autoencoder (SAE)	MLP	I/Q	PU detection	unsupervised	4	Detection prob. 0.8 at 1W and 0.1 False alarm probability	–	–	Emiting power (0.1 to 1)W	AWGN
[40]	Autoencoder	GRU+ CNN+ Atten.	I/Q	PU detection	unsupervised	8	Detection prob. 0.85	✓	–	$(-18$ to $-5)$dB	Rayleigh
[66]	VAE	MLP	Energy Levels	PU detection	unsupervised	2	Clustering accuracy 86.02%	–	–	AWGN, Laplacian $(-15$ to $5)$dB	Rayleigh fading
[39]	VAE	CNN + MLP + GRU	I/Q	PU detection	unsupervised	2	Detection probability 71% at 0.1 False alarm prob.	✓	–	AWGN $(-18$ to $-5)$dB	Rician model
[54]	cGAN	CNN	Spectrogram	PU localization	supervised	2	Error (33 to 15)m	–	–	AWGN (10 to 50)dB	–

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction

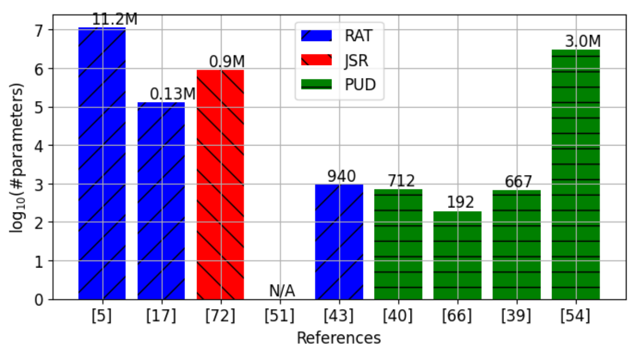


FIGURE 16. Size comparison of the models in signal sensing. Blue - RAT recognition, Red - JSR, Green - PUD.

data. Complexity-performance tradeoffs are analyzed, with a conclusion that the utilization of large CNN and Vision Transformer (ViT) modules for feature learning brings only minor performance improvement while significantly increasing the model size, thus imposing constraints on its application potential. The reason is the low-content pictorial data in the spectrogram images, which is captured by lower complexity models such as ResNet11 with comparable performance to the significantly larger models, such as VGG16 and ViT.

b: RAT

Another work for RAT categorization based on deep clustering is published in [17], where the target is even further lowering the model size. 1D FFT amplitude data is considered instead of spectrograms, thus allowing for significantly shorter end-to-end inference time of the 1D CNN-based models, such as depicted in Figure 10. The increased recognition performance was validated on three different datasets, each targeting a specific deployment challenge. These included the environment influence, the increased number of operating RATs, and the spectrum monitoring data of continuous operation with many unknown signals and noise. The proposed model was shown to be robust and easy to set up compared to the then-state-of-the-art AE-based models. Although the labeled validation datasets are from real-world transmissions, the very high performance (up to 0.99 F1 score) could be a result of using very clean data. The evaluation with the unlabeled, live stream spectrum data could be considered the most informative for the actual performance, considering the three separate evaluations.

c: JSR

In [72], AE architecture (see Section III-C) and classifier networks are jointly used to learn feature extraction from known (labeled) samples of jamming signal during training. In the deployment stage, only the feature extractor is used, combined with two-stage distance-based classification. Firstly, the binary classification of known and unknown signals is performed, and a more granular classification of different patterns is performed afterwards. A joint objective function containing center loss, cross-entropy, and reconstruction loss is employed with corresponding weighting coefficients. The distance-based classification is actually clustering signals based on Euclidean distance while also taking into consideration the known classes. The main advantage compared to the original implementation is obtained when classifying three unknown classes of jamming signals. However, the reported results appear pretty high with the accuracy of 100% for three known classes and one unknown class.

d: JSR

Similar to the approach in [72], and on a similar problem of unknown Jamming signal recognition, authors of [67] propose an AE-based architecture that can extract semantic attribute features. The only difference is in the definition of the clustering loss. They compare their approach to a significantly simpler (three-layer) MLP trained in a supervised setup and show the superior performance of the proposed method. Additionally, they also provide an ablation analysis of each of the losses contributing to the performance of the method.

2) RECONSTRUCTION

a: RAT

In [51], the authors used deep AEs (see Section III-C) on I/Q data, which contains samples from different modulations, in an attempt to address the problem of RAT classification between the LTE and a combination of WiFi signals (IEEE 802.11ax and IEEE 802.11ac). Interestingly, they show that the deep AE with CNN layers outperforms the LSTM-based and VAE architectures, although LSTM was meant for longer-term patterns apparent in time-series data such as I/Q sequence.

b: PUD

The PU detection task is addressed in [43] with Sparse AE (SAE) combined with Gaussian Mixture Model (GMM). Energy levels sensed by the SUs are used as input data. A cooperative setup is considered where the secondary users (SU) send their sensing data to the central coordination node, where PUs and SUs are spatially distributed in the operating environment. In the evaluation setup, it is considered that SUs can appear and disappear due to their contribution to the signal measurements of the central node. For the considered scenario, the authors prove that their proposed SAE-GMM model achieves performance comparable to a supervised neural network on the task of binary classification of active and inactive PUs in the monitored channel. The solution is realized by introducing sparsity in the embedding layer, by constraining the average activation rate of the neurons of the embedding layer of the AE architecture. The idea is that forcing only a small number of active neurons during training will lead to learning compressed and more meaningful representations. The sparsity is controlled by incorporating the Kullback-Leibler divergence between the required average activations and the actual average activations as part of the loss function, which originally contains only MSE loss. It is important to highlight that the claimed performance is achieved using a very simple, fully connected neural network with only five layers.

3) GENERATIVE

a: PUD

In [66], the authors propose using VAE (see Section III-D), a representative of generative methods, combined with GMM for distinguishing PU signals given unlabeled data of PU transmission signals and free channel signals. VAE is used for feature extraction, and GMM for clustering in the latent space. While the training is performed completely unsupervised, a small labeled subset of data is required for threshold adjustment and cluster identification. With this approach, they achieve comparable results to the supervised CNN counterpart, which requires significantly more labeled data for training.

b: PUD

VAE is also used for unsupervised learning of features in [39] for the PU detection in various single and multiple PU setups for the Non-Orthogonal Multiple Access (NOMA) technique. The feature learning with VAE is performed on denoised radio data with so-called recorrupted-to-recorrupted [97] denoising based on a proposed GRU-based (see Section IV-C) AE architecture for radio signals. The output classification is performed with the K-means++ algorithm. It is important to note that in this work, the authors propose a shallow encoder and corresponding decoder for the AE and VAE structures adapted to the characteristics of the radio signals, such as strong, stationary periodic properties and high frequency. Such architecture is shown to outperform the referenced approaches, which also contain a supervised CNN model, proving the significant contribution of the denoising function by the proposed GRU-based model.

c: PUD

In [54], the authors propose a two-stage solution based on GAN and U-shaped NN (U-Net) for privacy-aware localization of spectrum violators in the United States' Citizen Broadband Radio Service (CBRS) [98]. In their work, GAN architecture (Section III-D) is used for masking PUs in spectrograms acquired with multiple sensors, and U-NET [99] is used in an image-to-image translation setup based on labels. Thus, instead of the standard signal recognition setup where a given signal is classified, GAN is used for masking the privacy-sensitive PU transmissions in spectrograms, which implicitly requires learning their appearance in spectrograms. While the original setup of GAN is unsupervised, the way it is used in this work requires some sort of labeled data, consisting of spectrograms with and without PU transmissions.

D. DATASETS

In Table 5, we summarize some of the open, real-world datasets that could be used for future research in signal sensing challenges.

TABLE 5. Summary of real-world datasets for wireless signal recognition.

Ref.	Characteristics	Specific Features
[65]	Labeled dataset consisting of transmissions from LTE, WiFi, and DVB-T technologies operating in the 2.4 GHz band.	Data collected from 7 locations (6 in Ghent, Belgium, and 1 in Dublin, Ireland), enabling analysis of environmental effects on signal recognition.
[97]	Labeled dataset with signals from LTE, 5G, WiFi, ITS-G5, and C-V2X technologies operating in the ITS 5.9 GHz band.	Suitable for exploring model performance with coexisting RATs, including technologies with different or similar spectral shapes.
[19]	Partially labeled dataset from the LOG-a-TEC testbed in Ljubljana, Slovenia, covering the 868 MHz SRD band.	Contains spectrum traces with at least four technologies (LoRa, IEEE 802.15.4, SigFox, proprietary technologies). Evaluates model performance in scenarios with high variability and interference.

