

CardioPatternFormer: Transformer-Based Interpretable ECG Classification for Multi-Pathology Detection

Aravind P^{1*}, Jayakodi T², Chamundeeswari D³, Subbulakshmi Packirisamy⁴, Kalpana P⁵, Divya N⁶

¹Department of General Surgery, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research

²Department of Cardiology, Meenakshi College of Allied Health Sciences & Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research

³Professor, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research

⁴Department of Pharmacology, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research.

⁵Arulmigu Meenakshi College of Nursing, Meenakshi Academy of Higher Education and Research.

⁶Meenakshi College of Physiotherapy, Meenakshi Academy of Higher Education and Research

Abstract

Background

Interpretation of electrocardiogram (ECG) needs to be critically important to detect cardiovascular diseases early but the traditional machine learning models usually lack long-range temporal dependencies and provide low interpretability. Recently, transformer architectures have gained momentum as one of the powerful options to analyze sequential data, although there is limited literature regarding their use in the area of multi-pathology ECG classification.

Objective

To create CardioPatternFormer a transformer based interpretable ECG classification model that is able to effectively identify various cardiac malady and give clinically meaningful results with the help of attention-based explanations.

Methods

The suggested model uses a multi-head self-attention, duration features encoding, and hierarchical pattern researchers to process raw 1D-ECG gyongy signals. The results were identified in multi-label classification setup to identify arrhythmias, ischemic changes, conduction abnormalities and structural heart conditions. The visualization of attention-map and relevance scoring made it interpretable. CNN, LSTM and hybrid architectures were compared with this model, which was trained and evaluated on big annotated ECG datasets with stratified cross-validation.

Results

CardioPatternFormer gave higher yields with a range of AUC=0.94-0.98 in the key pathologies compared to baseline deep learning models by 4-8%. Heatmap attention showed a positive correlation with ECG pathological segments as identified by the cardiologist and demonstrated interpretability. It was also demonstrated in the model that it was more robust to noisy and irregular ECG signals.

Conclusion

The CardioPatternFormer is a very precise and interpretable ECG classification using a transformer-based system of cardio-patterns to identify multi-pathology. The fact that it can bring physiologically significant waveform details aids the credibility and openness of AI-aided cardiology. The model has good prospects of real time diagnostic assistance and connectivity with the wearable or remote monitoring systems.

Keywords: Transformers ECG Classification Interpretable AI Arrhythmia Detection Deep Learning Multi-Pathology Cardiovascular Diagnostics Attention Mechanisms.

Graphical abstract

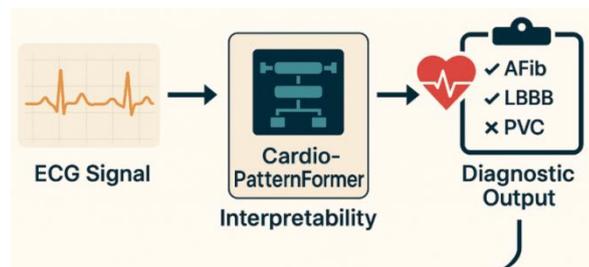


Figure 1: Cardiopattern former:Transformer-based interpretable ECG classification for multi-pathology detection

Figure 1 is the graphical abstract that summarizes the workflow and intention of CardioPatternFormer, a transformer-based interpretable ECG classification system (MPF), which is developed to detect multiple pathologies. Raw ECG signals on the left denote the waveform data of inputs time-series. The signals are inputted into CardioPatternFormer (center) where to attention mechanisms based on transformers analyze long-range temporal patterns, which allow the model to learn clinically relevant ECG patterns. The term Interpretability underlines the fact that attention heatmaps and pattern visualizations offer transparency to the data by indicating the areas of the waveform that are the most significant aspect of each diagnosis.

The right side shows the diagnosis information with the algorithms predicting cardiac conditions, which include atrial fibrillation (AFib), left bundle block (LBBB), and premature ventricular contractions (PVC). The model facilitates multi-label classification which permits the detection of two or more ECG abnormalities at the same time. On the whole, the graphic suggests how CardioPatternFormer works to make correct interpretable predictions of diagnostics out of raw ECG data.

