Rectified Adam Optimizer and LSTM with Attention Mechanism for ECG-Based Multi-Class Classification of Cardiac Arrhythmia

T. SIVARANJANI [1], B. SASIKUMAR [2], G. SUGITHA [3]

[1] Dept. of Computer Science and Engineering, Bharath Niketan Engineering College, Aundipatty, 625531, Theni City, India
[2]Dept. of Computer Science and Engineering, Infant Jesus College of Engineering, Kamarajar Nagar, 628851, Tuticorin City, India
[3]Dept. of Computer Science and Engineering, Muthayammal Engineering College (Autonomous), Rasipuram, 637408, Namakkal City, India

sivaranjani89@bnec.ac.in, sasi.csrrc@gmail.com, sugitha.g.cse@mec.edu.in

Submitted October 23, 2024 / Accepted January 11, 2025 / Online first April 9, 2025

Abstract. *Cardiac Arrhythmia (CA) is one of the most prevalent cardiac conditions and prime reasons for sudden death. The current CA detection methods face challenges in noise removal, R-peak detection, and low-level feature selection, which can impact diagnostic accuracy and signal stability. The research aims to develop an effective framework for detecting and classifying CA using advanced signal processing, feature extraction, feature selection, and classification for reliable medical diagnosis. The input electrocardiogram (ECG) signals are processed using hybrid noise reduction techniques such as cascaded variable step size normalized least mean square and sparse low-rank filter. The complex and high-level features are extracted using higher-order spectral energy distributed image, wavelet transform, and R-wave peak to R-wave peak interval to enhance the representation of cardiac data. Recursive feature elimination is applied to select the most relevant diagnostic features and the Rectified Adam optimizer is used to fine-tune parameters to achieve better training stability. The model integrates long-term memory with an attention mechanism to enhance the classification performance of arrhythmia detection. Simulation results demonstrate that the proposed model achieves 99.40% accuracy, outperforming existing models and showing its efficiency in classifying CA for better diagnosis and early treatments.*

Keywords

Cardiac arrhythmia, electrocardiogram, sparse low-rank filter, recursive feature elimination, long short-term memory, rectified Adam optimizer, attention mechanism

1. Introduction

Arrhythmias are common cardiovascular diseases characterized by abnormal heart activity, often stemming from underlying issues such as coronary heart disease. This disease occurs from the inadequate blood supply and coronary arteries narrowing, resulting in myocardial dysfunction [1]. Sudden cardiac death syndrome with a morphologically typical heart contributed to 18% of all cardiac deaths in 2022 [2]. Idiopathic ventricular fibrillation is a diagnosis performed after excluding other causes of sudden cardiac arrest, such as structural, metabolic, electric, or toxicological etiologies. Its incidence has decreased with improvement in testing and diagnosis of primary arrhythmia syndromes [3]. Current treatments include SSYX metoprolol, which provides cardio protection and antiarrhythmic effects for patients with CA [4]. Additionally, ruxolitinib has the potential to act as a CaMKII inhibitor in the treatment of CA [5], [6]. The latest scientific findings are incorporated into clinical practice by episodically updating the treatment of ventricular arrhythmia patients and preventing abrupt cardiac death [7]. Clinically significant arrhythmias like persistent ventricular, atrial fibrillation, and second-degree atrioventricular (AV) block need constant electrocardiogram (ECG) monitoring. Atrial fibrillation and flutter are particularly prevalent as perioperative complications in surgical settings [8].

Various scientific methods have been analyzed and developed to diagnose and classify CA efficiently. Some advancements include models that utilized innovative techniques like a Chi-square distance classifier with Particle Swarm Optimizer (Chi-Square-D-C PSO) and Generative Adversarial Networks-SkipNet (GAN-SkipNet) combined with an attention mechanism [9], [10]. Other models, such as improved ResNet and Bidirectional Long Short-Term Memory models (Res-BiANet), and Convolution, Attention, and Transformer-based Networks (CAT-Net), have been specially designed for detecting and classifying multiple arrhythmias using photoplethysmography (PPG) signals [11], [12]. Despite these advancements, challenges remain, in noise removal, feature extraction, feature selection, and optimization strategies. To overcome this, we propose a model that incorporates hybrid noise reduction, feature extraction,

DOI: 10.13164/re.2025.0195

feature selection, and optimization to improve arrhythmia detection and enhance model performance. By investigating the automated detection of CA, we analyze and explore solutions for the following research questions.

