An Optimized GRU Model with Callbacks for Efficient Classification of Cardiovascular Diseases Using ECG Signals

Abdulrahman Hamid Mahmood and Taha Mohammed Hasan
Department of Computer Science, College of Science, University of Diyala, 32001 Baqubah, Diyala,
Iraqscicompms232406@uodiyala.edu.iq, dr.tahamh@uodiyala.edu.iq

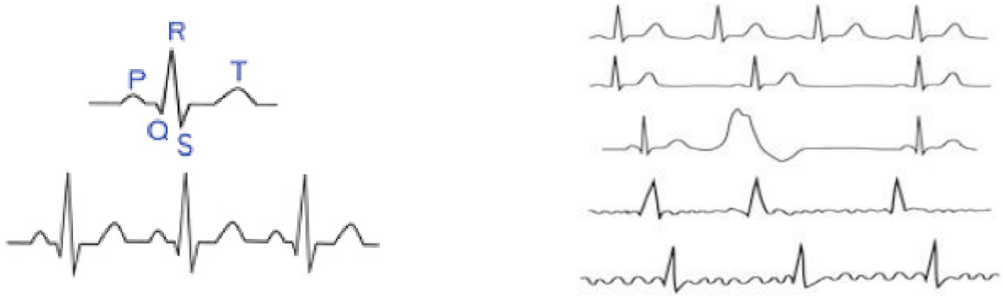
Keywords: Cardiovascular Disease Prediction, Adaptive Callbacks, Optimized GRU, Optimized LSTM, Multi-Layer 1D CNN, Arrhythmia Dataset, ECG Signals.

Abstract: Cardiovascular disease represents an ongoing and significant global health concern, and early, accurate cardiac arrhythmia identification is critical to the prevention of severe cardiovascular complications, underscoring the requirement for reliable prediction systems. In this paper, a deep learning-based system for cardiovascular disease classification using the benchmark MIT-BIH Arrhythmia Dataset is proposed, comparing the performance of Gated Recurrent Units (GRUs) with Long Short-Term Memory (LSTM) and one-dimensional Convolutional Neural Network (1D CNN). The utilized dataset is first balanced by class using resampling and separated into subsets (training and testing), and then the training set is normalized (scaled) without leakage to be fed to the deep learning models. To improve generalization and efficiency in training recurrent models, diverse adaptive callbacks are used, including Early Stopping, Reduce LROn Plateau, and Model Checkpoint. The results depicted that although the optimized LSTM and multi-layer 1D CNN models provided strong prediction ability in learning temporal sequence and capturing spatial features of electrocardiogram (ECG) signals, respectively, the optimized GRU model performed superiorly, achieving above 99% accuracy, precision, recall, and F1-score. These results confirm that the GRU model, when incorporated with carefully structured training methodologies, provides an effective and accurate system for predicting cardiovascular disease, with significant possibilities for consolidation into clinical decision support systems.

1 INTRODUCTION

Cardiovascular diseases refer to a variety of medical conditions that can affect the blood vessels and heart, leading to impaired cardiovascular function and causing deaths worldwide [1]. These conditions comprise a broad range of disorders, involving heart valve disease, heart failure, coronary artery disease, arrhythmias, and so on, which greatly affect a person's lifespan and survival [2]. As reported by the World Heart Federation, the number of deaths worldwide owing to cardiovascular diseases increased from 12.1 million in 1990 to 20.5 million in 2021. Moreover, as reported by the World Health Organization, these diseases have recently caused an estimated 17.9 million deaths annually. Physical inactivity, unhealthy diet, smoking, and excessive alcohol consumption are major risk factors for stroke and heart disease. The identification of cardiovascular risk early and supplying a suitable treatment can substantially reduce deaths. An accurate prediction with early recognition of cardiovascular risk helps avoid harmful conditions of cardiovascular disease and enhances patient health [1].

Electrocardiography is a comfortable, inexpensive, quick, and accessible diagnostic technique that is broadly utilized in medical practice by non-cardiologists and cardiologists. The electrocardiogram (ECG) signals reflect the electrical activity of the human heart, and any alteration in the heart rhythm disturbance (heart waveform) refers to underlying cardiovascular problems. The ECG examination provides a comprehensive view of the anatomical and physiological state of the heart and also provides beneficial diagnostic clues for disorders affecting the entire human body [3]. Figure 1 demonstrates the variance between normal and abnormal rhythm signals, in which "P-wave" indicates the duration for atrial contraction, "QRS-wave" indicates the period of ventricular contraction, and "T-wave" indicates the period of ventricular relaxation [4].



(a) (b)

Figure 1: ECG signal representation: a) Normal, b) Abnormal rhythm signals [4].

In spite of the clinical importance of ECG signals, their analysis represents a significant issue owing to their noise content and variable and complex nature. The conventional diagnosis method of cardiac arrhythmias depends on manual analysis of the ECG signal by medical professionals. Although this method is efficient, it requires extensive human effort, consumes time, and is prone to human error, which makes it unrealistic for large-scale screening. To conquer these restrictions, Artificial Intelligence-based ECG auto-classification systems have been developed as a pivotal research domain, bringing about essential advances in diagnostic accuracy and speed [5], [6].

Recently, deep learning models have received significant attention as encouraging schemes for automated classification or prediction in diverse research domains [7] - [10]. In particular, various deep learning-based prediction research studies demonstrated their ability to classify cardiovascular diseases using massive datasets [11]. However, cardiovascular disease auto-detection using ECG signals is a complicated issue; therefore, deep learning-based systems should achieve outstanding performance in diagnosing such diseases [12]. In this paper, the proposed system aims to evolve and assess three deep learning models for accurately predicting cardiovascular diseases from ECG signals, with a specific focus on comparing the efficiency of GRU, LSTM, and 1D CNN models, and evaluating the role of dataset balancing and its preprocessing, and adaptive callbacks in optimizing models' efficiency and predictive performance. The principal contributions of this suggested deep learning-based system are as follows:

1) Comparing several deep learning models for cardiac arrhythmia classification utilizing the MIT-BIH Arrhythmia Dataset, which highlights their efficiency in acquiring temporal and spatial features from ECG signals.
2) Resampling the utilized ECG dataset to provide balanced class distributions, and normalizing features in the training set using MinMaxScaler to inhibit the leakage of data.
3) Incorporating diverse adaptive Callbacks (Early Stopping, Reduce LROn Plateau, and Model Checkpoint) in the training of the GRU and LSTM pipeline to improve model generalization and decrease training time.
4) Achieving the superior predictive capability using an optimized GRU model, with an accuracy of 99%, exceeding the other implemented models (optimized LSTM and multi-layer 1D CNN), and verifying its efficacy in predicting cardiovascular diseases.
5) The results highlight the potential of the GRU model, combined with adaptive training methodologies, for incorporation into real-life clinical decision support systems that facilitate accurate and early detection of cardiovascular disease.