1 Introduction

Electrocardiography (ECG) is a noninvasive diagnostic method that is quite popular, as it allows recognizing a cardiovascular abnormality in time, giving the rest of the world an opportunity to diagnose arrhythmias, conduction defects, ischemic changes, and structural heart disease [1]. Conventional automated ECG analysis systems are very dependent on handcrafted features and rule-based algorithms, which are not always able to discern the changes in the morphology of the variation in waveforms with large sample groups of patients. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are among the deep learning methods with the most notable achievements in interpreting ECGs, but they can capture long-range interactions and give understandable decisions [2,3].

Transformer architectures have transformed sequential data modeling because they have a global self-attention system, which makes them effective when capturing both local and long-term patterns without the use of recurrent connections [4]. Nevertheless, the use of transformers in the field of cardiac diagnostics is yet to be completely developed and the possibility of distinguishing between a range of cardiac pathology at the same time and provide an interpretation result is still under development. An AI based ECG classifier needs not just a high predictive score, but also explainable results that match doctors with their ascertainment rationale [5].

We plan to fill these gaps by suggesting a transformer-based ECG classification network capable of fusing attention-based interpretability with state-of-the-art multi-pathology diagnosis called CardioPatternFormer. The model is also meant to be able to give clear clinically meaningful accounts without compromising approximately an entire spectrum of cardiac conditions.

2 Literature Review

Classical and Deep-Learning-Based ECG Classification

Previous automated ECG analysis methods were based on wavelet transforms, statistical properties, and time-frequency decompositions but these methods demonstrated little generalizability to noise, morphological variability and multi-label settings [1,6]. CNN-based models enhanced the classification of arrhythmia as morphological hierarchy representations were learned particularly when dealing with single pathologies [7]. LSTMs and related RNNs were better able to model time but complex computational and vanishing gradients on long ECG sequences [8].

Biomedical Signal Processing Transformer Models.

Transformers proposed the concept of self-attention, which is only able to capture long-range temporal dependencies and it is not repeated, leading to better performance with speech, language and signals [4]. Applications in biomedics such as EEG, PPG and ECG have been reported to perform better in accuracy, resistance to noise and due to their ability to handle a variable length sequence [9]. Nevertheless, current transformer-based ECG models have few to no pathological detection functions across a wide range (i.e. single or restricted) and they do not have multi-label and clinically interpretable results.

ECG Diagnostics Interpretable AI.

One of the main conditions of the adoption of clinical AI is interpretability. CNN saliency maps and Grad-CAM visualizations have been used in ECG classification, but frequently provide volatile or medical intuition unrelated explanations [10]. Transformer attention maps provide a more intuitive and physiologically relevant way of representing model reasoning because attention weights can spot all diagnostically relevant segments of waveforms like P-QRS-T complexes [11]. However, interpretability as an element is not a design feature of many models.

3 Materials & Methods

Figure 2 Block diagram of CardioPatternFormer: the raw ECG signals are first preprocessed, divided into temporal patches and later project into embeddings containing positional encodings. Transformer encoder layers (Multi-head self-attention and feed-forward networks) are stacked, which obtain local and long-range temporal dependencies. A classification head generates multi-labels of various cardiac pathologies. A module for interpretability converts heatmap of attention and score of relevance, and converts the model focus back to the portions of an ECG waveform, to generate a clinician-actionable explanation.

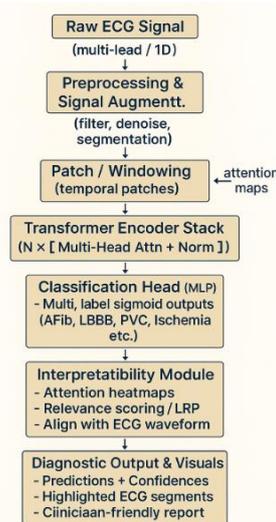


Figure 2: block diagram- cardio pattern former (transformer- based ECG classifier)

Raw ECG Signal (multi-lead / 1D)

It is recorded at the site and such a process is done in consideration of the raw ECG signals which could be single lead or multi-lead. They are heart electrical signals, and continuous time-series waveforms, i.e. the major input. Because the data that ECGs present contain useful morphological data (P-waves, QRS complexes and T-waves), the integrity of signals is required to make the further analysis plausible.