A labeled real-world dataset from [65] consists of transmissions of three RATs, namely LTE, WiFi, and DVB-T, which operate in the 2.4 GHz band. The complete data was collected from 7 different locations, including 6 in the city of Ghent, Belgium, and one in Dublin, Ireland. This dataset allows for a detailed analysis of the environmental effects on signal recognition models' performance.

Another dataset [96] of real-world measurements was collected in Antwerp, Belgium, and includes signals from five RATs: LTE, 5G, WiFi, ITS-G5, and C-V2X, all operating in the ITS 5.9 GHz band. This dataset provides an insight into scenarios with more coexisting RATs compared to the previous one, including technologies with significantly different or similar spectral shapes. This makes it suitable for exploring how the coexistence of more technologies affects the models' performance.

Yet another partially labeled dataset was collected from the LOG-a-TEC testbed in Ljubljana, Slovenia [19]. It comprises spectrum traces in the 868 MHz Short-Range Device (SRD) band with a bandwidth of 192 kHz and a sampling frequency of 5 Power Spectrum Density measurements per second, using 1,024 FFT bins. At least four technologies appear in this data: LoRa, IEEE 802.15.4, SigFox, and proprietary technologies. This dataset is suitable for evaluating model performance in real-world scenarios with significant variability, artifacts, and interference.

E. FINDINGS

1) PUD

PU detection is commonly formulated as a binary classification problem, focused on identifying the presence of a signal of the primary users. This formulation lends itself to relatively simple evaluation using standard metrics such as accuracy, Receiver Operating Characteristic / Area Under the Curve (ROC-AUC), and F1-score, which facilitates straightforward benchmarking across studies. While such simple setups allow for analysis on the model performance, considering also SNR and channel state, they are overly simplistic and do not reflect actual deployment conditions where many more users are present in the environment. Thus, following the examples of [40] and [43], multiple-user scenarios should be considered in future research.

2) JSR

Similarly to the PUD task, JSR is also addressed as a binary detection problem. Therefore, moving beyond simple presence detection to the classification of jamming types as in [67] is an important research direction. Such classification would support the development of targeted mitigation strategies, making this task increasingly relevant for operational deployments. Furthermore, setting the jamming recognition as an Open-Set problem, where the appearance of unseen jamming patterns is expected, should be a favoured approach in future works, following the examples in [67], [72]. Thus, further exploration of unsupervised deep learning approaches is highly relevant in this direction of work.

3) RAT

From a signal recognition perspective, RAT recognition parallels the challenges of JSR, where systems must identify and group multiple signal types without labeled data [5], [17], while favouring even faster decision making, considering potential applications in near-real-time sensing and network parameters' control. RAT recognition requires the identification of multiple signal classes or clusters without or with a very small set of labeled samples. This makes unsupervised evaluation often a necessity in estimating the performance, considering clustering metrics such as Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). When clean datasets are used [96], and the signals exhibit distinct frequency-domain features (e.g., unique FFT shapes), unsupervised learning can show extremely good performance [17], [40], which could be misleading. Therefore, future works in this direction should consider both supervised and unsupervised evaluations, and multiple metrics due to the uncertainty in the latter.

4) ROAD AHEAD

While supervised evaluation is well-established, the unsupervised, which is highly relevant in this direction of work, requires a more unified and even standard approach that uses a well-defined set of metrics, which could be beneficial for future benchmarking and development. Regarding the data, even when generated with over-the-air (OTA) capturing methods in a controlled manner, some datasets [65], [96] fail to fully reflect the dynamics, variability, and noise present in real-world RF environments. This results in significant performance differences compared to live continuously sensed data [17]. We recommend that in future research, both types of data, from controlled and uncontrolled environments, should be used for providing better performance estimation, especially for applications intended for deployment in live, dynamic wireless networks [19].

From a deployment perspective, signal sensing must be extremely fast, often within a few milliseconds, to meet the real-time requirements of modern wireless systems [96]. This constraint encourages the placement of sensing algorithms close to the radio unit, typically at the edge or within distributed units (DUs) in O-RAN architectures, where both latency and computational resources are limited. These limitations promote the use of lightweight models that minimize data movement and processing overhead. As a result, averaged FFT representations of the signal are often preferred over raw I/Q samples, leading to computational reductions by up to three orders of magnitude [2].

The rise of distributed applications (dApps) in edge computing and their increasing integration into O-RAN as standard components [100] will further support the development of near-real-time, low-level machine learning models for signal sensing.

VII. SPECIFIC EMITTER IDENTIFICATION

SEI relies on recognizing device-specific signal alterations caused by hardware imperfections [104], [105]. Deep learning techniques, especially supervised models, have shown exceptional performance [106] in distinguishing known devices, even in networks with thousands of transmitters [107], [108], [109]. However, real-world scenarios often lack labeled data and may also contain transmissions of unknown transmitters, which could pose security concerns. Thus, significant attention is paid to the development of models that can learn from data from known transmitters and distinguish potential unknown transmitters. The challenge could be addressed by using CNN-based feature extractors [84], [105], which are trained on known devices and aim to group unseen emitters in the latent space via clustering algorithms. So far, existing solutions approach the training of the feature extractor through a classification task [53], [110]. Unsupervised deep learning fits as a promising approach for this challenge [110], considering the ability to learn from large amounts of unlabeled data.

A. OVERALL SEI COMPARATIVE OVERVIEW

In Table 6, we summarize the works concerning SEI. Similar to AMC (Section V), most of them rely on contrastive learning approaches (Section III-B), two rely on deep clustering methods (Section III-A) while one employes reconstruction (Section III-C) and one predictive (Section III-E) approach, as can be seen from the second column of the table and also highlighted with different colors. As per the third column, all architectures contain CNNs (see Section IV-B) at least as part of the solution, and three also include MLPs (see Section IV-A). The fourth column of the table reveals that all works rely on I/Q data, which is expected since the goal is to capture the hardware-induced I/Q irregularities in the radio signal, specific to each device.

With regards to the *Problem type* column, SEI evaluations are performed in a zero-shot and few-shot settings, which directly corresponds to the unsupervised and supervised *Evaluation type*, accordingly. Small-sample evaluations are of great interest for the SEI task due to the similarity to the actual deployment conditions. The *Performance* column demonstrates consistently strong results, with detection and clustering accuracies typically above 90%. This confirms that unsupervised representations can effectively capture hardware-specific signal alterations. However, each publication reports evaluation with a different dataset and with different numbers of samples in the few-shot setup, which constrains direct comparison. According to the *Data* and *Code* columns, there is limited public availability of datasets and implementations, with only a few works providing accessible repositories. Regarding the *SNR* and *Channel* columns, only three of the studies consider the noise effect. Finally, regarding the channel model, only one study evaluates with antenna replacement, which means currently most of the focus is on the model's design and simplistic validations with varying numbers of seen and unseen devices.

1) REPORTED PERFORMANCE ANALYSIS

For the surveyed SEI works it is impossible to extract the performance comparison with respect to model size due to the large variety of evaluation conditions, such as different dataset classes, number of samples per class, and different technologies. Therefore, Figure 17 only compares estimated model sizes used across the SEI studies for zero-shot (blue) and few-shot (red) evaluation setups. Most of the proposed models in both setups have comparable sizes of up to 1 M parameters, with the exception of [42] and [52], reaching 11 M and 4.53 M parameters, respectively. Overall, the figure shows that SEI models achieve strong generalization with relatively lightweight architectures, and that model size scales primarily with the degree of supervision and the complexity of the experimental setup.

B. OVERVIEW PER APPROACH TO SEI

1) DEEP CLUSTERING

The study in [41] focuses on novel device discovery with deep clustering (see Section III-A) by extracting features

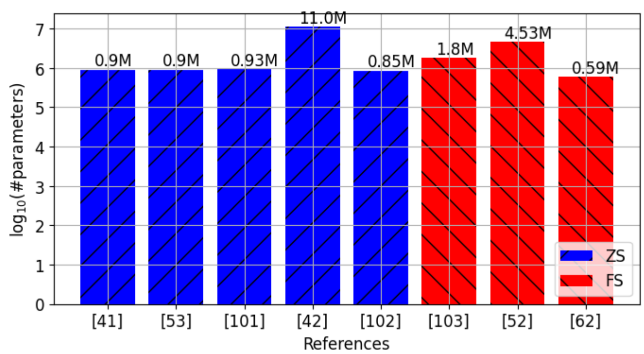


FIGURE 17. Estimated size comparison of the models in SEI. Blue - Zero-shot (ZS) evaluation, Red - Few-shot (FS) evaluation.

TABLE 6. Summary of unsupervised deep learning approaches and achievements for SEI.

