

Explainable Anomaly Detection of 12-Lead ECG Signals Using Denoising Autoencoder



Rok Hribar and Drago Torkar

Abstract Anomaly detection is an important task in the field of medical diagnostics, especially when it comes to ECG signals. Anomalies in ECG signals can be indicative of various cardiac conditions, such as arrhythmias, atrial fibrillation, ventricular tachycardia, and other, possibly life threatening, cardiac diseases. In medicine AI-assisted anomaly detection is favorable compared to diagnoses prediction because it can identify all out-of-the-ordinary patterns and not only those that are well represented in the available data bases. Also, AI-assisted anomaly detection can help reduce the risk of medical errors due to its ability to detect subtle abnormalities that may not be easily detected by humans. However, traditional anomaly detection methods are often limited in their ability to explain why an anomaly was detected. We propose an approach to an explainable anomaly detection and denoising of 12-lead ECG signals using a denoising autoencoder. The proposed approach is based on the idea of reconstructing the original signal from a noisy version of it, and then using the reconstruction error to detect anomalies, and also pinpoint them on the ECG in a visual way. We evaluate the proposed approach on a publicly available data sets and show that it is able to detect anomalies with high accuracy and explain why they were detected. The developed framework was also implemented as a cloud-based service that enables user-friendly ECG anomaly detection with minimal software and hardware capabilities.

1 Introduction

In recent years, a great potential for utilizing artificial intelligence (AI) in medicine has been recognized that stem from the immense advances made in the field including signal processing [1], image analysis [2] and drug discovery [3]. AI algorithms are able to quickly analyze large amounts of data, such as medical images, patient

R. Hribar (✉) · D. Torkar
Jožef Stefan Institute, Ljubljana, Slovenia
e-mail: rok.hribar@ijs.si

D. Torkar
e-mail: drago.torkar@ijs.si

© The Author(s) 2024
M. Karner et al. (eds.), *Intelligent Secure Trustable Things*, Studies in Computational Intelligence 1147, https://doi.org/10.1007/978-3-031-54049-3_7

records, and lab results, and provide insights that would otherwise take a human doctor much longer to identify. AI can also help with predictive analytics, allowing doctors to better anticipate potential health issues before they arise. Additionally, AI-driven tools can assist with decision support, providing doctors with evidence-based recommendations on the best course of action for each individual patient. Finally, AI can automate mundane tasks, freeing up time for doctors to focus more on their patients' needs. However, trust in AI models is a key factor in determining how widely they are used in medicine. If clinicians and patients trust the accuracy, reliability, and safety of AI models, then they will be more likely to use them as part of their medical decision-making process. To increase trust, explainable AI (XAI) offers a way for physicians and other medical personnel to understand the reasoning behind AI models. If one knows how AI came to some conclusion one is more likely to trust it which in effect brings possibilities to fully benefit from AI in medicine.

XAI is a concept that focuses on making the predictions of an AI model more transparent, understandable, and interpretable for both clinicians and patients. In addition to building trust, XAI is especially important in medicine because it helps to ensure that the decisions being made by AI-based systems are ethically sound, medically accurate, and consistent with patient safety. XAI allows medical professionals to understand how the system has come to its conclusions and gives them more confidence in using the technology for diagnosis, treatment planning, and other clinical tasks. It also provides a layer of transparency so that patients can better comprehend the decisions being made on their behalf. For example, XAI techniques can be used to explain why a certain diagnosis was made or why a particular treatment plan was recommended by the AI system. Additionally, it could potentially reduce medical errors due to increased transparency and accountability of the AI system. Overall, XAI has great potential to increase trust in AI models and enable them to be more widely used in medicine. By providing explanations for the decisions made by AI models, not only will this help to improve the accuracy, reliability, and safety of these models but also foster better communication between clinicians and patients. With increased trust, AI models are more likely to be used more widely as stakeholders become more comfortable with its capabilities.

In this work we focus on building an XAI pipeline, for processing electrocardiogram (ECG) recordings, able to detect anomalies, explain the results and provide visualizations. ECG is an important diagnostic tool in medicine as it provides a non-invasive method of assessing the electrical activity of the heart. The ECG records the electrical signals generated by the heart, which can be used to detect abnormal rhythms and other cardiac abnormalities such as arrhythmias, conduction delays, and myocardial infarction. By analyzing these signals, physicians are able to diagnose various cardiovascular diseases. Additionally, the ECG can provide valuable information about the patient's overall health, such as their risk factors for developing heart disease or stroke. The 12-lead version of ECG recording has the ability to provide a more comprehensive view of the electrical activity of the heart. By recording twelve different leads, it can detect abnormalities that may not be visible on other tests such as single lead ECGs or chest X-rays [4]. It also allows for comparison between multiple recordings over time, which can help identify changes in

cardiac function and diagnose arrhythmias. The 12-lead ECG is especially useful in detecting myocardial infarction (heart attack) because it can show areas of decreased blood flow to the heart muscle [5]. Additionally, it can provide valuable information about the size and shape of the heart chambers, helping to diagnose certain types of cardiomyopathy. As such, the 12-lead ECG is an invaluable tool in diagnosing cardiovascular conditions.