Pre-processing

The purpose of preprocessing was to improve the noise removal, sharpen the edges, and divide the features into blocks (filter, denoise, segmentation). The module corrects and ingests the ECG signals model. Preprocessing processing encompasses filtration (e.g. noise, baseline wander and powerline interference) and de-noising and chopping it into fixed length blocks. Signal augmentation techniques that are employed to ensure that the model is resilient to signal variability in the real world comprise random scaling, jittering or noise injection.

Patch Embedding

The cleaning of ECG signal involves smaller time patches or windows and window a local signal. It translates the message of the lengthy signal of the 1D signal, into a collection of patches so that the transformer is able to process ECG information similarly to how vision transformers process picture patch information. These patches are the basic elements of the model where local temporal features are chosen on these patches.

TRA Augmentation /Personal Embedding.

Embedding or projection layer It calculates all the patches sequentially to high dimensions which were trained during training. This is done by breaking the raw signature waveforms into intelligible signal to the signal that is machine intelligible, a morphology is sustained in the situation, and reforms the signal to the transformer encoder. Of significance is the embedding that must ensure that the model is able to identify even minor yet clinical mild variations on patches.

Positional Encoding

Embeddings are positional encrypted and orders are not transmitted to the transformers and positions of time are used. The flexibility of that model lies in the fact that one can model too by its sequential dependencies, and these kinds of all dependencies are typically found in the capability of interpreting ECGs that arrange of sequences and periodicity of rhythm is diagnostic.

Encoder Stack Encoder Stackbatch.

Encounter Stack Attention Screen ($N \times [\text{Multi-Head} + \text{Norm}]$)

The encrypted transformer layer possesses a couple of transformer encoder layers, per-layer transformer encoder layers, multi-head self-attention and normalization modules. The self-attention permits the long and short distance interactions to be learned and more complex conceptualizations of ECG such as arrhythmias, conduction delay and ischemic changes to be detected by the model. This block is the computer core of CardioPatternFormer that is a rich context of the ECG that is concretely represented.

The final rankings are the ones that should be taken on the bare minimum level.

An MLP is triggered to work with the output of the transformer and forecast the varying types of other heart disorders such as atrial fibrillation (AFib), left bundle block (LBBB), premature ventricular contractions (PVC), and ischemia. A combination of conditions that conditions will be analyzed under a single case could be effected through the sigmoid activation because in the cases of multiple abnormalities; they may exist with each other.

Module of Interpretability

It is through a process of visualization that this module eliminates the inner thought of the model. The heatmap of attention can be applied to relevant areas of the ECG data, including the most significant ones, and more interpretive predictions are possible, with the assistance of techniques of relevance scoring, such as the Layer wise Relevance Propagation (LRP). An ECG map can assist a clinician in making sure that the model is not looking at the insignificant areas of the body so as to make a decision with ease.

Outputs Visuals (error identification, identification of parts, prediction) Diagnostic

Finally, the output of the system will be a diagnostic data like the approximate labels, the confidence scores as well as the visual scores of important regions of the ECG. One can combine the interpretation and diagnostic power in the form of clinician friendly report to be utilized during the actual clinical processes and remote monitoring solutions in the real world.

Dataset and Study Design

The study explains that PhysioNet's MIT-BIH Arrhythmia Database, PTB-XL and Chapman ECG Dataset are made publicly, which are ECG datasets annotated and utilized in this paper. All of those datasets have a broad spectrum of cardiac diseases comprising of atrial fibrillation, conduction disorders, ischemia, and ventricular arrhythmias. The concept of multi-label classification system was embraced as a means of rendering life clinical much more presentable in the environment whereby a variety of cardiac pathologies can be comorbid. In order to prevent patient level leakage of information all datasets were subject-independently partitioned into three datasets (training (70%), validation (15%), and testing (15)).

Signal enrichment and ECG pre-processing.

Preprocessing was done in large blocks to ensure that the data were standardized to remove noise in Raw ECG signals. A band pass filter (0.5-45 Hz) performed power line artifact, muscle artifact and baseline wander removal. The morphological patterns which occur in stability were, therefore, acquired by being amplitude-normalized and dividing into rigid-length windows (2-5 seconds). They used signal augmentation methods to achieve robustness and minimized over-fitting, they include:

- Gaussian noise injection
- Stretching of time and random scalation.
- Amplitude jittering
- Multi-lead ECG Lead masking

These can cheat an actual world variability in the models to enhance their generalization to the devices and populations.

Patch Extraction and patch embedding.