Ref.	Approach	Modul	Input	Prob. Type	Evaluation Type	Performance	Data	Code	Small sample analysis	SNR (dB)	Channel
[41]	Deep Clustering	CNN+ MLP	I/Q	Zero-shot	Unsupervised	Clustering Purity WiFi: 0.564 ADS-B: 0.899	–	–	–	–	–
[53]	Deep Clustering	CNN	I/Q	Zero-shot	Unsupervised	Single known emitter detection Acc. 89.1%	–	–	n = 0, 50, 100, 200, 400, 800	–	–
[103]	Contrastive Learning	CNN+ MLP	I/Q	Zero-shot	Unsupervised	15 known 1 unknown, detection Acc. 91.79%	✓	–	0–30%	–	Antenna replacement
[42]	Contrastive Clustering	CNN+ MLP	I/Q	Zero-shot	Unsupervised	6 known 1 unknown, detection Acc. 91.21%	–	–	n = 1, 5, 10, 20, 30, 50	−10 to 20 by 10 incr.	AWGN
[104]	Meta-Contrastive	CNN	I/Q	Zero-shot	Unsupervised	30 known vs 15 unknown Accuracy: 90.5%	✓	✓	–	–	–
[105]	Contrastive	CNN	I/Q	Few-shot	Supervised	WiFi: 10 vs 6, ADS-B: 90 vs 10, 10-shot Acc. 86.33%, 62.73%	✓	✓	5 to 30 by 5 incr.	–	–
[52]	MAE (Reconstruction)	CNN	I/Q	Few-shot	Supervised	8 emitters, 10-shot Acc.: 41%	–	–	20 to 100 by 20 incr.	0 to 10 by 1 incr.	Real in-lab data.
[62]	Phase Prediction+ Decoupling	CNN	I/Q	Few-shot	Supervised	10 vs 6, Accuracy: ∼62%	✓	✓	5 to 30 by 5 incr.	0 to 30 by 5 incr.	–

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction, - Predictive

from known devices and clustering these features. The proposed architecture outperforms the baseline in metrics such as clustering purity, completeness, and isolation. For comparison, the baseline involves direct clustering of I/Q samples, with both approaches using DBSCAN as the clustering algorithm. The latent representations are taken from an intermediate (late but not final) network layer to avoid class-specific information. The clustering is performed on 2D data obtained via Uniform Manifold Approximation and Projection (UMAP) from 100-dimensional vectors provided by the MLP output layer.

Zero-shot learning is also performed in [53], where the proposed approach assumes adding clustering loss on the feature vectors extracted by the same architecture as used in supervised classification [111], with an additionally defined feature representation layer, thus creating a deep clustering method (Section III-A). Clustering loss is a complex function consisting of three separate parts, enforcing the separability of the clusters during training. Known and unknown transmitters are apparent in the training set, helping the model to create clear cluster boundaries. Single, known device detection is performed with 89.1% accuracy.

2) CONTRASTIVE

Contrastive learning is a recently introduced method in the SEI domain. Although some of the approaches require labels [102], [112] for the pre-training, they are suitable for few-shot learning, which is the most often used approach for evaluation of the unsupervised approaches [52], [62], [103]. Thus, we consider these methods as borderline cases and still include them in this review, as they rely on contrastive loss and meta-learning strategies, i.e., concepts which are applicable in unsupervised learning, and require a small amount of labeled data [102].

In [101], the authors propose a lightweight CNN-based architecture for contrastive learning (see Section III-B). They validate it in several scenarios on the open dataset with 20 transmitters, including the open-set recognition of new unknown transmitters [112]. Their method shows stable performance with the accuracy above 70%, considering various numbers of unknown transmitters (up to 5). It is important to note that, to the best of our knowledge, this is the only work investigating how the disappearance of known transmitters affects the performance. One significant constraint regarding clustering is the necessity of labeled datasets for the initial phase of training the feature extractor.

Contrastive clustering adaptation to I/Q signals for the SEI challenge is proposed in [42]. They use a CNN-based (Section IV-B) feature extractor and evaluate it with K-means on the extracted feature space. Such a model outperforms the baseline clustering methods, existing deep clustering, and generative approaches in novel device detection tasks. However, there is a constraining requirement of knowing the number of classes (transmitters) in the training phase.

Meta-contrastive learning method for SEI is proposed in [102], where the adaptation to the new unseen devices is approached as general domain adaptation, designing a dedicated loss function that enforces both device-specific features development and also general features. While the model shows remarkable performance, it is worth noting that labeled data is necessary for the training phase. This is an important direction of work, considering that there are many image-processing and time-series [26] processing frameworks that could be utilized with such domain adaptation.

Contrastive learning as a pretraining step of a deep CNN is also performed in [103], where, besides the common data augmentations such as rotation and flipping of samples, additional model-driven feature-space augmentation is proposed. Custom augmentation works by introducing offsets in the feature space and mapping the perturbed feature back to the original sample. Such an innovative approach allows for custom augmentation, which introduces maximal changes to the samples while still maintaining a stable feature space. It outperforms the standard contrastive learning methods with only standard augmentations by a significant margin in scenarios with a very small amount of labeled data.

3) RECONSTRUCTION

Masked AE (MAE) (see Section III-C) architecture for SEI is proposed in [52] for unsupervised pre-training of a feature encoder as part of the development of an SEI classifier in a scenario with a limited amount of training samples. The authors evaluate the approach using multiple masking techniques. They prove that the introduction of MAE leads to consistently better classification performance compared to the purely supervised models with different numbers of training samples, evaluating both simulated and real data.

4) PREDICTIVE

In [62], the authors propose a three-part pretext task for training an encoder on unlabeled data. Each of the three parts, I/Q sample rotation classification (see Section III-E), reconstructing the original sample from the rotated sample, and contrastive learning for distinguishing noise from real samples, has its own loss calculation, together building the total loss. The final model is capable of achieving significant classification accuracy above 90% on downstream tasks using only 30 labeled samples, significantly outperforming the purely supervised model on the same number of labeled samples. However, this work does not provide performance comparisons with other existing approaches, such as the pure contrastive ones proposed in [42] and [102].

C. DATASETS

In this section, we report on some of the most relevant datasets for exploring and addressing the SEI challenge, also summarized in Table 7. While there are many datasets,[1,2] available for this task, we will consider the most comprehensive ones.

In [112], the authors publish a large 7TB real-world dataset containing signals of 20 devices with similar hardware communicating over Wi-Fi. The data are collected in three setups, namely, in an anechoic chamber, in the wild, and using a cable for transmission between the transmitter and receiver. This is a very versatile dataset that could be used to explore how environmental conditions affect SEI performance, such as the multi-path and the channel effect. Considering the amount of data, it makes a valuable starting point for the development of large unsupervised learning models.

TABLE 7. Summary of datasets for SEI challenges.

Ref.	Characteristics	Specific Features
[114]	Real-world dataset of signals from 20 devices communicating over Wi-Fi. Data was collected in three setups: anechoic chamber, in the wild, and cable transmissions.	A versatile 7TB dataset that explores environmental effects on SEI performance, such as multi-path and channel effects. Suitable for developing large unsupervised models.
[22]	ADS-B airplane signals operating at 1090 MHz. Includes short and long signals with 1713 and 1670 device categories.	Automatic data collection procedure. Evaluation of deep learning models on the dataset could guide feature extraction module selection for unsupervised approaches.
[115]	RF fingerprinting dataset with transmissions of 174 Wi-Fi transmitters and 41 USRP receivers in the 2.4 GHz band. Captures made over a month.	Collected in a closed space using the ORBIT[1] testbed. Includes small subsets for fast experiments tackling challenges like channel variations and the effect of multiple receivers.

Over-the-air captured data of airplanes, using the ADS-B system operating at 1,090 MHz, is described in [22]. The authors propose an automatic data collection procedure and demonstrate it by creating two subsets of data, with short and long signals, with 1,713 and 1,670 categories/devices. They also evaluate various deep learning models on the data. Although they consider only supervised approaches, such an evaluation could be useful in the selection of the deep learning module, which most unsupervised approaches also contain for the feature extraction process.

Another open and large RF fingerprinting dataset is reported in [113], containing transmissions of 174 Wi-Fi transmitters and 41 USRP receivers, with the captures made over a month. The transmissions are made in the 2.4 GHz band. The data is captured in a closed space and controlled environment using the ORBIT[4] testbed. The authors also provide small subsets of the data that could be used for fast experiments, tackling different challenges, such as channel variation influence over time and the effect of utilization of multiple receivers.

D. FINDINGS

Unsupervised deep learning shows significant potential for tackling the SEI challenge, especially given the promising early results achieved with architectures adapted from other domains such as contrastive learning [101] and predictive learning [62]. These approaches demonstrate the capacity to extract meaningful representations from unlabeled RF data and distinguish among signal sources with minimal supervision.