1. How can enhancing ECG signal quality lead to more reliable detection of arrhythmias, especially in noisy or low-quality data?
2. Does the feature extraction impact the early diagnosis of arrhythmias often missed in standard diagnostic processes?
3. How can optimizing the learning process improve the consistency and accuracy of arrhythmia detection across patients with varying heart conditions?
4. In what ways does a more precise analysis of time-based patterns in ECG signals improve the detection of arrhythmia, particularly in patients with irregular symptoms?

This research aims to improve the detection of CA by enhancing the quality of ECG signals, particularly for noisy data. The work enables the detection and classification of arrhythmias through feature extraction, leading to timely intervention. Optimizing the learning process ensures consistent accuracy across different patients. Time-based pattern analysis can help in identifying arrhythmias in patients with irregular symptoms, potentially improving the diagnostic process and patient care.

The main contributions of the proposed model are presented as follows:

- The research introduces hybrid noise reduction techniques Cascaded Variable Step Size Normalized Least Mean Square (CVSS-NLMS) and Sparse Low-Rank Filter (SLRF) to improve the quality of ECG signals, making them more suitable for the subsequent feature extraction process.
- The model employs three sophisticated feature extraction techniques: the Higher Order Spectral Energy Distributed Image (HSDI), Wavelet transform, and R-wave peak to R-wave peak (R-R) interval. This approach captures the non-linear characteristics, analyzes transient features, and calculates intervals between R-wave peaks analysis to provide a comprehensive and richer representation of ECG signals, leading to better detection of subtle arrhythmias.
- The proposed approach employs the Recursive Feature Elimination (RFE) method to reduce the irrelevant features and enhance the model’s classification performance.
- The incorporation of a Rectified Adam (RAdam) optimizer addresses the issue related to instability during the early training phase and provides a more effective learning process. This also improves the model performance, minimizes loss, and leads to faster convergence.
- Unlike traditional Convolutional Neural Networks (CNNs), which excel at identifying spatial patterns, the proposed model uses Long Short-Term Memory (LSTM) networks to capture temporal dependencies and emphasize significant features in the ECG data. The attention mechanism helps the model prioritize significant time steps in the ECG signal, leading to improved arrhythmia detection.

The structure of the research paper is as follows: Section 2 examines the related work of CA diagnosis and classification, Section 3 explains the pipeline of the proposed model, Section 4 depicts the implementation of the proposed model and compares it with the existing models, and Section 5 concludes the study.

2. Related Work

Numerous researchers and clinicians have analyzed and developed models for detecting and categorizing the CA, each with a unique objective, methodology, advantages, and limitations. We discuss a few of those recent studies here. Ahmed et al. [13] developed a fast and automated method for diagnosing and classifying CA using a one-dimensional CNN (1D-CNN). It offered high accuracy and automatic feature extraction but faced data imbalance and generalization issues. An automated classification model [14] was developed to address data imbalance issues in classifying CA using a combination of CNN and Recurrent Neural Network (RNN). This showed high accuracy but needed further refinement by exploring different neural network architectures. In another work, Ruben et al. [15] aimed to enhance an Internet of Things (IoT)-based system for monitoring ECG signals and detecting arrhythmia. The work employed the Machine Learning (ML) algorithm K-Nearest Neighbor algorithm (KNN) for arrhythmia detection, offered high accuracy, and provided real-time alerts, which were limited by data misclassification due to the quality of ECG data. For this, Dhyani et al. [16] used a 3D wavelet transform for an ECG-based arrhythmia detection system and a Support Vector Machine (SVM) for accurate classification, achieving a high accuracy of 99.02%. However, the model used a shorter ECG signal, which could not be sufficient for comprehensive detection.