In our work, we chose to build an anomaly detection model as opposed to a diagnoses prediction model because it can identify all out-of-the-ordinary patterns and not only those that are well represented in the available data bases which is more prudent in critical applications such as medicine and especially cardiology. Anomaly detection is an unsupervised learning approach that attempts to detect unusual behavior or outliers in data. Unlike diagnosis prediction, this technique does not require labeled data as it looks for anomalies within existing data sets. Even though the training of anomaly detection model can be more computationally expensive and sensitive as opposed to to diagnoses prediction models, it can be used in a more general setting and requires less domain knowledge to correctly present explanations in the scope of XAI.

2 Anomaly Detection and Explainability in Deep Learning

ECG recordings involve extremely complicated patterns and are somewhat unique from person to person. This is why a powerful, high capacity model needs to be selected for processing them. A model that is able to work with very large data bases, noise and perform very heterogeneous pattern recognition. Deep learning models are an excellent choice for such a task since they scale well and are very flexible [6].

To perform anomaly detection an autoencoder (AE) is the most common class of deep learning architectures which works by compressing the input data into a low-dimensional latent space, then reconstructing the output from that latent space [7]. In this sense it is an unsupervised learning model that learns the multidimensional manifold on which our data is distributed and can be easily modified to assess whether a given data instance falls to that manifold or whether it is an outlier/anomaly. In this methodology, the AE is trained on normal data instances and learns to reconstruct only the normal patterns. When applied to a general data instance, the reconstructed output can be very similar to the input (data instance is normal) or significantly different from the input (data instance is likely to be anomalous). Such way of detecting anomalies is especially suitable for XAI because it does not only tell us whether an anomaly is present but also which part of the signal is anomalous which offers a way to visualize the anomaly in human-understandable way. See Fig. 1 for a depiction of AE concept for anomaly detection.

However, there are many flavors of AEs. The most basic version (shown in Fig. 1) consists of an encoder and a decoder, both with multiple layers of neurons. The input data is compressed by the encoder into a low-dimensional latent representation which is then reconstructed by the decoder. This vanilla AE sometimes suffers from

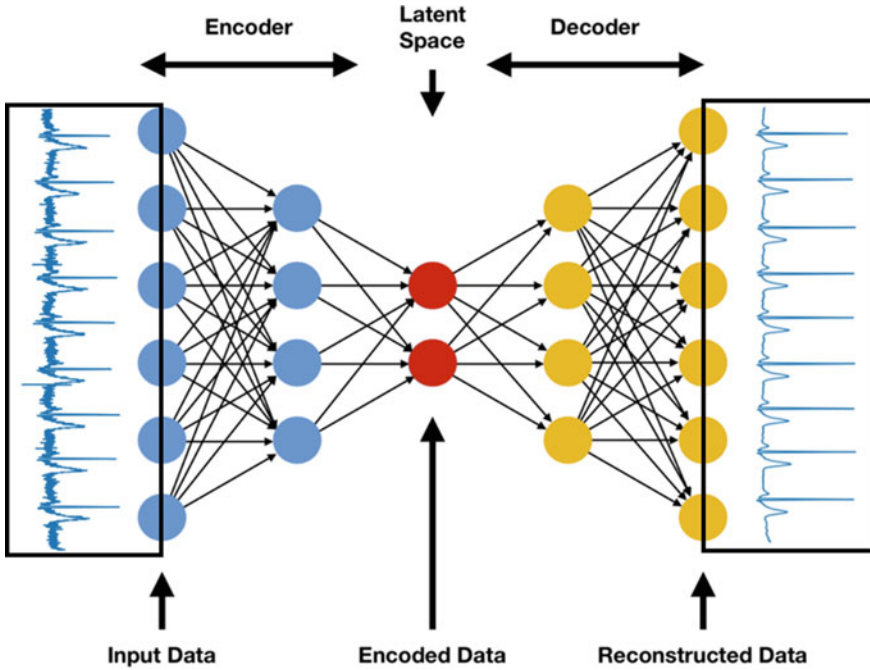


Fig. 1 Conceptual depiction of AE model when applied to ECG data. On the left AE receives a real-world ECG recording as an input and processes it so that each layer of the encoder decreases its dimensionality until a bottleneck is reached known as the latent space. If AE is trained properly small-dimensional version of the data instance encoded in the latent space includes enough relevant information to meaningfully reconstruct the ECG recording with a decoder part of AE using only the information from the latent space so that the output on the right is approximately equal to the input on the left