[1]https://genesys-lab.org/mldatasets
[2]https://ece.northeastern.edu/wineslab/datasets.php
[4]https://orbit-lab.org/

1) EVALUATION PRACTICES

In general, two dominant evaluation strategies have emerged in the literature. The first corresponds to zero-shot setups, where models are tested on devices unseen during training, with or without the use of labels in evaluation. The second covers few-shot learning approaches, which involve fine-tuning with a small number of labeled samples (e.g., 5, 10, or 30) before evaluating classification accuracy. While both strategies are useful for assessing generalization and sample efficiency, there is substantial inconsistency in dataset selection, the number of devices, and task configurations. Many studies limit their evaluations to a small number of transmitters, often fewer than 10 to 20. This does not reflect the growing density and heterogeneity of wireless deployments in real-world indoor and outdoor environments. Although this limitation does not require a fundamental change in evaluation methodology, scaling the number of devices in benchmark evaluations is essential. Future research should emphasize large-scale validation experiments involving tens or even hundreds of devices, which would bring current research efforts closer to realistic deployment scenarios.

2) ROAD AHEAD

Despite the encouraging technical results, most existing solutions remain at the proof-of-concept stage, primarily evaluated in laboratory settings or fixed testbeds with static channel conditions and controlled interference [113]. While such setups are useful for dissecting the impact of specific channel or device-level features (e.g., hardware impairments), they fail to capture real-world complexities such as device mobility, spectrum congestion, and dynamic multipath environments. Consequently, the robustness of these models to environmental variation remains underexplored. There is a clear need to expand evaluations to real-world deployments, such as those examined in [114], [115] and [116], and to consider different channel models [117]. They should include both indoor and outdoor scenarios, particularly in use cases involving mobile devices, such as uncrewed aerial vehicles, for which the channel variation is very significant. Incorporating such conditions into benchmarking would better inform both model limitations and design priorities for real-world deployment.

A further critical issue is the lack of reproducibility. Many studies do not release their datasets, code, or full experimental details, impeding fair comparison across approaches and hindering progress toward standardized evaluation. This fragmentation not only limits scientific transparency but also slows the development of unified benchmarks. To address this, future research should prioritize the use and extension of open, community-maintained datasets, such as WISIG [113] and RF-DNA [116], and adhere to transparent evaluation protocols. Establishing widely accepted benchmarks will facilitate fair comparison, reproducibility, and ultimately, progress toward deployable and robust unsupervised SEI systems.

VIII. ANOMALY DETECTION

Anomaly detection in radio signal sensing helps identify unusual events and irregular patterns. These anomalies may indicate general malfunctions, equipment failures, or potential security threats that require immediate and appropriate actions [121]. However, in real-world settings, the presence of anomalies is sparse, so we often lack labeled data [122]. To address this challenge, unsupervised learning techniques have gained prominence in anomaly detection for radio signal sensing. These methods can learn normal patterns from unlabeled data and subsequently identify deviations that may correspond to anomalous activities [123], [124]. Moreover, the dynamic and often unpredictable nature of radio environments adds another layer of complexity to anomaly detection. Factors such as interference, signal attenuation, and device heterogeneity can lead to fluctuations that resemble anomalies but are not necessarily indicative of security threats or system failures. In general, reconstruction or generative methods are mostly utilized for unsupervised anomaly detection.

A. OVERALL AD COMPARATIVE OVERVIEW

In Table 8 we summarize the unsupervised anomaly detection works, noticing that most of them rely on generative methods (Section III-D), while two rely on predictive methods (Section III-E) as can be seen from the second column of the table and also highlighted with different colors. According to the third column, six of the architectures contain CNNs (see Section IV-B), out of which two are in combination with LSTM (see Section IV-C), and one is in combination with MLP (see Section IV-A. The fourth column reveals that four of the works rely on spectrogram data, three on I/Q data, and one on Stockwell Transform amplitudes, which are pictorial data.

Overall, the reviewed studies demonstrate that unsupervised deep learning methods can effectively detect radio frequency anomalies even under challenging SNR conditions. Most of the evaluated systems achieve high detection accuracy or strong anomaly discrimination (e.g., AUC above 0.9) using synthetic or limited real-world data. However, differences in input representations, SNR ranges, and evaluation metrics make direct quantitative comparison difficult. The limited availability of public datasets and code repositories, as indicated in the last two columns of Table 8, also constrains reproducibility and benchmarking across studies. In the subsequent sections, we further analyze the generative and predictive approaches in terms of their underlying architectures, training objectives, and generalization capabilities under variable noise and interference conditions.

B. OVERVIEW PER APPROACH TO AD

1) GENERATIVE

The reconstruction error of GANs and AEs (see Section III-D) can also serve as a metric for anomaly detection, allowing for pinpointing unusual behaviors effectively. In [69], the authors propose SAIFE, an adversarial

TABLE 8. Summary of works concerning unsupervised deep learning approaches and achievements for RF Anomaly detection.

Ref.	Approach	Module	Input	SNR	Num. clas.	Performance	Data	Code	Num. params.
[69]	AE + GAN	CNN + LSTM	I/Q	-20dB to 20dB	2	Accuracy: synthetic data 92.86%, real data 100%	✓	-	-
[55]	VAE	CNN	Spectrogram	-	4	PER [55] score AUC 0.98	✓	✓	≈22.5M
[120]	Enhanced GAN (E-GAN)	CNN + MLP	Spectrogram	-20dB to 20dB	2	Probability of detection 1 at -3dB ISR	-	-	-
[121]	GAN	CNN	FFT amplitudes TS	-20dB to 20dB	2	F1 score > 0.95	✓	-	17.5M
[122]	GAN and VAE	CNN	Stockwell Transform (amplitudes)	-	2	AD AUC > 0.97	-	-	-
[63]	Predictive learning	LSTM + MDN	I component (I/Q)	40dB	2	/	-	-	-
[64]	Predictive learning	CNN + LSTM	Spectrogram data	>5dB	3	Probability of detection 90%	-	-	4.7M

Legend: - Generative, - Predictive

AE-based model for wireless spectrum anomaly detection, achieving high accuracy with interpretable feature learning. Similarly, [55] demonstrated that VAE reconstruction can be used to detect anomalies in unauthorized bands by identifying the rise in the spectrogram's noise floor after reconstruction.

The study in [118] proposes a radio anomaly detection algorithm using a modified GAN with an encoder (E-GAN). It applies the Short Time Fourier Transform (STFT) to convert RF signals into spectrograms and detects anomalies based on reconstruction error and discriminator loss. This method also enables anomaly localization in the time-frequency domain. In [119], a similar result was shown where the authors demonstrated the potential of GANs (See Section III-D) in detecting anomalies in the spectrum and mitigating security attacks. Additionally, [120] also applied unsupervised approaches on spectrum anomaly detection in mmWave radios; more specifically, they utilised the Conditional GAN and Auxiliary Classifier, GAN, and VAE.

2) PREDICTIVE

In [63], authors proposed an unsupervised anomaly detection method using a combination of LSTM and mixture density networks (MDNs), i.e., LSTM-MDN, for time-series data in digital radio transmissions. The model learned the expected signal distributions and detected anomalies using negative log-likelihood. On the other hand, the authors in [64] utilized a Deep Predictive Coding Neural Network (see Section III-E) on spectrograms of RF signals to detect anomalies such as jamming, chirping of transmitters, spectrum hijacking, and node failure.

C. DATASETS

In this section, we report on some of the openly available datasets, summarized in Table 9 for exploring and addressing the anomaly detection.

In [125], a wideband spectrum monitoring dataset is introduced with annotated anomalous signals. This dataset uses licensed frequency bands and STFT spectrograms to capture both time and frequency characteristics. Next, [126] proposed an RSSI-based synthetic anomaly detection dataset. The authors injected 4 different types of anomalies into the Rutgers dataset and also provided code for their generation. Finally, the dataset used in [127] focuses on RF jamming scenarios. Experimentally measured spectral scan data provides a baseline for the evaluation of jamming detection algorithms.

TABLE 9. Summary of open datasets for Anomaly detection.

Ref.	Characteristics	Specific Features
[127]	Real-world dataset of 5G spectrum data	This dataset comprises anomalous signals that have been manually annotated and leverages licensed frequency bands along with STFT spectrograms to capture both temporal and spectral features.
[128]	Synthetic dataset of RSSI wireless signals	Provides a script for generating "unlimited" samples of 4 different anomaly types.
[129]	Measured RF jamming data	Comprises of spectrum data with jamming anomalies

D. FINDINGS

Predictive and reconstruction-based approaches each have distinct strengths and challenges when applied to anomaly detection in the wireless spectrum. Predictive methods, such as LSTM-MDN and Deep Predictive Coding Neural Networks [63], [64], are particularly effective in detecting temporal anomalies by learning expected signal behaviors and identifying deviations. These methods excel in time-series analysis and can capture anomalies that emerge as disruptions in the expected signal evolution. However, the predictive methods are, in general, less robust, hence the reconstruction-based techniques appear more frequently in the literature.