The remainder of the paper is structured as follows: the "Related Work" section surveys prior and relevant research studies on cardiovascular disease detection using a benchmark arrhythmia dataset. The "Proposed System" section presents the proposed methodology, including dataset description, preprocessing, and optimized classification. The "Experimental Results and Discussion" section exhibits the models' evaluation and results analysis. Eventually, the "Conclusion" provides the core conclusions about the proposed system. The remainder of the paper is structured as follows: the "Related Work" section surveys prior and relevant research studies on cardiovascular disease detection using a benchmark arrhythmia dataset. The "Proposed System" section presents the proposed methodology, including dataset description,

preprocessing, and optimized classification. The "Experimental Results and Discussion" section exhibits the models' evaluation and results analysis. Eventually, the "Conclusion" provides the core conclusions about the proposed system.

2 RELATED WORKS

Recent elevations in deep learning models have substantially strengthened the classification of cardiac arrhythmias based on ECG signals. Diverse systems have evolved to cope with diverse issues like computational efficiency, noise, and class imbalance, all of which contribute to enhancing the performance of classification systems [13]. Temporal modeling represents a critical component of ECG analysis, as cardiac rhythms inherently display sequential dependencies. GRU models supply an effective scheme for capturing long-term dependencies and short-term variations from ECG signals, providing performance comparable to or superior to LSTMs while employing fewer parameters and minimizing the cost of computation. While CNN models have been broadly utilized owing to their capability in capturing spatial patterns from ECG signals, resulting in acceptable performance in classifying cardiac arrhythmias [14]. A comparative review of some of the latest systems of arrhythmia detection is provided in Table 1.

K. Lee et al. [15] studied beat score map-based systems for multi-class arrhythmia ECG signals classification. These ECG signals were first de-noised using the discrete wavelet transformation (DWT), and their Peaks were detected using either ground truth or the Pan-Tompkins method. They were then segmented, transformed into 2D scalograms, converted into beat score vectors, and created peak interval vectors to construct 2D images. Finally, these images were normalized. The lightweight 2D CNNs (VGG16 or MobileNet) were implemented to classify the yielded pre-processed images into multi-class arrhythmia. VGG16-based system achieved superior performance with accuracies of 98.96% for the MIT-BIH dataset and 91% for the SPH dataset. This system has difficulty transferring between datasets. Additionally, the uneven distribution of classes in MIT-BIH significantly impacted the detection of minority rhythms.

M. Akkuş et al. [16] proposed an arrhythmia auto-classification system that incorporated an adapted Spindle Autoencoder with a CNN model. In this system, meaningful and complicated features were first extracted from the ECG signals through the symmetric and deeper hidden layers of the Autoencoder utilized. The CNN then analyzed these learned representations to acquire spatial relations over convolution layers. This system was trained and assessed using a benchmark Arrhythmia dataset named MIT-BIH. This dataset was filtered using a Bandpass to eliminate high-frequency noise and baseline drift, normalized using Z-score to standardize amplitudes, segmented into separate beats, and split into 80% train and 20% test, then fed to the extraction of essential features to enhance system performance. This system achieved high robustness and accuracy (98.78%) across diverse arrhythmia classes, handled noise efficiently, and accelerated inference and reduced storage by combining compression with classification. However, its drawbacks include its complex architecture, sensitivity to class imbalance, and potential overfitting.

T. Anitha et al. [17] presented a hybrid ECG signals-based arrhythmia classification system that combines the CNNs' strengths, RNNs' sequential sensitivity, and intelligent construction of Capsule Networks. In this system, ECG signals captured from the MIT-BIH arrhythmia dataset were preprocessed through several steps, including removing baseline dispersion using DC Drift, scaling signal amplitude, eliminating high-frequency noise and artifacts, converting them to time-frequency representations, split into 60% train and 40% test, and balancing the training set across heartbeat classes. These preprocessed signals were then fed to the CNN-RNN scheme to extract the essential features. Eventually, the deep model of the bidirectional capsule network was implemented to classify cardiac arrhythmia into five classes. This presented system outperformed standalone models (CNN and Capsule Network) with an accuracy of 97.19%. Although this system had better noise handling and reduced overfitting, its complex architecture required a higher computational cost in contrast to standalone models.

Y. Zhang et al. [18] proposed a self-supervised arrhythmia classification system that combined generation and clustering models. In this system, the utilized 1D-time series ECG signals were segmented, normalized, and converted into a 2D matrix utilizing Gramian Angular Summation Field to produce images (size 64×64). Subsequently, these matrices are handled utilizing the Sequential Masked Autoencoder and Gaussian Mixture Clustering to classify cardiac arrhythmia into five classes of the MIT/BIH dataset. The proposed system achieved an accuracy of 97.84%. Although this type of learning excluded reliance on large labeled datasets and extracted essential features across diverse classes, it was computationally heavy and provided poor detection for minority classes (low sensitivity).

A. Gupta et al. [19] presented an LSTM-based system for classifying arrhythmia ECG signals into five classes. These signals were first separated into particular beats, normalized to obtain a consistent amplitude, and fed as sequences to the LSTM. The presented system was trained and approved utilizing the MIT-BIH dataset, and the findings attained were remarkable (accuracy of 98.85%), which far outperformed other state-of-the-art SVM, CNN, or hybrid CNN-LSTM-based systems. Although this system is promising for supporting clinical decisions, improvements in noise handling and interpretability are necessary for applying in real-world systems.

C. K. K. Reddy et al. [20] developed a deep-learning-based system for enhancing ECG signals' cardiac arrhythmia classification performance. The ECG signals from the MIT-BIH dataset were preprocessed to be ready for the system training, including eliminating noise using wavelet de-noising, normalization using Z-score, and balancing the dataset via randomly down-sampling and over-sampling of majority and minority classes, respectively. This system incorporated three layers of 1D CNN, dense blocks, bidirectional LSTM, and two dense layers with a Softmax to achieve accurate classification across five arrhythmia classes. Although the developed system provided high classification accuracy (98.22%), its complexity increased the memory demand and time required for training, and there is a potential for overfitting.

S. Mangaraj et al. [21] proposed a hardware-aware two-branch CNN framework to improve the inherent parallelism of field programmable gate arrays and decrease the complexity of data reuse in classifying ECG signals into five classes. Instead of ReLU activation and fully-connected layers, an unsigned 4-bit integer was utilized to ensure nonnegative intermediary results. The used arrhythmia signals were pre-processed before going deeper into the classifier, including filtering, normalization, peak detection, segmentation, 64×64 image formation, and quantization. This proposed system achieved an accuracy of 97.79% using the MIT-BIH dataset. This system delivered efficient performance with a balance between hardware and accuracy, low latency, and energy efficiency. However, it may require larger inputs beyond memory and handle class imbalance.