overfitting problems. These problems can be alleviated by using a sparse AE [8] where additional constraints are imposed on the weights in order to induce sparsity on the activations of the hidden layers during training. This helps ensure that only a few important features are encoded. Variational AE is another type which upgrades the AE into a generative model [9], meaning that it can generate new data samples that resemble the training data. It does this by learning a probability distribution over the inputs. Another type is a contractive AE [10]. This type of AE adds a regularization term to the loss function to enforce a contractive behavior on the hidden units. This helps make the model more robust to small changes in the input. Finally, there is a denoising AE (DAE) [11] which was found most useful in our work. Unlike other AEs, this type attempts to reconstruct the original input from a corrupted version of it. It turns out to be more robust to noise and have superior manifold learning capabilities. It can also be used for tasks such as removing noise from data as complex as images or text [12].

As stated before, AE can provide explanations for detecting anomalies by pinpointing regions of the ECG signal that was poorly reconstructed by the AE. Those regions are the reason that the signal was flagged as anomalous and we can also build a visual depiction for the locations of anomalies. This method of explainability, however, is not the only known method in deep learning. In recent years XAI advanced substantially, not only in deep learning but also in general machine learning. Widely applicable methods include Local Interpretable Model-Agnostic Explanations (LIME) [13] (which works by approximating model behavior locally around a prediction by creating simplified surrogate models), SHapley Additive exPlanation (SHAP) [14] (which uses game theory to explain the contribution of each feature to the model's predictions) and Anchors [15] (instance-level explanations that identify key features that lead to specific predictions). Some methods were, on the other hand, specifically designed for neural networks, such as Layer-wise relevance propagation (LRP) [16], and specifically for convolutional neural networks such as Grad-Cam method [17]. Despite the wide assortment of methods none of them was explicitly designed for time series data. Therefore, explainability for ECG processing models has been in literature performed by either adding additional explainable features designed by physicians [18], which is costly and error prone [19], or by employing methods originally designed for image data [20] which did not fully align with time series nature of ECGs [21].

3 Denoising Autoencoder as an Explainable Anomaly Detection Model for ECGs

DAE is a type of neural network used for unsupervised learning with strong connections to manifold learning. Suppose that the data set on which DAE is trained on lies on a low-dimensional manifold embedded in a full feature space of the data set. This embedded manifold can be learned by DAE by training an AE so that it can reconstruct input data that has been corrupted by noise, it denoises the instances. This means that DAE learns to project noisy data instance to a closest point on a manifold on which the full data set resides. Simply put, DAE is an AE where input data instances have added noise. The type of noise injected is typically Gaussian or salt and pepper noise (randomly sets some of the input values to one of its extreme values). In the specific case of ECG data Gaussian random walk noise was found to be useful as well. For illustration, Fig. 2 shows all three stages of DAE for ECGs, a real-world ECG input, the same ECG with injected noise, and the output ECG (i.e., the reconstructed, denoised ECG).