Reconstruction-based techniques, including AEs, VAEs, and GANs [55], [69], [118], [119], [120], leverage the inability of generative models to accurately reconstruct anomalous signals. These models have demonstrated robustness against varying noise levels and the ability to operate in unsupervised settings. The advantage of using GAN-based and VAE-based methods lies in their ability to model complex signal distributions. Additionally, these approaches do not require any labeled data.

1) ROAD AHEAD

Despite their advantages, both approaches face challenges in real-world deployments. Predictive models may struggle with unseen spectrum events that do not follow expected patterns, while reconstruction-based models are harder to train and learn proper latent representation of non-anomalous examples. Future research could focus on hybrid models that integrate predictive and reconstruction-based mechanisms to leverage the strengths of both paradigms, improving anomaly detection reliability in dynamic spectrum environments.

IX. CHALLENGES AND FUTURE DIRECTIONS

In this section, we concisely summarize the identified challenges and potential future research directions, based on the analysis in each of the four research topics tied to signal recognition in wireless communications. Furthermore, we also provide a visual summary highlighting the most important aspects in Figure 18.

1) AUTOMATIC MODULATION CLASSIFICATION

In AMC, research covers a broad range of SNR levels, channel conditions, and datasets, yet most studies still rely on simulated benchmarks such as RadioML, which only partially reflect real-world variability. Future work should focus on evaluations using over-the-air (OtA) collected data that reflect realistic propagation, hardware, and interference effects. Another important direction is small-sample analysis, where current inconsistencies in labeled data ratios constrain the direct comparison. Establishing standardized benchmarks with fixed sample sizes and unified training protocols would improve reproducibility and comparability. Among unsupervised methods, contrastive learning stands out for its strong performance and ability to learn discriminative features through data-specific augmentations; thus, designing effective, domain-relevant augmentations is another promising research direction. Finally, hybrid architectures that include sequence-processing components such as RNNs or attention mechanisms have shown superior accuracy and label efficiency, warranting further exploration using real-world datasets under practical operating conditions.

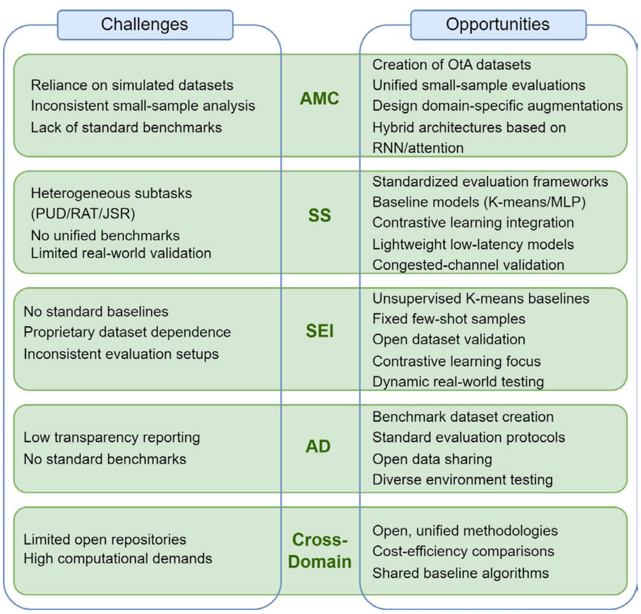


FIGURE 18. Summary of existing challenges and future research directions for each domain of work.

2) SPECTRUM SENSING

Spectrum sensing encompasses three complementary tasks: PU detection, RAT recognition, and JSR, each with distinct challenges that complicate unified evaluation. Developing benchmark datasets and standardized evaluation frameworks for these tasks would greatly enhance comparability across studies. Including simple reference models, such as K-means for unsupervised and MLP for supervised settings, along with common metrics like the Silhouette score, could streamline benchmarking and reduce dependence on highly application-specific designs. Future work should also explore contrastive learning, which has shown strong feature extraction capability in AMC and SEI, as a promising direction for spectrum sensing. Given the prevalence of interference and noise in this domain, rigorous validation under congested channel conditions with multiple PUs is essential for assessing deployment readiness. In terms of architectures, lightweight models are preferable for low-latency PU and RAT detection.

3) SPECIFIC EMITTER IDENTIFICATION

For the SEI task, both unsupervised and supervised evaluation setups are common, corresponding to zero-shot and few-shot learning, respectively. Establishing a standard unsupervised baseline, such as K-means, together with consistent clustering metrics, would help define lower performance boundaries and improve comparability across studies. For the supervised few-shot setting, standardizing the number of labeled samples would enhance transparency and facilitate fair evaluation, similar to practices in AMC research. Although most studies already use real-world data collected from proprietary testbeds, transparency could be improved by including validation with open datasets, such as the ones reviewed in this work. In terms of learning approaches, contrastive learning remains the most widely adopted and effective paradigm, while hybrid architectures that include sequence-processing modules such as LSTM or self-attention show strong potential for capturing temporal and contextual features. Future research should emphasize real-world evaluations under dynamic conditions involving mobile devices and changing propagation environments, and should aim at complementing proprietary experiments with open, standardized datasets to ensure reproducibility and comparability.

4) ANOMALIES DETECTION

The AD domain currently exhibits the least transparency, with limited information available regarding model architectures, datasets, and code accessibility. This gap highlights an urgent need for the introduction of standard benchmark datasets, baseline algorithms, and evaluation methodologies to enable comparability and reproducibility.

5) FINAL REMARKS

Across all four domains, a recurring challenge is the lack of open science practices such as open-source code, open data repositories and standardized benchmarks. This limits reproducibility and slows progress, as researchers must repeatedly replicate experimental setups that could otherwise serve as shared baselines. Adopting standardized methodologies, including open data, baseline algorithms, unified metrics, and public access to implementation materials, would greatly benefit future research. Together with the channel conditions and the SNR analysis, they represent the key common, domain-agnostic resources based on which a transparent and fair comparison of different models is possible for each of the research topics surveyed in this work. In Figure 19, we show the degree of availability of the corresponding resources, relative to the number of surveyed works, for each specific topic. As can be seen, and also noticed before, the AMC and SEI domains are the most thoroughly analysed, and the AD domain is the least transparent one.

It is important to recognize that all unsupervised deep learning models depend on the intensive processing of large volumes of unlabeled data, which is often taken for granted in the literature. The substantial computational and memory resources required for such training should be carefully considered in future work, besides the size of the models. If a supervised model can achieve comparable performance while requiring only a fraction of the computational cost, the unsupervised approach may not always represent the most efficient or practical solution.

Finally, we note that current works lack considerations regarding throughput/latency trade-offs, largely due to the focus on the model's performance. However, such considerations that are more related to a model operation once it is deployed, may affect the performance of network applications, therefore should be considered in the future, together with energy consumption and CO_2 footprint of models.

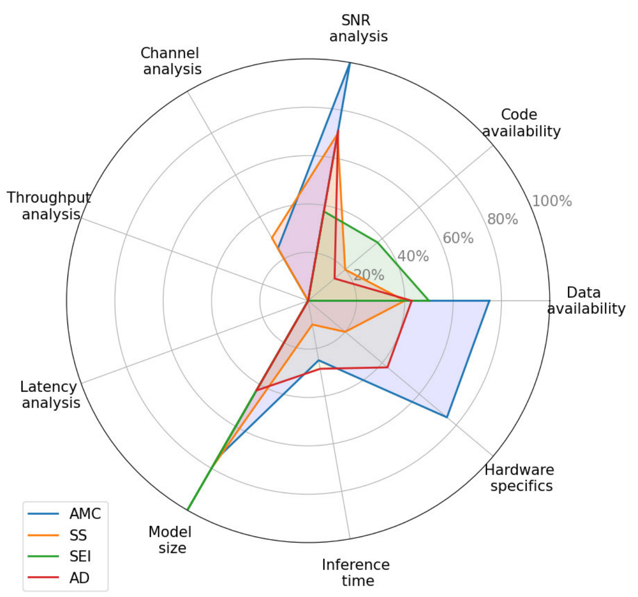


FIGURE 19. **Web plots of the key common requirements for comparative analysis of the published works across the four signal recognition challenges.**

X. CONCLUSION

This survey has provided a comprehensive overview of the current state of unsupervised deep learning techniques applied to wireless signal recognition, focusing on key challenges, i.e., AMC, signal sensing, SEI, and anomaly detection. We have systematically reviewed deep clustering, contrastive learning, autoencoder-based reconstruction, generative models, and predictive learning, highlighting their potential to address the limitations of supervised methods, particularly the reliance on large labeled datasets. We have complemented this method-oriented review with an overview of the most representative openly available datasets that can be utilized for further research in the area of a given wireless signal recognition challenge.