R. Dong and L. Xie [22] presented an incorporation of CNN and LSTM models with a cross-attention for classifying arrhythmia ECG signals. The CNN and LSTM models were utilized for extracting local features and sequential dependencies, respectively, and cross-attention was utilized for fusing time and frequency domains of data, which improved classification performance. This hybrid system was trained and assessed using the MIT-BIH dataset for classifying arrhythmia into five classes and the PTB dataset for classifying arrhythmia into two classes (normal and abnormal). The imbalance issue associated with the MIT-BIH dataset was addressed using SMOTE and Tomek links techniques to improve the system's capability to accurately classify minority and majority classes. This system was able to handle imbalanced data efficiently and achieve a high degree of robustness and accuracy (98.97% using the balanced MIT-BIH dataset and 99.69% using the PTB dataset), despite the high complexity and sensitivity of pre-processing.

Some of the previously mentioned relevant systems for cardiovascular disease classification using ECG signals frequently concentrated on transforming 1D ECG signals into 2D images or representations, leveraging CNN models. Although those image-based systems can efficiently identify morphological features, they lack access to long-term rhythm dynamics, which are vital for characterizing arrhythmia. While other relevant systems tried to recover temporal context via incorporating recurrent components (like LSTM or bidirectional units) with hybrid models (like CNNs or Capsule Networks). These frameworks are highly compute-intensive, need complex parameter tuning, and are less appropriate for real-time deployment. Therefore, in this proposed system, lightweight learners are trained on balanced training, leakage-free pre-processing, and stabilization callbacks, bypassing complicated pipelines while maintaining ease of deployment.

3 PROPOSED SYSTEM

The proposed optimized GRU-based system for cardiovascular disease classification focuses on direct temporal learning while preserving the ECG signals in their original one-dimensional form. This system ensures interpretability and computational efficiency through comparing end-to-end learning models (optimized GRU, optimized LSTM, and multi-layer 1D CNN) via a transparent framework, as shown in Figure 2. This framework specifically handles the class imbalance issue by using resampling prior to training models, which mitigates bias for majority classes and enhances the identification of uncommon arrhythmias. Moreover, preprocessing is implemented in a leakage-free form, with expansion limited to the training set of data, guaranteeing unbiased assessment of new testing samples. Training convergence and stability are also enhanced through callbacks to inhibit overfitting and maintain persistent performance across runs.

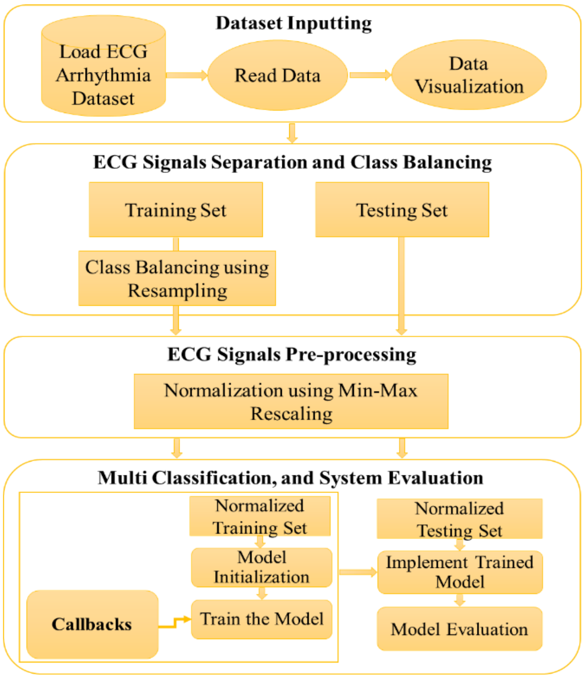


Figure 2: Proposed system framework.

3.1 Arrhythmia Dataset Utilized

This proposed system utilizes an accessible dataset produced by the Massachusetts Institute of Technologies and hosted on "PhysioNet", specifically the Arrhythmia MIT-BIH dataset [23]. This dataset involved 47 subjects' annotated 2-channel ECG recordings gathered for 48 patients (23 females and 25 males). Each of those ECG recordings is roughly half an hour long, amplitude of 10 mV, features an eleven-bit resolution, and a 360 Hz sampling rate.

This dataset is broadly utilized in cardiac arrhythmia classification and prediction. Its standardized and varied detailed nature has contributed to substantial progress in health-related signal processing and provided a significant basis to develop our proposed arrhythmia classification system.

Table 1: Comparison of some of the latest arrhythmia detection systems using the MIT-BIH dataset.

Authors, Ref., Year	MIT-BIH Dataset Preprocessing	Dataset Balancing Techniques	Feature Extraction and Classification Models	Findings
K. Lee et al. [15], 2024	DWT, Pan-Tompkins method, Segmentation, Scalograms, 2D images formation, and Normalization	-	Pre trained 2D CNN (VGG16)	Accuracy= 98%, F1-score= 97%
M. Akkuş et al. [16], 2025	Bandpass filter, Z-score Normalization, and Segmentation	-	Adapted Spindle Autoencoder with CNN	Accuracy= 98%, Precision, Recall, and F1-score= 96%
T. Anitha et al. [17], 2025	DC Drift Elimination, Normalization, Low-Pass Filtering, Artifact Elimination, and Spectrogram Analysis	Data Augmentation	Ensemble CNN-RNN, and BiDirectional Capsule Network	Accuracy= 97%, Precision=95%, Recall=97%, and F1-score= 96%
Y. Zhang et al. [18], 2025	Segmentation, Normalization, and Transformation using Gramian Angular Field	-	Sequential Masked Autoencoder, and Clustering	Accuracy= 97%, Recall= 65%, and Specificity= 94%
A. Gupta et al. [19], 2025	Segmentation and Normalization	-	LSTM	Accuracy= 98%
C. K. K. Reddy et al. [20], 2025	Wavelet Denoising, and Normalization	Down-Sampling and Over-Sampling	1D CNN, Dense Blocks, Bidirectional LSTM, and Two Dense Layers	Accuracy= 98%; Precision, Recall, and F1-score= 97%
S. Mangaraj et al. [21], 2025	Band-pass filtering, Scaling, Pan–Tompkins method, Segmentation, 2D image conversion, and Affine Quantization	-	Two-branch CNN	Accuracy= 97%
R. Dong and L. Xie [22], 2026	Segmentation, Down-sampling, and Normalization	SMOTE and Tomek links	CNN and LSTM models with a cross-attention	Accuracy= 98%, Precision= 94%, Recall= 93%, and F1-score= 93%