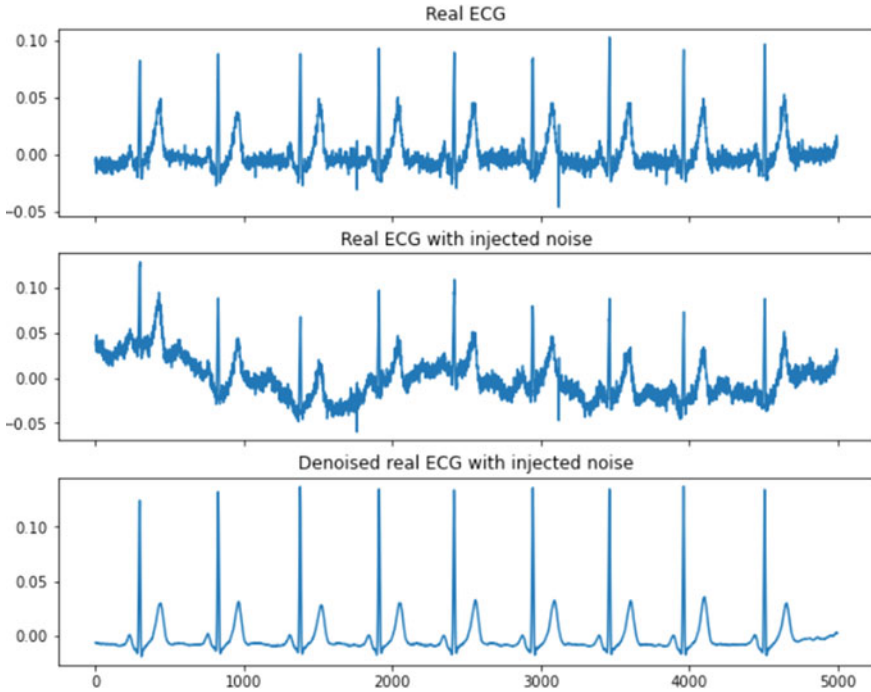


Fig. 2 How a real-world ECG recording changes when passed through a trained DAE. On top is a real-world ECG given as an input to DAE, in the middle is the same ECG but with injected noise and on the bottom is the output of DAE, a denoised ECG which is reconstructed from a low-dimensional latent space

3.1 ECG Data Sets

The data set collection used in this work was compiled for the purpose of The PhysioNet/Computing in Cardiology Challenge 2021 and includes the following data sets: CPSC Database and CPSC-Extra Database, INCART Database, PTB and PTB-XL Database, The Georgia 12-lead ECG Challenge (G12EC) Database, Augmented Undisclosed Database, Chapman-Shaoxing and Ningbo Database, The University of Michigan (UMich) Database. The most prevalent form of ECG recording in this collection is a 12-lead, 500 Hz recording that lasts for 10 s. There are minute amount of other types of ECG recordings present in the collection, e.g. with other frequencies, however, we decided to discard those in order to generate a data set with uniform properties. Each ECG recording is equipped with supplementary data including patient age, gender and diagnoses. The resulting data set we use includes 81,100 data instances of which 14,419 are pure sinus rhythm without any anomalies. There are 132 possible diagnoses in the collection and a given patient can have multiple of them.

3.2 Model Architecture and Training

To build an efficient DAE model we now need to select appropriate layers that will process ECG recordings to increasingly smaller dimension and back again. The choice of layers need to reflect the properties of the data. A single ECG data instance in our collection is a time series of 5,000 steps with each step containing a vector of 12 numerical values corresponding to the 12 ECG leads. Without considering any of the data properties we would build a DAE that gets 60,000 input numbers and transform them using dense layers (general matrix multiplication) to ever smaller dimension. However, such a model would be extremely large and computationally expensive to both use and train. It is, therefore, prudent to consider the fact that the ECG signal is a time series which means that our recordings are “continuous” in time. This means that we can reduce the dimensionality of our ECGs by contracting the number of steps in them and at the same time obtaining as much information as possible. In other words, in any given time point it is only important how the 12 leads crudely change in time and not what are the exact values that surround a given point. This leads us to use layers that act locally which means to use multiplication with a band matrix instead of the general dense matrix. If we also consider the fact that the way to process ECG should not depend on time, we are left with convolutional layers. It is important to note that recurrent and transformer layers are also well suited for time series processing, however, it is known that they are difficult to train when sequences are long which is especially true for our ECG recordings. Another drawback is that they offer less parallelism compared to convolutional layers which results to less opportunities for extensive training usually needed for (D)AE models.

We performed manual fine tuning of a DAE with convolutional layers with various properties such as kernel and filter sizes, number of layers, type of activations and so on. The most efficient model was DAE with encoder composed of 9 layers with filter and kernel sizes dropping from 64 to 4 and 16 to 4, respectively. To reduce the time dimensionality of our ECGs one can use either max-pooling or striding, however, we found that both perform similarly on our data. Because striding is more computationally efficient we used convolutional layers with stride equal to 2. To ensure well defined striding we zero-padded the input ECGs to the number of time steps equal to $5120 = 10 \cdot 2^9$. By applying 9 layers that reduce the number of time steps by factor of 2 we produced a latent space of shape (10, 4). Zero-padding also resolved edge artefacts that were present in former models.