By presenting both a challenge-driven and a method-oriented perspective, this survey aims to inspire further research and provide a roadmap for the development of data-efficient, self-supervised models that can enhance the adaptability and intelligence of future wireless networks.

Despite notable progress in recent years, the application of unsupervised deep learning in wireless communications remains relatively underexplored compared to some other domains such as computer vision and natural language processing. Significant research opportunities exist in improving the scalability and adaptability of these models to dynamic and complex radio environments, some of them also identified in discussions within sections corresponding to key challenges.

REFERENCES

[1] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an LSTM auto-encoder," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 370–382, Jan. 2022.

[2] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cognit. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.

[3] J. Robinson, S. Kuzdeba, J. Stankowicz, and J. M. Carmack, "Dilated causal convolutional model for RF fingerprinting," in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2020, pp. 0157–0162.

[4] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," *Comput. Commun.*, vol. 151, pp. 331–337, Feb. 2020.

[5] L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," *Comput. Commun.*, vol. 212, pp. 353–365, Dec. 2023.

[6] 3GPP. (2016). *Release 13*. Accessed: Jul. 11, 2025. [Online]. Available: https://www.3gpp.org/release-13

[7] G. Reus-Muns, P. S. Upadhyaya, U. Demir, N. Stephenson, N. Soltani, V. K. Shah, and K. R. Chowdhury, "SenseORAN: O-RAN-based radar detection in the CBRS band," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 326–338, Feb. 2024.

[8] I. R. Sector, *Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond*, document ITU-R M.2160, 2023.

[9] *Representative Use Cases and Key Network Requirements for Network 2030*, document NET2030-O-027, 2020.

[10] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network automation through orchestrated intelligence in the open RAN," in *Proc. IEEE Conf. Comput. Commun.*, May 2022, pp. 270–279.

[11] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Comput. Netw.*, vol. 182, Dec. 2020, Art. no. 107516.

[12] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1376–1411, 2nd Quart., 2023.

[13] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, and J. Van der Merwe, "Digital spectrum twinning for next-generation spectrum management and metering," in *Proc. IEEE 2nd Int. Conf. Digit. Twins Parallel Intell. (DTPI)*, Oct. 2022, pp. 1–6.

[14] S. Tadik, K. M. Graves, M. A. Varner, C. R. Anderson, D. M. Johnson, S. K. Kasera, N. Patwari, J. Van der Merwe, and G. D. Durgin, "Digital spectrum twins for enhanced spectrum sharing and other radio applications," *IEEE J. Radio Freq. Identificat.*, vol. 8, pp. 376–391, 2024.

[15] R. Schwarz. (2023). *The Metaverse and Extended Reality—Implications for Wireless Communications*. Accessed: May 14, 2025. [Online]. Available: https://www.rohde-schwarz.com/us/solutions/wireless-communications-testing/wireless-standards/5g-nr/extended-reality-xr-testing/white-paper-the-metaverse-and-extended-reality_257899.html

[16] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for extended reality (XR)," *ITU J. Future Evolving Technol.*, vol. 3, no. 2, pp. 273–287, 2022.

[17] L. Milosheski, M. Mohorčič, and C. Fortuna, "Spectrum sensing with deep clustering: Label-free radio access technology recognition," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 4746–4763, 2024.

[18] S. Atapattu, C. Tellambura, and H. Jiang, *Energy Detection for Spectrum Sensing in Cognitive Radio*, vol. 6. Cham, Switzerland: Springer, 2014.

[19] T. Gale, T. Šolc, R.-A. Mosoi, M. Mohorcic, and C. Fortuna, "Automatic detection of wireless transmissions," *IEEE Access*, vol. 8, pp. 24370–24384, 2020.

[20] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *Proc. 2nd IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, Apr. 2007, pp. 212–215.

[21] F. Paisana, N. Prasad, A. Rodrigues, and R. Prasad, "An alternative implementation of a cyclostationary detector," in *Proc. 15th Int. Symp. Wireless Pers. Multimedia Commun.*, Sep. 2012, pp. 45–49.

[22] T. Ya, L. Yun, Z. Haoran, J. Zhang, Y. Wang, G. Guan, and M. Shiwen, "Large-scale real-world radio signal recognition with deep learning," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 35–48, Sep. 2022.

[23] D. Liu, P. Wang, T. Wang, and T. Abdelzaher, "Self-contrastive learning based semi-supervised radio modulation classification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2021, pp. 777–782.

[24] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, "A survey on self-supervised learning: Algorithms, applications, and future trends," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9052–9071, Dec. 2024.

[25] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-supervised and unsupervised deep visual learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1327–1347, Mar. 2024.

[26] Q. Meng, H. Qian, Y. Liu, Y. Xu, Z. Shen, and L. Cui, "Unsupervised representation learning for time series: A review," 2023, *arXiv:2308.01578*.

[27] S. N. Syed, P. I. Lazaridis, F. A. Khan, Q. Z. Ahmed, M. Hafeez, A. Ivanov, V. Poulkov, and Z. D. Zaharis, "Deep neural networks for spectrum sensing: A review," *IEEE Access*, vol. 11, pp. 89591–89615, 2023.

[28] L. J. Wong, W. H. Clark, B. Flowers, R. M. Buehrer, W. C. Headley, and A. J. Michaels, "An RFML ecosystem: Considerations for the application of deep learning to spectrum situational awareness," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2243–2264, 2021.

[29] A. Upadhye, P. Saravanan, S. S. Chandra, and S. Gurugopinath, "A survey on machine learning algorithms for applications in cognitive radio networks," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2021, pp. 01–06.

[30] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[31] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[32] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine learning for wireless communications in the Internet of Things: A comprehensive survey," *Ad Hoc Netw.*, vol. 93, Oct. 2019, Art. no. 101913.

[33] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 77–100, 1st Quart., 2023.

[34] M. Usama, J. Qadir, A. Raza, H. Arif, K.-L.-A. Yau, Y. Elkhatib, A. Hussain, and A. Al-Fuqaha, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, 2019.

[35] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. B. Le, K. Lee, and W.-J. Hwang, "Intelligent radio signal processing: A survey," *IEEE Access*, vol. 9, pp. 83818–83850, 2021.

[36] H. Zhou, J. Bai, Y. Wang, J. Ren, X. Yang, and L. Jiao, "Deep radio signal clustering with interpretability analysis based on saliency map," *Digit. Commun. Netw.*, vol. 10, no. 5, pp. 1448–1458, Oct. 2024.

[37] E. Perenda, S. Rajendran, G. Bovet, M. Zheleva, and S. Pollin, "Contrastive learning with self-reconstruction for channel-resilient modulation classification," in *Proc. IEEE Conf. Comput. Commun.*, May 2023, pp. 1–10.

[38] K. Zhang, Q. Wen, C. Zhang, R. Cai, M. Jin, Y. Liu, J. Y. Zhang, Y. Liang, G. Pang, D. Song, and S. Pan, "Self-supervised learning for time series analysis: Taxonomy, progress, and prospects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 10, pp. 6775–6794, Oct. 2024.

[39] N. Liao, Y. Zhang, Y. Wang, and Y. Liu, "Deep denoising and clustering-based cooperative spectrum sensing for non-orthogonal multiple access," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 5, pp. 1831–1842, Oct. 2024.

[40] Z. Su, K. C. Teh, S. G. Razul, and A. C. Kot, "Deep non-cooperative spectrum sensing over Rayleigh fading channel," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4460–4464, Apr. 2022.

[41] J. Stankowicz and S. Kuzdeba, "Unsupervised emitter clustering through deep manifold learning," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 0732–0737.

[42] X. Hao, Z. Feng, R. Liu, S. Yang, L. Jiao, and R. Luo, "Contrastive self-supervised clustering for specific emitter identification," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20803–20818, Dec. 2023.

[43] N. A. Khalek and W. Hamouda, "DeepSense: An unsupervised deep clustering approach for cooperative spectrum sensing," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 1868–1873.

[44] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales, "Self-supervised representation learning: Introduction, advances, and challenges," *IEEE Signal Process. Mag.*, vol. 39, no. 3, pp. 42–62, May 2022.

[45] S. Zhou, H. Xu, Z. Zheng, J. Chen, Z. Li, J. Bu, J. Wu, X. Wang, W. Zhu, and M. Ester, "A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions," *ACM Comput. Surv.*, vol. 57, no. 3, pp. 1–38, Mar. 2025.

[46] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.

[47] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan, "A comprehensive survey on contrastive learning," *Neurocomputing*, vol. 610, Dec. 2024, Art. no. 128645.