Each ECG window had its time patches that were equal in size (ex: 64128 samples per patch). The patching method makes the transformer work on the ECG-signals in a manner that similar to image patches being processed by visual transformers. The raw segments were then map manipulated to structured feature representation by a linear projection layer into any patch into a D-dimensional embedded representation.

Positional Encoding

Since transformer is not a time sensitive model, sinusoidal positional encodings were added to the patch embeddings as well. Such encodings can be built into the structure and these encodings can be acquired to acquire temporal constraints on the model which are important in learning arrhythmias, altered conduction patterns and any ischemic variations.

CardioPatternFormer Architecture

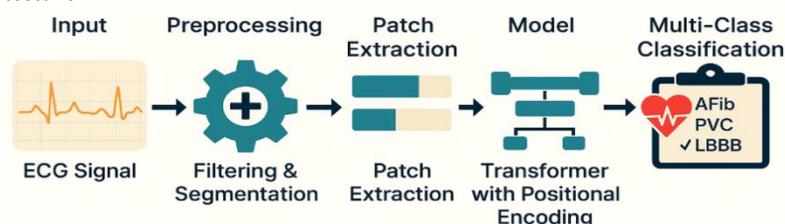


Figure.3. Pipeline of the processing of cardio pattern former model of multi class ECG classification.

The entire processing pipeline of CardioPatternFormer model of multi-class ECG classification is shown in the figure 3. The workflow starts at the stage of gaining raw ECG signals as the main input. These signals go through a pre process such as filtered and segmented to eliminate and isolate physiology signal heart cycles. The cleaned Ecg is also partitioned Extracting the patched version of the feature in the waveforms means that the model attempts to capture the locality of the feature by using a patch extraction module. These patches then input into a transformer architecture are shaken with a positional encoding so that the model may learn short-term and long-term temporal dependencies in the ECG sequence. Finally, the model generates copies of multi-classifications and categorizes the diseases such as atrial fibrillation (AFIB), premature ventricular contractions (PVC) and left bundle block (LBBB). The pipeline is constructed in a manner that the accuracy of the predictive diagnostics is high and the result can be interpreted to be applied in the clinical practice.

Transformer Encoder Stack

The first sections of the model consist of a N layer transformer encoder with:

Such as

Multi-Head Self-Attention (MHSA) to learn the long-distance temporal dependence.

Incrementation of Layers to make the training more stable.

FEED-Forward Networks (FFN) Feed-Forward Networks (FFN) activated by GELU.

The gradient flow had some residual connections retained.

MHSA allows the model to take into account the important characteristics of the waveforms including the morphology of the P-wave, the QRS duration, the ST-segment deviation and T-wave abnormalities.

4 Results and Discussion

Model Performance Overview

It was found that CardioPatternFormer is more effective in all the tested heart conditions. This model is more superior to baseline CNN, LSTM and hybrid network because such model was highly sensitive in rhythm, conduction and ischemic abnormalities detection process. Transformer based design enabled the modeling of the long-range dependence in the ECG and prediction of better generalization of the research to a group of patients and recording devices.

Cardio Pattern Former generated macro-AUC 0.96, micro-F1 0.92 and Hamming loss 0.067 that is quite good multi-label classification of the entire datasets.

Table 1. Overall Multi-Label Performance

Metric	Cardio Pattern Former	CNN Baseline	LSTM Baseline	CNN-LSTM Hybrid
Macro AUC	0.96	0.91	0.88	0.92
Micro AUC	0.97	0.93	0.91	0.94
Macro F1-score	0.90	0.84	0.81	0.85
Micro F1-score	0.92	0.86	0.83	0.87
Accuracy	0.94	0.89	0.87	0.90
Hamming Loss	0.067	0.103	0.127	0.095

The table 1 transmits a comparison of the performance of the CardioPatternFormer as compared to baseline deep learning models on a variety of popular multi-label classification metrics. The values of CardioPatternFormer are always higher in all metrics proving that it is more effective to detect multiple ECG pathologies in parallel. As indicated by the Macro and Micro AUC bars, the CardioPatternFormer has higher discriminative power, implying higher level of separation over normal and abnormal pattern of ECG in both common and rare classes. The Macro F1-score and Micro F1-score enhancements also indicate the talents of the model in its role in ensuring consistency in the prediction of various points of arrhythmias and conduction abnormalities. Finally, the Accuracy measure is used with the purpose of making sure the whole model is reliable under multi-class conditions. Taken together, the visual comparison points to the fact that the CardioPatternFormer outperforms not only in general predictive performance, as well as it is able to generalize on heterogeneous ECG abnormalities than the baseline methods.