Each convolutional layer is followed by a batch normalization and than a leaky rectified linear unit (ReLU) layer. We also tested models with skip connection with various topologies, however, contrary to the findings in literature, they did not improve the training in any way. Since skipping produces additional computational load and does not bring any benefits in our case we omit it from our end model. The decoder used is simply a mirror image of the encoder with convolution layers substituted for transposed convolution layers and zero-padding for cropping. For training we used absolute value metric and Adam algorithm. The model was trained for 20 epochs with batch size of 32 on a set of 11,535 (80% for training and 20% for testing) recordings

with pure sinus rhythm normalized with a constant factor of $1/2^{12}$. It is important to note that model accuracy was quite sensitive to the type and level of noise injected before the input layer. In our case the following noise was found the most useful:

$$\text{noise}_t = \frac{1}{\sqrt{t + 3000}} \left[\sum_{i=-3000}^t U(-0.1, 0.1) \right] + N(0, 0.003), \quad (1)$$

where $t \in \{1, \dots, 5000\}$ is a time step index, $U(a, b)$ is a variate drawn from uniform distribution on interval $[a, b]$ and $N(\mu, \sigma)$ a variate drawn from normal distribution with mean μ and standard deviation σ . The random walk part of the noise was found to be important, we suspect that this is due to its similarity to the noise actually observed in ECGs. The shift of 3000 steps has a role of burn-in that guarantees statistical independence.

3.3 *Results of Denoising and the Exploration of the Latent Space*

The trained DAE model is able to process a real-world ECG recording that may include noise (e.g., due to random body movements and respiration, power line and external electromagnetic interference) or medical anomalies and returns the closest approximation of the same ECG as if it would not have any noise or anomalies. The denoised ECG is reconstructed from only 40 floating point numbers, however, comparing it to the original ECG shows that it holds enough information and captures positions and shape of peaks really well together with patient specific artefacts. In Fig. 2, we show three examples of real-world ECG alongside the denoised, reconstructed ECG given as the output of our DEA. The height of the peaks does not align completely with the real-world ECG, however, we expect to improve that in future work. If medical anomalies are present in the ECG our DAE is unable to reconstruct them because DAE was not trained on such ECGs, it can only reconstruct normal ECGs which makes it a suitable anomaly detection model. In Fig. 2, we show two anomalous ECGs and how our DAE model attempts to reconstruct them as if they are normal ECGs. One anomaly type in Fig. 2 is localized while the other is not. We can observe in what way the differences between original and the reconstructed ECG manifest and how the explainability model put on top of our DAE can be implemented (Fig. 3).

We see that the 40-dimensional latent space can encode all the relevant features of the ECG which means that it can be used further as a compact encoding for ECGs. As far as we know from literature [22], latent space possesses special structure that is semantically meaningful, therefore, it is interesting to explore how ECGs in our data base are distributed in this 40-dimensional space. DAE reduces the raw ECG recording of shape (5, 000, 12) to an encoding of shape (10, 4) in the latent space.

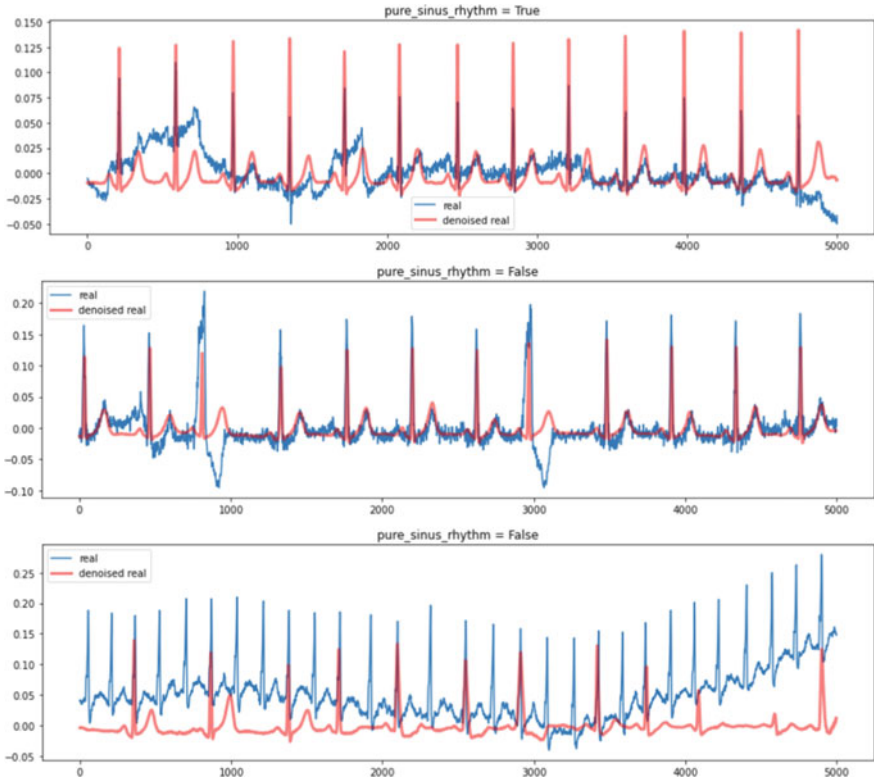


Fig. 3 Three examples of application of our DAE on real-world test ECGs. On top there is an ECG without anomalies with clearly observable natural noise, in the middle there is an ECG with two localized visually observable anomalies, and at the bottom an ECG with non-localized anomaly. All plots show only the first ECG lead