[48] K. Davaslioglu, S. Boztas, M. C. Ertem, Y. E. Sagduyu, and E. Ayanoglu, "Self-supervised RF signal representation learning for NextG signal classification with deep learning," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 65–69, Jan. 2023.

[49] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[50] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Appl. Soft Comput.*, vol. 138, May 2023, Art. no. 110176.

[51] S. Subray, S. Tschimben, and K. Gifford, "Towards enhancing spectrum sensing: Signal classification using autoencoders," *IEEE Access*, vol. 9, pp. 82288–82299, 2021.

[52] K. Huang, J. Yang, H. Liu, and P. Hu, "Deep learning of radio frequency fingerprints from limited samples by masked autoencoding," *IEEE Wireless Commun. Lett.*, vol. 14, no. 12, pp. 3842–3846, Dec. 2025.

[53] J. Robinson and S. Kuzdeba, "Novel device detection using RF fingerprints," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 0648–0654.

[54] C. Ezemaduka and A. A. Abouzeid, "Privacy-aware deep learning based localization of spectrum violators," in *Proc. IEEE 20th Int. Conf. Mobile Ad Hoc Smart Syst. (MASS)*, Sep. 2023, pp. 98–106.

[55] Y. Tian, H. Liao, J. Xu, Y. Wang, S. Yuan, and N. Liu, "Unsupervised spectrum anomaly detection method for unauthorized bands," *Space, Sci. Technol.*, vol. 2022, 2022, doi: 10.34133/2022/9865016.

[56] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–49, Nov. 2022.

[57] N. Van Huynh, J. Wang, H. Du, D. T. Hoang, D. Niyato, D. N. Nguyen, D. I. Kim, and K. B. Letaief, "Generative AI for physical layer communications: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 3, pp. 706–728, Jun. 2024.

[58] C. Zhao, H. Du, D. Niyato, J. Kang, Z. Xiong, D. I. Kim, X. Shen, and K. B. Letaief, "Generative AI for secure physical layer communications: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 1, pp. 3–26, Feb. 2025.

[59] X. Zhou, Z. Sun, and H. Wu, "Wireless signal enhancement based on generative adversarial networks," *Ad Hoc Netw.*, vol. 103, Jun. 2020, Art. no. 102151.

[60] J. Carmack, A. Bhatia, J. Robinson, J. Majewski, and S. Kuzdeba, "Neural network generative models for radio frequency data," in *Proc. IEEE 12th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Dec. 2021, pp. 0577–0582.

[61] J. Wang, H. Du, D. Niyato, J. Kang, S. Cui, X. Shen, and P. Zhang, "Generative AI for integrated sensing and communication: Insights from the physical layer perspective," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 246–255, Oct. 2024.

[62] L. Xu, W. Shi, X. Fu, H. Xu, Y. Wang, B. Adebisi, and G. Gui, "Few-shot specific emitter identification method using rotation feature decoupling for secure 6G," in *Proc. IEEE 23rd Int. Conf. Commun. Technol. (ICCT)*, Oct. 2023, pp. 490–494.

[63] M. Walton, M. Ayache, L. Straatemeier, D. Gebhardt, and B. Migliori, "Unsupervised anomaly detection for digital radio frequency transmissions," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 826–832.

[64] N. Tandiya, A. Jauhar, V. Marojevic, and J. H. Reed, "Deep predictive coding neural network for RF anomaly detection in wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[65] J. Fontaine, E. Fonseca, A. Shahid, M. Kist, L. A. DaSilva, I. Moerman, and E. De Poorter, "Towards low-complexity wireless technology classification across multiple environments," *Ad Hoc Netw.*, vol. 91, Aug. 2019, Art. no. 101881.

[66] J. Xie, J. Fang, C. Liu, and L. Yang, "Unsupervised deep spectrum sensing: A variational auto-encoder based approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5307–5319, May 2020.

[67] N. Zhang, J. Shen, Y. Shi, and Y. Li, "CNN-zero: A zero-shot learning framework for jamming identification," in *Proc. IEEE 22nd Int. Conf. Commun. Technol. (ICCT)*, Nov. 2022, pp. 1126–1131.

[68] S. Kuzdeba, J. Robinson, and J. Carmack, "Transfer learning with radio frequency signals," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–9.

[69] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, "SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Oct. 2018, pp. 1–9.

[70] W. Kong, X. Jiao, Y. Xu, B. Zhang, and Q. Yang, "A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition," *IEEE Trans. Cognit. Commun. Netw.*, vol. 9, no. 4, pp. 950–962, Aug. 2023.

[71] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2025, pp. 5998–6008.

[72] H. Han, W. Li, Z. Feng, G. Fang, Y. Xu, and Y. Xu, "Proceed from known to unknown: Jamming pattern recognition under open-set setting," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 693–697, Apr. 2022.

[73] J. Xie, J. Fang, C. Liu, and X. Li, "Deep learning-based spectrum sensing in cognitive radio: A CNN-LSTM approach," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2196–2200, Oct. 2020.

[74] C. Xiao, S. Yang, Z. Feng, and L. Jiao, "MCLHN: Toward automatic modulation classification via masked contrastive learning with hard negatives," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14304–14319, Oct. 2024.

[75] D. Wu, J. Shi, Z. Li, M. Du, F. Liu, and F. Zeng, "Contrastive semi-supervised learning with pseudo-label for radar signal automatic modulation recognition," *IEEE Sensors J.*, vol. 24, no. 19, pp. 30399–30411, Oct. 2024.

[76] Y. Shi, H. Xu, Y. Zhang, Z. Qi, and D. Wang, "GAF-MAE: A self-supervised automatic modulation classification method based on gramian angular field and masked autoencoder," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 1, pp. 94–106, Feb. 2024.

[77] Y. Wang, J. Bai, Z. Xiao, Z. Chen, Y. Xiong, H. Jiang, and L. Jiao, "AutoSMC: An automated machine learning framework for signal modulation classification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 6225–6236, 2024.

[78] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 34–46, Mar. 2021.

[79] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, T. T. Nguyen, R. Ruby, M. Zeng, and D.-S. Kim, "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142950–142971, 2021.

[80] S. Hu, Y. Pei, P. P. Liang, and Y.-C. Liang, "Deep neural network for robust modulation classification under uncertain noise conditions," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 564–577, Jan. 2020.

[81] S. Zheng, X. Zhou, L. Zhang, P. Qi, K. Qiu, J. Zhu, and X. Yang, "Toward next-generation signal intelligence: A hybrid knowledge and data-driven deep learning framework for radio signal classification," *IEEE Trans. Cognit. Commun. Netw.*, vol. 9, no. 3, pp. 564–579, Jun. 2023.

[82] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[83] Y. Wang, G. Gui, H. Gacanin, T. Ohtsuki, H. Sari, and F. Adachi, "Transfer learning for semi-supervised automatic modulation classification in ZF-MIMO systems," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 231–239, Jun. 2020.

[84] Y. Dong, X. Jiang, H. Zhou, Y. Lin, and Q. Shi, "SR2CNN: Zero-shot learning for signal recognition," *IEEE Trans. Signal Process.*, vol. 69, pp. 2316–2329, 2021.

[85] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.* Cham, Switzerland: Springer, 2016, pp. 213–226.

[86] X. Chen, S. Xie, and K. He, "An empirical study of training self-supervised vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9640–9649.

[87] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, vol. 1, 2024, pp. 1597–1607.

[88] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[89] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[90] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. (2017). *Automatic Differentiation in Pytorch*. [Online]. Available: https://api.semanticscholar.org/CorpusID:40027675

[91] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[92] J. Wang, Q. Zhu, Z. Lin, J. Chen, G. Ding, Q. Wu, G. Gu, and Q. Gao, "Sparse Bayesian learning-based hierarchical construction for 3D radio environment maps incorporating channel shadowing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14560–14574, Oct. 2024.

[93] J. Downey, B. Hilburn, T. O'Shea, and N. West, "Machine learning remakes radio," *IEEE Spectr.*, vol. 57, no. 5, pp. 35–39, May 2020.

[94] T. Šolc, C. Fortuna, and M. Mohorčič, "Low-cost testbed development and its applications in cognitive radio prototyping," in *Cognitive Radio and Networking for Heterogeneous Wireless Networks*. Cham, Switzerland: Springer, 2015, pp. 361–405.

[95] *Intelligent Transport Systems (ITS); Access Layer Specification for Intelligent Transport Systems Operating in the 5 GHz Frequency Band; Technical Report en 302 663; Version 1.2. 0; 650 Route Des Lucioles F-06921 Sophia Antipolis Cedex-France*, ETSI, Sophia Antipolis, France, 2012.

[96] M. Girmay, V. Maglogiannis, D. Naudts, M. Aslam, A. Shahid, and I. Moerman, "Technology recognition and traffic characterization for wireless technologies in ITS band," *Veh. Commun.*, vol. 39, Feb. 2023, Art. no. 100563.