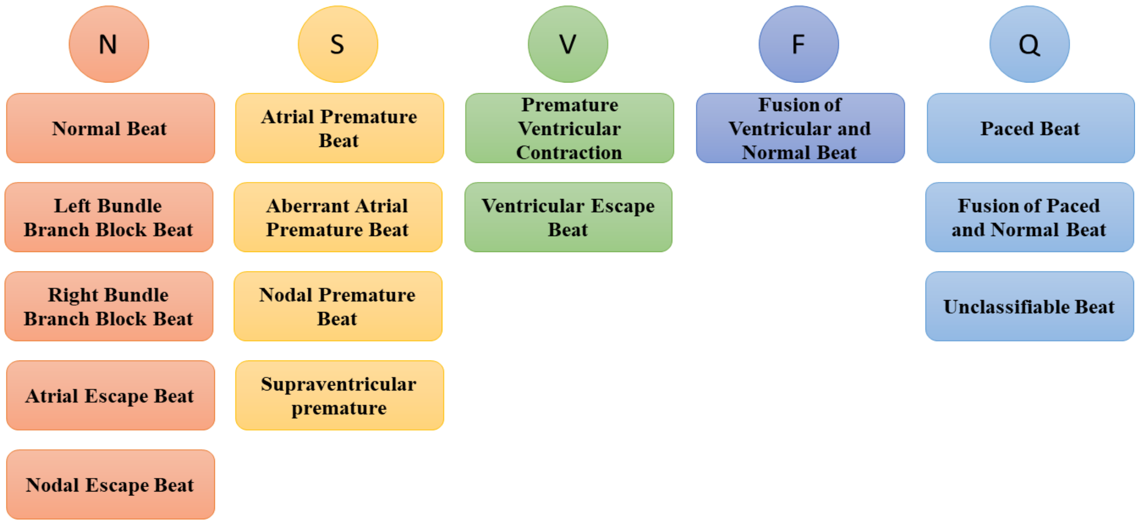


Figure 3: Five classes (which included 15 beat kinds) of ECG rhythms of the MIT-BIH dataset.

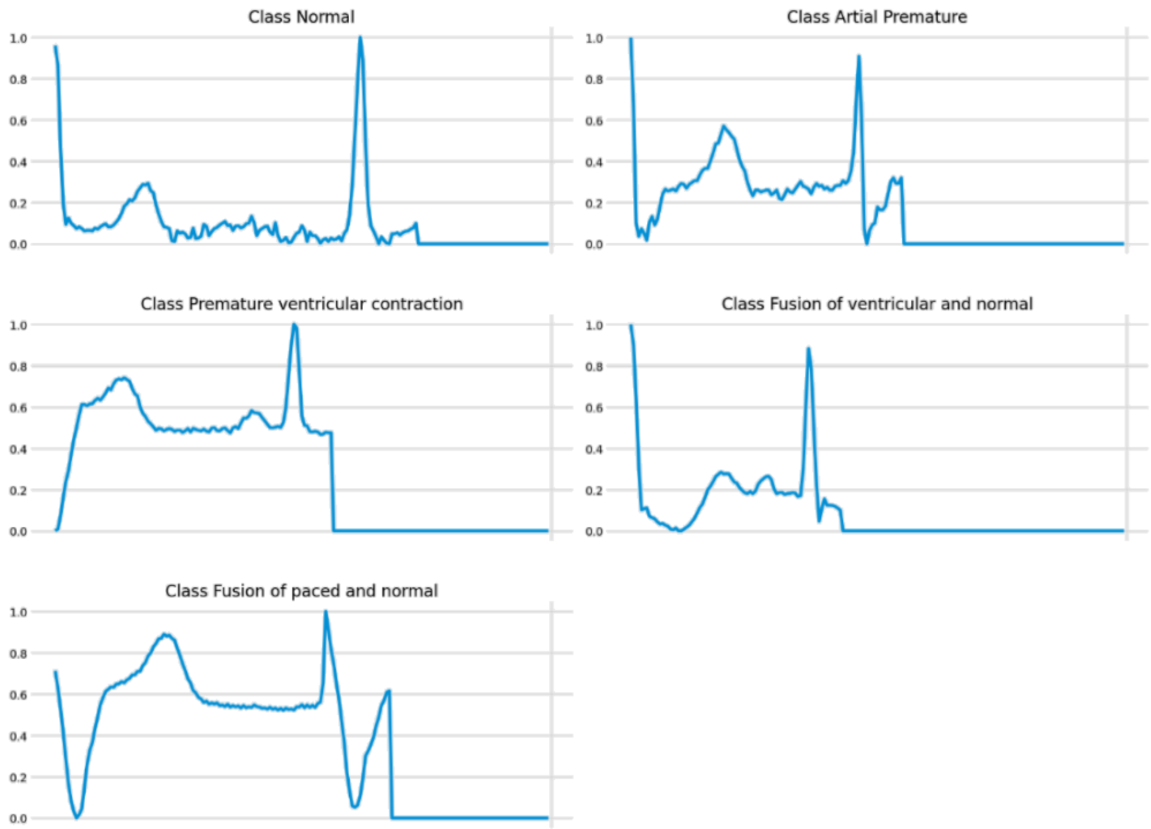


Figure 4: ECG Samples of MIT-BIH dataset.

Several cardiologists have explained the heart rhythms included in this dataset, leading to a classification of 15 normal and abnormal rhythm types, which were in turn categorized into 5 classes, as depicted in Figure 3, in which Q, F, V, S, and N indicate unknown, fusion, ventricular ectopic, supraventricular ectopic, and non-ectopic cardiac beats, respectively. Figure 4 depicts five ECG signal samples for the MIT-BIH dataset classes.

3.2 ECG Signals Separation and Class Balancing

The utilization of the MIT-BIH dataset and its diversity enables us to handle fundamental issues in ECG signals by implementing and validating pre-processing techniques with class balancing, which are essential to improve signals' quality, thus allowing the proposed system to efficiently classify essential features extracted from ECG signals.

The dataset utilized encompasses 90,587 normal and 18,857 abnormal heartbeats, for a total of 109,444 samples. To provide a robust assessment, dataset was separated into 80% training (87,553 samples), and 20% testing (21,891 samples). Figure 5 depicts the count of training and testing abnormal samples.