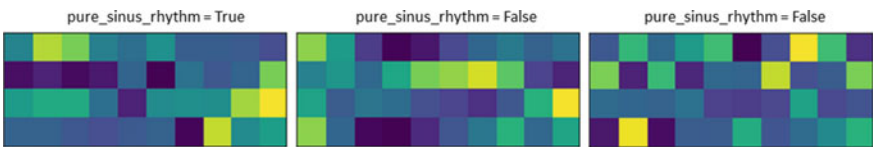


Fig. 4 Visual depiction of the same three examples as in Fig. 2 but as encoded in our 40-dimensional latent space in DAE. Instead of using original ECG of shape (5000, 12) we can represent them with a reduced shape of (10, 4). Even though the reduction is enormous this encoding holds all the relevant information to reasonably reconstruct the original ECG (minus the noise)

Figure 4 shows three examples of ECGs encoded in this space, however, it is difficult to see any evident structure.

What is informative is to study the distribution of the ECGs in our data base. Figure 5 shows a 2-dimensional projection of this distribution using UMAP method [23] which generates a scatter plot in such a way that the distances between the

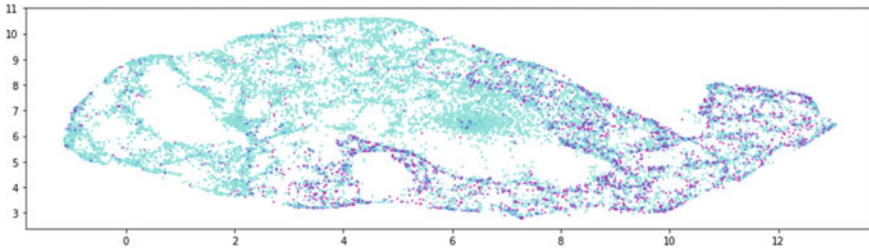


Fig. 5 2-dimensional depiction of the distribution of both anomalous (cyan) and non-anomalous (magenta) test ECGs as encoded in the 40-dimensional latent space of our DAE. We can observe partial separation between the two classes

points are as close as possible to the ones in full 40-dimensional latent space. What we can see is that normal ECGs are distributed in a special subregion of the space and partially overlaps with anomalous ECGs. In other words normal ECGs are a special kind of anomalous ECG, such as whiteout or an anomaly.

Given that the latent space encoding of an ECG holds enough information to reasonably reconstruct it, it can also be used for other ECG processing tasks. Instead of using raw ECG we could use this compact 40-dimensional encoding to perform, for example, parameter extraction and diagnosis prediction. To demonstrate this strength of the latent space encoding produced in our DAE we built a very simple classifier model that predicts whether an ECG is a pure sinus rhythm or not. The input to this simple model is solely the encoding as seen in Fig. 4. The simple classification model is a neural network with two dense hidden layers and returns a probability that the ECG is a pure sinus rhythm. The structure of the model can be seen in Fig. 6 alongside with the model performance.

4 Cloud-Based Service and Visualization of Explainable Anomaly Detection on ECGs

The denoising autoencoder model can now be used on any standard 12-lead 500 Hz ECG recording to denoise and detect anomalies. In order to maximise the widest possible use of this methodology we aspire to make it as user-friendly as possible. To this end we provide all the necessary software inside a docker image [24] and is ready to use without any other software requirements. This image includes several trained models, functions to apply the model and visualize the result alongside with all libraries that are required. We also provide a cloud-based service to perform ECG anomaly detection on a server which can be used without any programming knowledge [25]. This service uses FastAPI and applies our model to a desired ECG recording and returns a visualization of the ECG alongside with the annotations of regions where the model has detected anomalies. Currently the service operates on