The work of Yang et al. [17] utilized 12-lead ECG signals to improve arrhythmia classification and a dual-channel Deep Neural Network (DNN) to retrieve characteristics for enhancing accuracy and effectively fused multimodal features but had high computational complexity and potential overfitting. Considering these issues, an automated DL model was developed by Daydulo et al. [18] to classify ECG signals to detect CA. This model achieved high classification accuracy and demonstrated robustness in detection, reducing diagnosis time and errors. The work needed real-time data validation for improved robustness. To improve the real-time validation, a tiny CNN for the real-time model was developed by Farag [19] for classifying ECG and detecting arrhythmias at the edge, leveraging matched filter theory. It achieved high accuracy and efficiency but faced noise and inter-patient variability challenges.

Mora et al. [20] developed a model that automated the hyperparameter selection process using a hybrid model combining CNN and PSO to classify CA and achieved a high accuracy of 98%. However, balancing the dataset and including more CA classes could improve the model's performance. To overcome this, Kim et al. [21] used 12-lead ECG data with 45 different classes to develop an automated multi-label CA detection model. They employed CNN with a Global Channel Attention Block (GCAB) and Short Residual Block (SRB), efficiently using end-to-end learning to combine feature extraction and categorization. The insufficient data samples for some arrhythmia classes limited the model's performance. To improve this, Din et al. [22] aimed to enhance ECG-based arrhythmia detection by fusing CNN, LSTM, and transformer features and achieved a high accuracy of 99.56% but was limited to binary classification and could not extend to multi-class classification.

Thus, by analyzing the existing models, we noticed some issues still prevailing, such as effective noise removal, feature extraction, feature selection, and standard optimizers for classification. The proposed model is developed to address these concerns using hybrid preprocessing techniques, feature extraction, and selection methods, an improved optimizer, and an innovative model architecture for improving classification performance.

3. Proposed Cardiac Arrhythmia Detection and Classification

The proposed model enhances CA detection by employing hybrid noise reduction techniques CVSS-NLMS and SLRF to improve the quality of ECG signals, followed by feature extraction using HSDI, wavelet transform, and R-R intervals analysis for capturing detailed and non-linear features. RFE is applied to eliminate irrelevant features, improving classification accuracy. The RAdam optimizer ensures a stable and efficient training process while integrating LSTM with an attention mechanism that captures temporal dependencies and focuses on significant features, improving arrhythmia detection. The workflow of the proposed diagnostic model is depicted in Fig. 1.

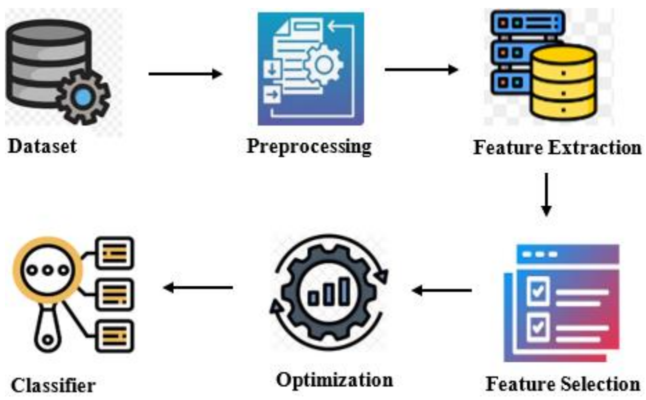

Fig. 1. The proposed model for detecting and classifying cardiac arrhythmia.

$\mu(b)$	Adaptive filter at iteration b
z	Input signal
α	Step size
E	Error signal
$\vartheta_{\max}$	Major eigenvalue of the auto-correlation matrix
$\mathbf{C}$	Original data
$\mathbf{LR}$	Low-rank matrix
$\mathbf{SP}$	Sparse matrix
$\mathbf{N}$	Noise matrix
$\mathbf{H}^{3d}$	HSDI matrix of 3D-CNN
H_{in}^{patch}	Input patch of the data
$H^{reshape3d}$	Reshaped 2D-CNN input
$\mathbf{H}^{2d}$	HSDI matrix of 2D-CNN
H^{out1}	Output 1 of 2D-CNN
H^{out2}	Output 2 of 2D-CNN
M_0S	Zero-order scattering coefficient
M_1S	First-order scattering coefficients
M_2S	Second-order scattering coefficients
M_nS	n-th order scattering coefficients
P_1	First-order wavelet modulus transforms
P_2	Second-order wavelet modulus transforms
S	Convolving time
∂_d	Wavelet function
φ_d	Scaling function
d_i	The scale at index i
r_i	Set of scale
c	Time
$\mathbf{MS}$	Final scattering matrix
O	The target function
T	Actual value
$\hat{T}$	Predicted value
T^i	Definite value
$\hat{T}^i$	Projected value
G_s	Gradient of the objective function
$\beta^{\propto}$	The maximum length
v_s	First-moment moved
w_s	Second-moment moved
$\hat{v}_s$	First-moment moving reverse
$\hat{w}_s$	Second-moment moving reverse
ρ_s	Decay rate
rect_s	Rectifier
ω_s	Model parameter
n	Samples sequence

Tab. 1. Glossary of parameters.