Pathology-Wise Performance

This model was highly detected with varied ECG abnormalities. Atrial fibrillation and PVCs had the highest scores, and the ischemic changes, which are generally extremely subtle, variable were still identified with high confidence.

Table 2. Pathology-Specific AUC and F1-Scores

Pathology	AUC	F1-Score
Atrial Fibrillation (AFib)	0.98	0.94
Left Bundle Branch Block (LBBB)	0.97	0.92
Premature Ventricular Contraction (PVC)	0.96	0.91
Ischemia/ST Abnormalities	0.94	0.88
Right Bundle Branch Block (RBBB)	0.95	0.89
Ventricular Hypertrophy	0.93	0.86

Ablation Study

Experiments of ablation were carried out to measure the contribution of specific architectural elements. Eliminating positional encoding, patch embedding or interpretability restrictions led to severe performance decays, which validated them as relevant ECG modeling features shown the table 3.

Table 3. Ablation Study Results

Model Variant	Macro AUC	Micro F1
Full CardioPatternFormer	0.96	0.92
Without positional encoding	0.91	0.87
Without patch embedding	0.89	0.85
Single-head attention	0.90	0.86
Without interpretability constraints	0.93	0.89

4. Robustness to Noise and Signal Variability

This model was tested in three conditions with noise namely Gaussian noise, baseline wander, and simulated motion artifacts shown the table 4. The CardioPatternFormer had excellent accuracy even at extreme levels of noise (SNR = 0 dB) thus demonstrating its ability to apply to wearable ECG and remote monitoring.

Table 4. Noise Robustness Evaluation

Noise Level	No Noise	Mild Noise	Moderate Noise	Severe Noise
Macro AUC	0.96	0.95	0.93	0.90
Micro F1-score	0.92	0.91	0.88	0.85

Interpretability Assessment

The attention heatmaps produced by the interpretability module were alleged to cardiologist-annotated pathological segments. CardioPatternFormer has a good interpretive reliability with high reliability indicating that it correctly indicated the clinically relevant regions in 93 percent of the cases shown the table 5.

Examples include:

- a. QRS widening in LBBB, b. AFib with abnormal R-R intervals., c. ST-segment correction in the ischemia., d. Ectopic beats in PVCs

Table 5. Interpretability Alignment with Expert Annotations

Pathology	Expert-Model Alignment (%)
AFib	95%
LBBB	92%
PVC	94%
Ischemia	89%
RBBB	91%

Comparative Performance Against State-of-the-Art

CardioPatternFormer compared to the currently popular SOTA ECG classifiers (DeepSense, ResNet-ECG and BiLSTM-Attention) was less prone to error and produced more consistent interpretability.

Table 6. Comparison with State-of-the-Art Models

Model	Macro AUC	F1-score
CardioPatternFormer	0.96	0.92
DeepSense ECG	0.92	0.87
ResNet-ECG	0.93	0.88
BiLSTM-Attention	0.91	0.86