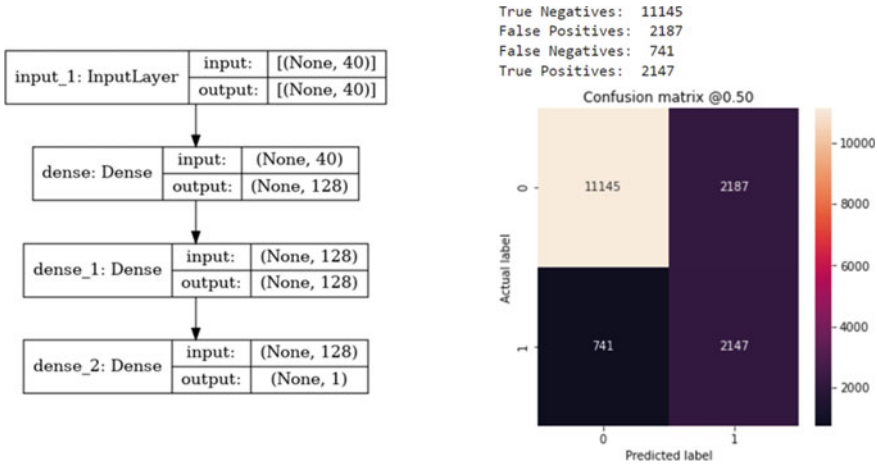


Fig. 6 The structure of a simple prediction model that takes the latent space encoding of ECG as input and returns the probability that the ECG is non-anomalous (left). A confusion matrix showing different classification metrics for such a simple model (right)

an ECG data base that is stored on the server, however, we intend to expand this to allow uploading of ECGs in FHIR compatible format in safe and secure way.

The visualization of the explainable anomaly detection is performed by first observing deviation between the real and the reconstructed ECG recording. We empirically found that the relevant anomalies exceed a threshold of 820mV. We color code the degree of deviation that exceeds this threshold and superimpose it alongside the original ECG recording. An example of such visualization is shown in Fig. 7. To reduce the noisiness of the visualization, which results from the noise present in the original ECG recording, we convolve the colors with the time window of 60ms which results to visually pleasing representation of the positions of anomalies present in the ECG.

5 Conclusion

In this work, we show that anomaly detection can be done using pure data-driven methodology without using expert knowledge and even on signals as complicated as ECG recordings. We found denoising autoencoder model to be the most effective for this task and it is able to compress 60,000-dimensional ECG to a mere 40-dimensional vector that holds all the relevant information including heart rate, respiratory rate, PR and QT interval, PR and ST segments, shape of QRS complex and other, even distinct features seen in individual patients. What is interesting is that this catalog of shapes that are statistically common in ECGs were extracted directly from the data. We found that this 40-dimensional encoding holds even an information about the

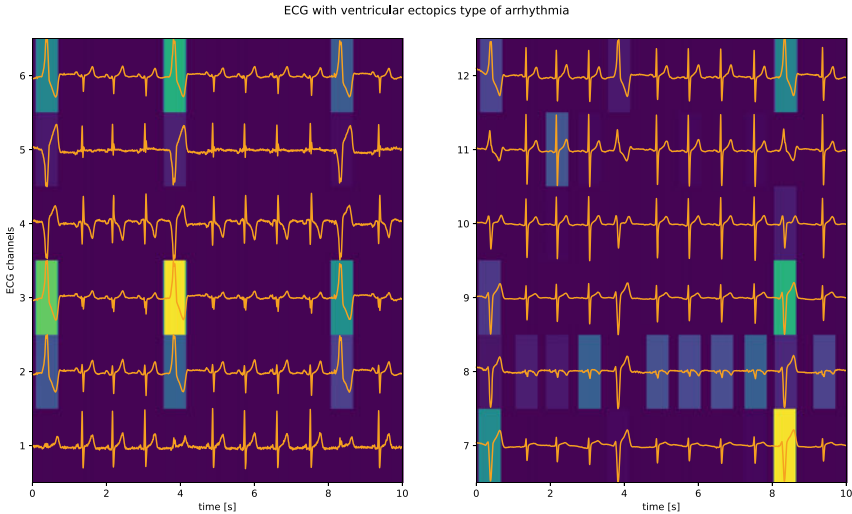


Fig. 7 A visualization of the result of explainable anomaly detection with denoising autoencoder on an example of a ECG recording with ventricular ectopics type of arrhythmia. The sections of ECG where the deviation between original and reconstructed ECG is large is color coded with yellow tones indicating large discrepancy and with blue tones low discrepancy. By this we can show the positions of medical anomalies in the ECG recording and provide not only whether the anomaly is present but also where to look for it in the recording

diagnoses and can be used to construct a simple prediction model. The signal can be reconstructed back from this 40 numbers, however, in a way that the reconstructed ECG does not include neither noise or medical anomalies present in the original ECG recording. We use this fact to construct an explainable anomaly detection model that can both tell if ECG includes anomalies and where exactly in the ECG they are positioned. To maximize the use of our methodology we provide a way to visualize the ECG with anomalies annotated using color codes and provide a user-friendly cloud-based service to perform it using the simplest possible hardware and software.