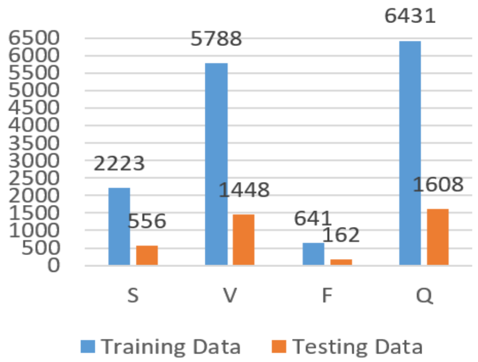


Figure 5: Distribution of abnormal training and testing data.

In the dataset used, we noticed that some kinds of rhythms appeared less frequently than others; abnormal rhythms appeared in less than 5% of the data, while normal rhythms made up to 80%. When the proposed deep learning models are trained on this dataset without processing, they tend to predict the majority class most of the time, since they can provide superior accuracy by guessing “Normal” for all occasions. However, this is not useful for classifying severe and rare arrhythmias. Accordingly, in this proposed system, the training set was first balanced using resampling, where all classes in the training set had the same count of samples; therefore, the system is not biased towards the class with the largest count of samples.

Resampling can make all classes have the same count of samples (20,000). This is done by over-sampling minority classes (repeating samples) and under-sampling the majority class (randomly eliminating samples), resulting in a new training set of 100,000. This equal representation enables deep learning models to learn all kinds of cardiovascular patterns fairly and enhances their capability in accurately detecting rare heart conditions.

3.3 ECG Signals Pre-Processing

ECG signals from the MIT-BIH dataset exhibited variable amplitude extents over electrodes and patients. Without prior normalization, deep learning models can focus on amplitude magnitude instead of the waveform. Therefore, after training set blanching, normalization utilizing Min-Max Scaling is implemented on the training and testing sets independently for rescaling each (numeric) feature f to a determined extent (between zero and one), as given in the following equation:

$$f_{Scaled} = \frac{f - f_{minimum}}{f_{maximum} - f_{minimum}}\,. \qquad (1)$$

This pre-processing step ensures that the input features receive comparable scaling, which is vital for deep learning models, where high-value features can significantly influence the learning process. Additionally, it works on speeding up training convergence and improving generalization.

3.4 Proposed Optimized GRU Model

In this system, the optimized GRU model is proposed for classifying five cardiovascular classes. It is constructed utilizing "Tensor Flow’s Keras API". This model involves several layers, explained as follows:

- A 128-unit GRU as the 1st layer is firstly utilized for capturing temporal dependencies in rhythm ECG signals. Every cell in the GRU contains gating mechanisms to determine which information from former time steps should be retained or forgotten, making (long-term) pattern learning efficient without the issue of the vanishing gradient that plagues traditional recurrent neural networks. The output of the 1st layer represents the whole hidden states' sequence (instead of the final one), allowing the subsequent layer of GRU to process the entire temporal dynamics.
- A dropout is added after the 1st GRU to randomly abandon “20%” of the neurons during training, inhibiting the model from remembering the training data, and helping decrease overfitting.
- The 64-unit GRU is then utilized as the 2nd layer to process the sequence produced by the 1st GRU and output a single vector of summarized temporal information.

Table 2: Explanation of the proposed GRU model hyper-parameters.

Processes	Hyper-Parameters	Setting or Value	Explanation
Compilation	Optimizer	Adam	Adaptive Learning Rate optimizer
	Loss function	Sparse categorical cross entropy	Appropriate for integer-encoded multi-classes
Training	Epochs	20	Highest training iterations
	Batch size	32	Samples per gradient update
Callbacks – Early Stopping	Monitor	Validation loss	Metric utilized for trigger early stopping
	Patience	6	Stops after Six epochs without improvement
Callbacks – Reduce LROn Plateau	Monitor	Validation loss	Reduces Learning Rate when validation loss plateaus
	Factor	0.2	Learning rate reduced by this factor
	Patience	2	Waits Two epochs before Learning Rate reduction
	Min learning rate	0.000001	Minimum Learning Rate
Callbacks – Model Checkpoint	Monitor	Validation loss	Metric for superior model selection

- A dropout is added after the 2nd GRU, serving as a regularizer.
- A Dense (of 5 output neurons) with a Softmax function as a final layer to perform arrhythmia multi-classification.

After model construction, a prevalent optimizer, called "Adam," is utilized to compile the model by adjusting the learning rate separately for every weight throughout the training period, along with a sparse class cross-entropy loss (which is suitable since the targeted labels are integer indices (from 0 to 4) instead of one-hot vectors). Furthermore, this optimizer works on tracking an accuracy metric to assess how well the model's predictions match the correct class labels.

The proposed GRU model is optimized using three Keras callbacks to enhance the training process's efficiency and stability:

- Early Stopping. It is utilized for monitoring the validation loss and terminating the training process when it has not improved for six epochs. This prohibits the overfitting issue via stopping training once the model no longer generalizes well to new data.
- Reduce LROn Plateau. It works on dynamically adjusting the learning rate. If the validation loss ceases improving for two epochs, then it will reduce it to "20%" of its value by multiplying the learning rate by "0.2". Reducing this rate assists the optimizer in tuning local minima in the loss landscape to improve the stability of convergence. The learning rate will never drop under "1e-6", avoiding getting stuck at a rate close to zero.
- Model Checkpoint. It keeps the model weights anytime the validation loss gets a new minimum. This is a safeguard against performance degradation later during training, allowing us to always re-load the optimal version.

This setup represents a solid foundation for modeling medical ECG signals, resulting compact and regular model, tuned to be stable in a domain known for its noise. The detailed explanation of the hyper-parameters used in this proposed model is depicted in Table 2.

3.5 Applied Models for Comparison

Along with the optimized GRU model, two efficient deep learning models are also implemented.

- The optimized LSTM model involves 128-unit LSTM and 64-unit LSTM layers, each layer with dropout to inhibit the overfitting issue. These upper layers are concluded with a Dense (of 5 output neurons) with a Softmax function as the final layer to perform arrhythmia multi-classification. This model utilizes the "Adam" as optimizer and sparse categorical cross-entropy loss to be trained on scaled data for "twenty" epochs and "thirty-two" batch size, with model performance validated on an individual validation set. Similar to the GRU, the LSTM model was optimized using three callbacks.
- Multi-layer 1D CNN model was constructed for ECG signal classification into five classes. This model initiates with an input layer formed

to match the normalized training data, accompanied by three convolutional blocks, each of which begins with a convolution of 64 filters for extracting temporal features from the arrhythmia signals. These are combined with batch normalization and maximum-pooling layers for stabilizing learning and minimizing the dimensionality of features. This model was ended with flatten and two dense layers (of 64 and 32 neurons) utilizing ReLU function, and a Softmax function to produce the probabilities for the targeted classes.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the experimental assessment of the optimized GRU and LSTM models, and a multi-layer 1D CNN model, using the MIT-BIH arrhythmia dataset, is presented.