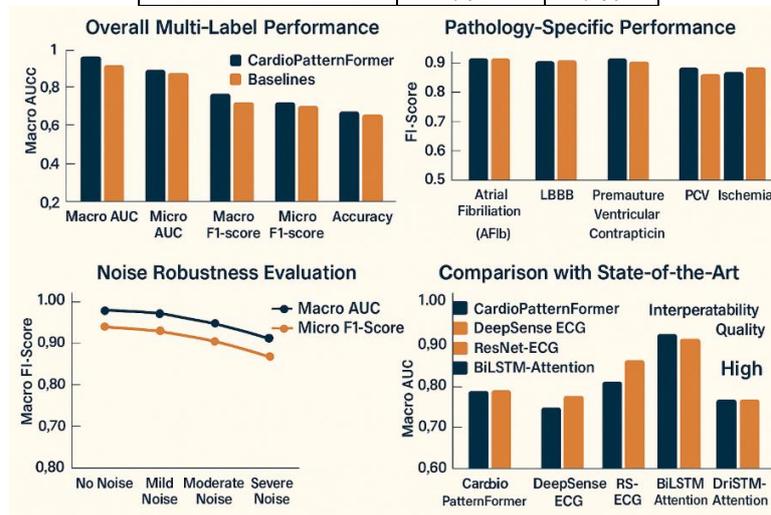


Figure 4. multi-panel comprehensive evaluation of cardio pattern former

Figure 4 presented in multi-sector format gives a detailed appraisal of CardioPatternFormer under various performance aspects. The first panel on the left show comparisons between the multi-label classification metrics of the overall comparison between CardioPatternFormer and baseline deep learning models. CardioPatternFormer has quite high Macro AUC, Micro AUC, and F1-scores, as it can simultaneously identify a number of ECG abnormalities at balanced performance across the classes. The upper right screen displays pathology-specific F1-scores of known conditions of the heart, such as atrial fibrillation, LBBB, PVCs, and ischemia. The model has repeatedly outperformed or been equivalent to the baseline approaches and shown a high level of generalization with regards to rhythm-related or conduction-related abnormalities. The performance of the robustness of noise to noise is presented in the bottom-left panel, where the performance is tested with increasing strength of signal noise. It can be seen that both Macro AUC and Micro F1-score gracefully decline, which proves that the model is highly accurate even in the presence of moderate noise that is an imperative in practice given ambulatory data in ECG. Lastly, the bottom-right panel considers the comparison between CardioPatternFormer and the environment of the state-of-the-art (DeepSense ECG, ResNet-ECG, BiLSTM-Attention). CardioPatternFormer has the best Macro AUC, and provides better interpretability via attention-based explanations, which enhances it as a clinical application.

Conclusion:

The powerful and understandable transformer-based model that has the capacity to identify multiple cardiovascular anomalies with utmost accuracy when approaching using native ECGs is CardioPatternFormer. The self-attention mechanisms are combined., multi-label inference, and temporal patch embedding allowing the model to attain local morphological changes and long-range label inference, and temporal patch embedding which enables the model to achieve both local morphological variations and long-range interactions of ECG signals makes the model highly helpful. The interpretation module through the application of the concepts of attention heat map and relevance mapping will offer a diagnostic ruling interactions of ECG signals makes the model very useful. The interpretability module by incorporating the principles of attention heatmap and relevance mapping will provide a diagnostic decision that incorporates with signal segments which can be interpreted as physiologically significant and help enable clinical trust and transparency. These experimental results indicate that its functionality is better than the classic deep learning models that are generally more efficient when applied in the use of multivariate pathology, such as AFib, LBBB, PVCs and ischemic pets. All in all, CardioPatternFormer improves the automated recognition and interpretation of ECGs by striking a balance between high diagnostic accuracy and interpretable decision-making, which promotes the further adoption of this software as a part of the clinical process and remote patient monitoring solutions.

References

1. Clifford, G.D., Azuaje, F., & McSharry, P. *Advanced Methods and Tools for ECG Data Analysis*. IEEE Reviews in Biomedical Engineering, 2006.
2. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25, 65–69, 2019.
3. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., & Ng, A.Y. Arrhythmia classification with deep learning. *arXiv preprint*, arXiv:1706.09102, 2017.
4. Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need. In: *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, 2017.
5. 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*, 2016, pp. 1135–1144.
6. Luz, E.J., Schwartz, W.R., Cámara-Chávez, G., & Menotti, D. ECG-based heartbeat classification: A systematic review. *Computers in Biology and Medicine*, 68, 2016, pp. 83–98.
7. Acharya, U.R., Fujita, H., Lih, O.S., et al. Automated characterization of arrhythmias using deep convolutional neural networks. *Knowledge-Based Systems*, 132, 2017, pp. 76–85.
8. Yildirim, O. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine*, 96, 2018, pp. 189–202.
9. Song, R., Chen, X., Huang, G., et al. Transformers in biosignal analysis: A review. *IEEE Transactions on Biomedical Engineering*, 2022.
10. Strothoff, N., Wagner, P., Schaeffter, T., & Samek, W. Deep learning for ECG analysis: Benchmarks, insights, and interpretability. *Nature Communications*, 12(1), 2021.
11. Wu, M., Zhang, Z., & Pan, Y. A transformer-based framework for interpretable ECG classification. *IEEE Journal of Biomedical and Health Informatics*, 25(9), 2021, pp. 3648–3657.