For future work it is important to further explore the possibilities for disentanglement of medical anomalies and anomalies that result from noise. One way to find more understanding is to stack two denoising autoencoders where each of them is trained in a way so that it can only reconstruct one type of anomaly. So that we have one model for removing noise and one for removing medical anomalies. Another possibility is to use semi-supervised learning and train a model that not only removes anomalies but at the same time predicts the diagnoses. This couples the representation in the latent space more tightly to the medical interpretation and possibly result to disentanglement of noise anomalies and medical anomalies in the latent space itself which can be studied due to its low dimension.

References

1. Rim, B., Sung, N.-J., Sedong, M., Hong, M.: Deep learning in physiological signal data: A survey. *Sensors* **20**(4), 969 (2020)
2. Singh, A., Sengupta, S., Lakshminarayanan, V.: Explainable deep learning models in medical image analysis. *J. Imaging* **6**(6), 52 (2020)
3. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R.K., Kumar, P.: Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers.* **25**(3), 1315–1360 (2021)
4. Tomašić, I., Trobec, R.: Electrocardiographic systems with reduced numbers of leads-synthesis of the 12-lead eeg. *IEEE Rev. Biomed. Eng.* **7**, 126–142 (2013)
5. Miranda, D.F., Lobo, A.S., Walsh, B., Sandoval, Y., Smith, S.W.: New insights into the use of the 12-lead electrocardiogram for diagnosing acute myocardial infarction in the emergency department. *Can. J. Cardiol.* **34**(2), 132–145 (2018)
6. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016)
7. Chen, Z., Yeo, C.K., Lee, B.S., Lau, C.T.: Autoencoder-based network anomaly detection. In: 2018 Wireless telecommunications symposium (WTS), pp. 1–5. IEEE (2018)
8. Ng, A., et al.: Sparse autoencoder. In: CS294A Lecture Notes, vol. 72, pp. 1–19 (2011)
9. An, J., Cho, S.: Variational autoencoder based anomaly detection using reconstruction probability. In: Special Lecture on IE, vol. 2, no. 1, pp. 1–18 (2015)
10. Rifai, S., Mesnil, G., Vincent, P., Muller, X., Bengio, Y., Dauphin, Y., Glorot, X.: Higher order contractive auto-encoder. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 645–660. Springer (2011)
11. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., Bottou, L.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**(12) (2010)
12. Kim, J.-C., Chung, K.: Multi-modal stacked denoising autoencoder for handling missing data in healthcare big data. *IEEE Access* **8**, 104933–104943 (2020)
13. Ribeiro, M.T., Singh, S., Guestrin, C.: “Why should I trust you?” explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016)
14. Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774. Curran Associates, Inc. (2017)
15. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: High-precision model-agnostic explanations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
16. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K.-R.: Layer-wise relevance propagation: an overview. In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pp. 193–209 (2019)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
18. Maweu, B.M., Dakshit, S., Shamsuddin, R., Prabhakaran, B.: Cefes: a cnn explainable framework for eeg signals. *Artif. Intell. Med.* **115**, 102059 (2021)
19. Jo, Y.-Y., Cho, Y., Lee, S.Y., Kwon, J., Kim, K.-H., Jeon, K.-H., Cho, S., Park, J., Oh, B.-H.: Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *Int. J. Cardiol.* **328**, 104–110 (2021)
20. Ganeshkumar, M., Ravi, V., Sowmya, V., Gopalakrishnan, E.A., Soman, K.P.: Explainable deep learning-based approach for multilabel classification of electrocardiogram. *IEEE Trans. Eng. Manag.* (2021)
21. Raza, A., Tran, K.P., Koehl, L., Li, S.: Designing eeg monitoring healthcare system with federated transfer learning and explainable ai. *Knowl. Based Syst.* **236**, 107763 (2022)

22. Oh, M., Zhang, L.: Deepmicro: deep representation learning for disease prediction based on microbiome data. *Sci. Rep.* **10**(1), 6026 (2020)
23. McInnes, L., Healy, J., Melville, J.: Umap: uniform manifold approximation and projection for dimension reduction (2018). [arXiv:1802.03426](https://arxiv.org/abs/1802.03426)
24. Explainable ECG anomaly detection playground. <https://repo.ijs.si/hribarr/ecg-anomaly-detection>. Accessed 10 May 2023
25. Explainable ECG anomaly detection webapp. <https://repo.ijs.si/cs/insectt-webapp>. Accessed 10 May 2023

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