The analysis here follows a comparative approach, where results for temporal and spatial modules are presented and discussed to illustrate the individual contributions of each model. Results for performance metrics (like precision, recall, specificity, F1 score, and accuracy [24]) are presented for all models applied, as demonstrated in Table 3.
Following balanced and strictly controlled experiments on the MIT-BIH dataset, the optimized GRU model delivers 99.49% accuracy, outperforming other implemented models and more complex pipelines for related systems, while maintaining a lightweight and easy-to-deploy structure appropriate for real-world medical applications.

The training loss versus validation loss curves and their accuracy relative to the applied models, shown in Figure 6, demonstrate the superiority of the improved GRU model across twenty epochs. Thanks to the callbacks used, the optimizer, and the loss, the model was able to learn efficiently, retain the best weights throughout the training period, and prevent underfitting and overfitting.

The confusion matrices depicted in Figure 7 demonstrated high accuracy achieved by the proposed GRU-based system's capability of efficiently classifying ECG signals.

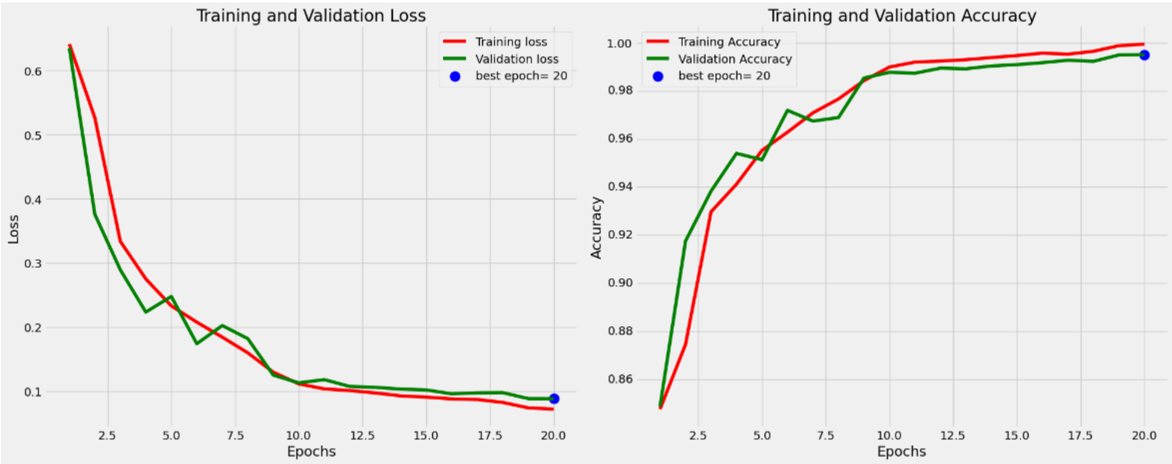
Considering the results in Figures 6 and 7, the proposed system demonstrates strong learning capabilities, considerable stability, and outstanding generalization to unseen or new data, especially the proposed optimal GRU model (achieved about 99% accuracy with a low loss of about 0.1), which is an indication of a well-tuned model and efficient hyper-parameters.

In summary, the findings signify a robust capability of classifying arrhythmia kinds. The high macro and weighted average (M_{avg} and W_{avg}), and F1-scores for the whole classes indicate that the dataset balancing and callbacks performed well for the optimized GRU model in real-world scenarios. Figure 8 demonstrates a comparison between our proposed GRU model and closely relevant works that utilized the same dataset and reported all the adopted performance metrics.

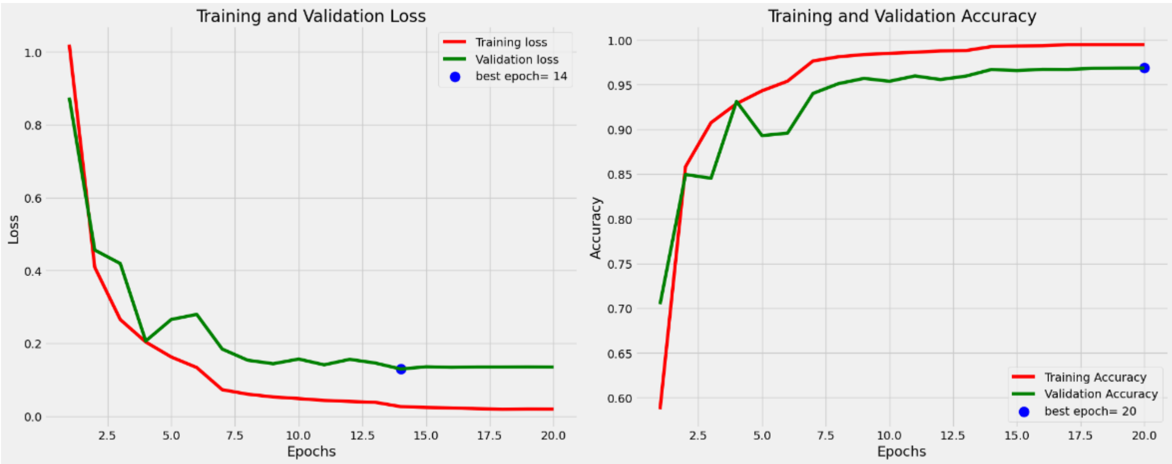
However, there is a slight room for improvement in Classes 1 and 3, where the precision results are slightly lower.

Table 3: Metrics assessment for the implemented models.