[97] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2043–2052.

[98] "Amendment of the commission's rules with regard to commercial operations in the 3550–3650 MHz band," Federal Communications Commission, Washington, DC, USA, GN Docket 12-354, 2015.

[99] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[100] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "DApps: Distributed applications for real-time inference and control in O-RAN," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 52–58, Nov. 2022.

[101] X. Zhang, Y. Huang, M. Lin, Y. Tian, and J. An, "Transmitter identification with contrastive learning in incremental open-set recognition," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 4693–4711, Feb. 2024.

[102] M. Zhan, Y. Li, H. Cui, B. Li, J. Zhang, C. Li, and W. Wang, "MCRFF: A meta-contrastive learning-based RF fingerprinting method," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2023, pp. 391–396.

[103] C. Liu, X. Fu, Y. Wang, L. Guo, Y. Liu, Y. Lin, H. Zhao, and G. Gui, "Overcoming data limitations: A few-shot specific emitter identification method using self-supervised learning and adversarial augmentation," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 500–513, Oct. 2024, doi: 10.1109/TIFS.2023.3324394.

[104] L. Ding, S. Wang, F. Wang, and W. Zhang, "Specific emitter identification via convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2591–2594, Dec. 2018.

[105] X. Wen, C. Cao, Y. Li, and Y. Sun, "DRSN with simple parameter-free attention module for specific emitter identification," in *Proc. IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2022, pp. 192–200.

[106] Z. Xu, G. Han, L. Liu, H. Zhu, and J. Peng, "A lightweight specific emitter identification model for IIoT devices based on adaptive broad learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 7066–7075, May 2023.

[107] Y. Luo, X. Chen, N. Ge, W. Feng, and J. Lu, "Transformer-based device-type identification in heterogeneous IoT traffic," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 5050–5062, Mar. 2023.

[108] S. Basak, S. Rajendran, S. Pollin, and B. Scheers, "Combined RF-based drone detection and classification," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 1, pp. 111–120, Mar. 2022.

[109] J. Wang, B. Zhang, J. Zhang, N. Yang, G. Wei, and D. Guo, "Specific emitter identification based on deep adversarial domain adaptation," in *Proc. 4th Int. Conf. Inf. Commun. Signal Process. (ICICSP)*, Sep. 2021, pp. 104–109.

[110] L. J. Wong, W. C. Headley, S. Andrews, R. M. Gerdes, and A. J. Michaels, "Clustering learned CNN features from raw I/Q data for emitter identification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2018, pp. 26–33.

[111] J. Robinson and S. Kuzdeba, "RiftNet: Radio frequency large population classification," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–6, doi: 10.1109/CCNC49032.2021.9369455.

[112] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, "Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 646–655.

[113] S. Hanna, S. Karunaratne, and D. Cabric, "WiSig: A large-scale WiFi signal dataset for receiver and channel agnostic RF fingerprinting," *IEEE Access*, vol. 10, pp. 22808–22818, 2022.

[114] T. Zhao, B. W. Domae, C. Steigerwald, L. B. Paradis, T. Chabuk, and D. Cabric, "Drone RF signal detection and fingerprinting: UAVSig dataset and deep learning approach," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2024, pp. 431–436.

[115] S. Mohanti, N. Soltani, K. Sankhe, D. Jaisinghani, M. Di Felice, and K. Chowdhury, "AirID: Injecting a custom RF fingerprint for enhanced UAV identification using deep learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[116] N. Soltani, G. Reus-Muns, B. Salehi, J. Dy, S. Ioannidis, and K. Chowdhury, "RF fingerprinting unmanned aerial vehicles with non-standard transmitter waveforms," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15518–15531, Dec. 2020.

[117] B. Hua, L. Han, Q. Zhu, C.-X. Wang, K. Mao, J. Bao, H. Chang, and Z. Tang, "Ultra-wideband nonstationary channel modeling for UAV-to-ground communications," *IEEE Trans. Wireless Commun.*, vol. 24, no. 5, pp. 4190–4204, May 2025.

[118] X. Zhou, J. Xiong, X. Zhang, X. Liu, and J. Wei, "A radio anomaly detection algorithm based on modified generative adversarial network," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1552–1556, Jul. 2021.

[119] E. Ayanoglu, K. Davaslioglu, and Y. E. Sagduyu, "Machine learning in NextG networks via generative adversarial networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 2, pp. 480–501, Jun. 2022.

[120] A. Toma, A. Krayani, L. Marcenaro, Y. Gao, and C. S. Regazzoni, "Deep learning for spectrum anomaly detection in cognitive mmWave radios," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.

[121] A. Parmar, K. Shah, K. M. Captain, M. López-Benítez, and J. R. Patel, "Gaussian mixture model-based anomaly detection for defense against Byzantine attack in cooperative spectrum sensing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 2, pp. 499–509, Apr. 2024.

[122] H. Wang, L. Muñoz-González, D. Eklund, and S. Raza, "Non-IID data re-balancing at IoT edge with peer-to-peer federated learning for anomaly detection," in *Proc. 14th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, New York, NY, USA, Jun. 2021, pp. 153–163, doi: 10.1145/3448300.3467827.

[123] L. Ruff, R. A. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[124] B. Zong, S. Qi, M. R. Min, W. Cheng, C. Lumezanu, D.-K. Cho, and H. Chen, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: https://openreview.net/forum?id=BJJLHbb0-

[125] J. Kim, H. Kim, and B. Kim, "Wireless anomaly signal dataset (WASD): An open dataset for wireless cellular spectrum monitoring and anomaly detection," *IEEE Access*, vol. 12, pp. 196240–196248, 2024.

[126] G. Cerar, H. Yetgin, B. Bertalanic, and C. Fortuna, "Learning to detect anomalous wireless links in IoT networks," *IEEE Access*, vol. 8, pp. 212130–212155, 2020.

[127] A. S. Ali, W. T. Lunardi, G. Singh, L. Bariah, M. Baddeley, M. A. Lopez, J.-P. Giacalone, and S. Muhaidat, "RF jamming dataset: A wireless spectral scan approach for malicious interference detection," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 114–120, Nov. 2024.

LJUPCHO MILOSHESKI (Member, IEEE) is currently pursuing the Ph.D. degree with the International Postgraduate School Jožef Stefan. He is a Research Assistant at SensorLab, as part of the Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia. He is involved in Horizon Europe projects and has co-authored several peer-reviewed conferences and journal publications. His research interests include spectrum sensing, self-supervised learning, and generative AI for wireless networks.

BLAŽ BERTALANIČ (Member, IEEE) received the Ph.D. degree (Hons.) from the Faculty of Electrical Engineering, University of Ljubljana. He is currently a Researcher with SensorLab, Jožef Stefan Institute. His research interests include the advancement of machine learning and AI algorithms, especially in the context of time series analysis and smart infrastructures. He held several leadership positions in Slovenian Chapter with over 15 IEEE publications.

CAROLINA FORTUNA was a Postdoctoral Researcher with Ghent University, Ghent, Belgium. She was a Visiting Researcher at InfoLab, Stanford University, Stanford, CA, USA. She is currently a Research Associate Professor with Jožef Stefan Institute, where she leads SensorLab. She has led and contributed EU-funded projects, such as H2020 NRG5, eWINE, WISHFUL, FP7 CREW, Planetdata, ACTIVE, and under various positions. She has advised/co-advised more than six M.Sc. and Ph.D. students. She has consulted public and private institutions. She has co-authored over 100 articles, including IEEE Communications Surveys and Tutorials, *IEEE Wireless Communications Magazine*, IEEE Open Journal of the Communications Society, and IEEE Access. Her research interests include developing the next generation of smart infrastructures that surround us and improve the quality of our lives. She contributed to community work as a TPC Member, the Track Chair, and a Reviewer at several IEEE conferences, including Globecom and ICC.

MIHAEL MOHORČIČ (Senior Member, IEEE) is currently the Head of the Department of Communication Systems and a Scientific Counselor with Jožef Stefan Institute, as well as a Full Professor at Jožef Stefan International Postgraduate School. His research interests include advanced wireless communication systems, including mobile, satellite, and stratospheric networks; heterogeneous and ad hoc networks; wireless sensor networks; and the Internet of Things. His current research interests include AI-driven resource management in wireless communications, smart infrastructure connectivity, and intelligent sensing applications. He has contributed to over 25 international research projects in mobile and satellite communications, UAV communication systems, and wireless sensor networks, along with more than 15 national basic and applied research projects. He has co-authored more than 230 journals and conference publications, three books, and one patent. He actively serves on conference organizing committees, including as the general chair and the TPC chair.