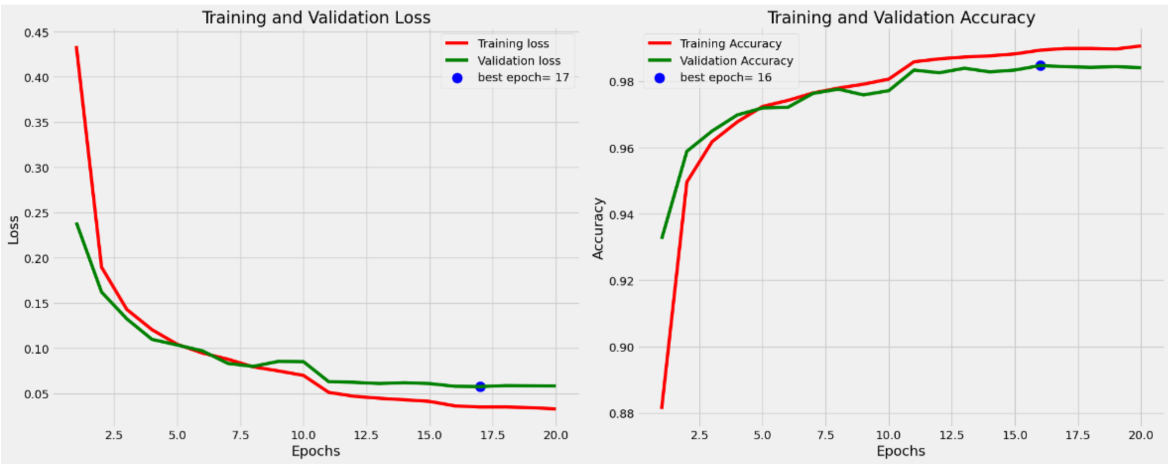
	Optimized GRU Model			Optimized LSTM Model			Multi-layer CNN Model			Supported
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
0	1.00	0.98	0.99	1.00	0.97	0.98	0.99	0.99	0.99	9059
1	0.99	1.00	0.99	0.58	0.86	0.69	0.81	0.85	0.83	278
2	1.00	0.99	0.99	0.90	0.96	0.93	0.95	0.96	0.95	724
3	0.99	1.00	1.00	0.53	0.89	0.66	0.72	0.86	0.79	81
4	1.00	1.00	1.00	0.98	0.99	0.98	0.99	0.99	0.99	804
AC	0.99			0.97			0.98			10946
M_{avg}	0.99	0.99	0.99	0.80	0.93	0.85	0.89	0.93	0.91	10946
W_{avg}	0.99	0.99	0.99	0.97	0.97	0.97	0.98	0.98	0.98	10946



(a)



(b)



(c)

Figure 6: Training versus validation losses and accuracies concerning: a) optimized GRU model, b) optimized LSTM model, c) multi-layer CNN model.

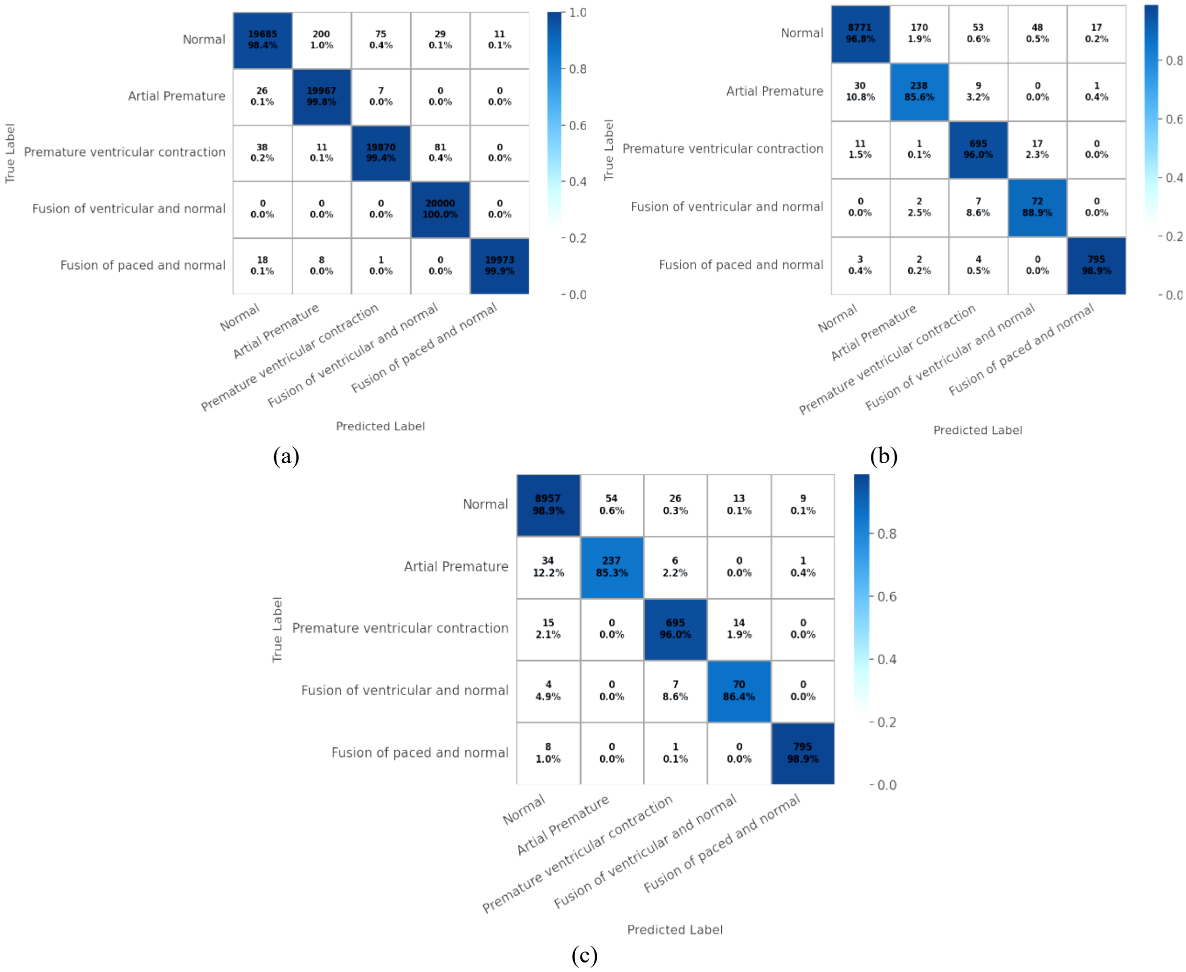


Figure 7: Confusion matrices concerning: a) omptimized GRU model, b) optimized LSTM model, c) multi-layer CNN model.

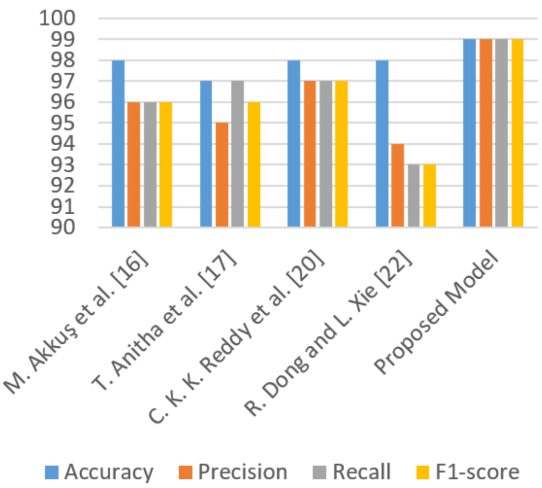


Figure 8: A comparison with closely relevant works.

5 CONCLUSIONS

This proposed deep learning–based system presented a cardiovascular disease prediction framework utilizing ECG Arrhythmia signals captured from the MIT-BIH Dataset, which compares an optimized GRU model with optimized LSTM and multilayer 1D CNN models. Based on balanced data resampling, preprocessing with MinMax Scaler, and incorporating diverse adaptive callbacks (Early Stopping, Reduce LROn Plateau, and Model Checkpoint), the proposed optimized models achieved stable and efficient training. Among the implemented models, the optimized GRU achieved superior classification results (99% accuracy, precision, recall, and F1-score), confirming its outstanding capability to acquire temporal dependencies while preserving computation efficiency. These findings confirm the potential of the

GRU-based system as an effective and accurate model for arrhythmia auto-classification and early diagnosis of cardiovascular diseases.

Future work should concentrate on extending the generalizability of the system via assessing its performance on more diverse and larger ECG datasets. Incorporating hybrid architectures or attention mechanisms could enhance spatiotemporal feature extraction.

REFERENCES

[1] M. N. Hasan, M. A. Hossain, and M. A. Rahman, "An ensemble based lightweight deep learning model for the prediction of cardiovascular diseases from electrocardiogram images," Engineering Applications of Artificial Intelligence, vol. 141, 109782, 2025.

[2] Z. Safdar, J. Sheng, M. U. Saeed, M. Ramzan, and A. Al-Zubaidi, "Empowering cardiovascular diagnostics with SET-MobileNet: A lightweight and accurate deep learning based classification approach," Image and Vision Computing, vol. 162, 105684, 2025.

[3] M. Tayyeb, M. Umer, K. Alnowaiser, S. Sadiq, A. A. Eshmawi, R. Majeed, A. Mohamed, H. Song, and I. Ashraf, "Deep Learning Approach for Automatic Cardiovascular Disease Prediction Employing ECG Signals," CMES - Computer Modeling in Engineering and Sciences, vol. 137, no. 2, pp. 1677-1694, 2023.

[4] S. Sowmya and D. Jose, "Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model," Measurement: Sensors, vol. 24, 100558, 2022.

[5] R. A. Marnani, R. Jaros, J. Pavlicek, R. Martinek, and R. V. Kahankova, "Advancements and Challenges in Non-Invasive Electrocardiography for Prenatal, Intrapartum, and Postnatal Care: A Comprehensive Review," in IEEE Access, vol. 12, pp. 44730-44747, 2024.

[6] M. A. Talukder, "A hybrid multiscale feature fusion model for enhanced cardiovascular arrhythmia detection," Results in Engineering, vol. 25, 104244, 2025.

[7] S. Albawi, M. H. Arif, and J. Waleed, "Skin cancer classification dermatologist-level based on deep learning model," Acta Scientiarum. Technology, vol. 45, no. 1, e61531, 2022.

[8] A. Mir, A. U. Rehman, T. M. Ali, S. Javaid, M. F. Almufareh, M. Humayun, and M. Shaheen, "A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques," ESC Heart Failure, vol. 11, no. 6, pp. 3742-3756, 2024.

[9] H. Q. Flayyih, J. Waleed, and A. M. Ibrahim, "Indoor Air Quality Prediction in Sick Building Using Machine and Deep Learning: Comparative Analysis," Diyala Journal of Engineering Sciences, vol. 18, no. 1, pp. 203-218, 2025.

[10] J. Waleed, T. M. Hasan, A. J. Abdullah, and A. Alkhayyat, "Precious Metal Prices Forecasting Using Optimally Configured Hybrid Deep Learning Approach," Journal of Machine and Computing, vol. 5, no. 3, pp. 1823-1838, 2025.

[11] G. M. Rao, D. Ramesh, V. Sharma, A. Sinha, M. M. Hassan, and A. H. Gandomi, "AttGRU-HMSI: enhancing heart disease diagnosis using hybrid deep learning approach," Scientific Reports, vol. 14, no. 7833, 2024.

[12] M. S. Haleem, R. Castaldo, S. M. Pagliara, M. Petretta, M. Salvatore, M. Franzese, and L. Pecchia, "Time adaptive ECG driven cardiovascular disease detector," Biomedical Signal Processing and Control, vol. 70, 102968, 2021.

[13] B. S. Tchinda, D. Tchiotsop, and R. Tchinda, "Classification of cardiac arrhythmias using ECG shape-adapted wavelets and BiLSTM network," Measurement, vol. 257, part C, 118832, 2026.

[14] P. Mahajan and A. Kaul, "Graph-enhanced deep learning for ECG arrhythmia detection: An integration of CNN-GNN-BiLSTM approach," Medical Engineering & Physics, vol. 145, 104418, 2025.

[15] K. Lee, J. Lee, and M. Shin, "Lightweight beat score map method for electrocardiogram-based arrhythmia classification," Biocybernetics and Biomedical Engineering, vol. 44, no. 4, pp. 844-857, 2024.

[16] M. Akkuş, M. Karabatak, and R. Tekin, "Spindle Autoencoder-CNN hybrid model for cardiac arrhythmia classification," Computers in Biology and Medicine, vol. 195, 110593, 2025.

[17] T. Anitha, S. Aanjankumar, R. K. Dhanaraj, D. Pamucar, and V. Simic, "A deep Bi-CapsNet for analysing ECG signals to classify cardiac arrhythmia," Computers in Biology and Medicine, vol. 189, 109924, 2025.

[18] Y. Zhang, X. Li, L. Zhang, J. Wang, S. Jiang, Y. Ma, and D. Li, "A sequential MAE-clustering self-supervised learning method for arrhythmia detection," Expert Systems with Applications, vol. 269, 126379, 2025.

[19] A. Gupta, S. Panday, and C. Bhatnagar, "Arrhythmia Detection from ECG Signal Using Long Short Term Memory Network Model," Procedia Computer Science, vol. 258, pp. 1102-1111, 2025.

[20] C. K. K. Reddy, A. Daduvy, V. S. Kaza, M. Shuaib, M. Mohzary, S. Alam, and A. Sheneamer, "A multi-scale convolutional LSTM-dense network for robust cardiac arrhythmia classification from ECG signals," Computers in Biology and Medicine, vol. 191, 110121, 2025.

[21] S. Mangaraj, K. Mahapatra, and S. Ari, "HLS-compiled PYNQ-based cardiac arrhythmia detection system leveraging quantized ECG beat images," Biomedical Signal Processing and Control, vol. 109, 108063, 2025.

[22] R. Dong and L. Xie, "A multimodal cross-attention-based CNN-LSTM network for arrhythmia classification and clinical diagnostic support," Biomedical Signal Processing and Control, vol. 112, 108697, 2026.

[23] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," in IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, 2001.

[24] J. Waleed et al., "An Effective Deep Learning Model to Discriminate Coronavirus Disease From Typical Pneumonia," International Journal of Service Science, Management, Engineering, and Technology, vol. 13, no. 1, pp. 1-16, 2023.