3.1. Dataset Overview

The proposed model uses the MIT-BIH Arrhythmia dataset, which encompasses 48 ECGs from 47 patients examined in Beth Israel Hospital Arrhythmia Laboratory between 1975 and 1979. There are 48 Comma-Separated Values (CSVs), each with 30 minutes at 360 Hz ECG from a single patient providing high-resolution time-series data for analysis. The dataset providing a total of 109,446 annotated heartbeats also includes reference annotations for each beat, which two or more cardiologists determined, and any discrepancies were resolved. The annotations provided with the datasets are considered to be reliable and accurate. The dataset consists of five classes: N-normal, L-left bundle branch block, R-right bundle branch block, A-atrial premature beat, and V-ventricular premature beat.

3.2. Signal Processing and Noise Reduction

Signal processing and noise reduction ensure that the gathered data is in a suitable format for further analysis. The proposed model utilizes hybrid noise reduction techniques, such as CVSS-NLMS and SLPF, to remove the noise baseline wander, powerline interface, and muscle artifacts from the ECG signal. These methods are applied sequentially to improve signal quality by reducing noise, making them more suitable for feature extraction.

3.2.1. Cascaded Variable Step Size Normalized Least Mean Square

The normalized least mean square adaptive filter is cascaded to achieve a higher Signal-to-Noise Ratio (SNR) value, constancy, and quick convergence rate. Inadequate excitation in an input sequence may result in excessive parameter estimations, which might cause problems like overflow, inferior filter performance, and unbounded prediction error [23]. To mitigate these issues, the Least Mean Square (LMS) filter's normalized component helps stabilize the system. The NLMS technique increases stability and tracking capabilities when adapting to changes in the input signal. Equation (1) illustrates how the normalized component is added to the weight update equation to increase performance. The parameters involved in the proposed work are introduced in Tab. 1.

$$\mu(b+1)=\mu(b)+\frac{\alpha}{|z(b)|^2}E(b)z(b) \tag{1}$$

where $\mu(b + 1)$ represents updated weight vector for the adaptive filter at iteration $b + 1$, $\mu(b)$ represents the weight vector at the current iteration b, $z(b)$ denotes the input signal at iteration b, $E(b)$ represents error signal at iteration b, α is the step size, and $|z(b)|^2$ represents the power of the input signal.

Choosing a step size for the LMS algorithm is challenging because a large step size allows fast adaptation, but it also leads to a significant increase in Mean Squared Error (MSE). If the step size is too large, stability will be lost. A small step size leads to delayed convergence even with the small additional MSE. To maintain stability in the LMS algorithm, it is necessary to impose the upper boundary for the step size, as shown in (2):

$$0<\alpha<\frac{2}{\vartheta_{\max}} \tag{2}$$

where $\vartheta_{\max}$ denotes the major eigenvalue of the autocorrelation matrix of the input signal.

3.2.2. Sparse Low-Rank Filter

After the first filtering process, the data goes to the SLRF filtering process for further processing. The data usually contains noise, which degrades the model's performance. A sparse low-rank matrix decomposition method solves the problem of noise interference in structured data [24]. In this method, the data is represented as the sum of a low-rank, sparse, and noise matrix, as shown in (3):

$$\mathbf{C}=\mathbf{LR}+\mathbf{SP}+\mathbf{N} \tag{3}$$

where $\mathbf{C}$ is the original data, $\mathbf{LR}$ represents the low-rank matrix, $\mathbf{SP}$ represents the sparse matrix and $\mathbf{N}$ represents the noise matrix. These preprocessing steps work together to improve the signal quality and suitable for supplementary analysis.

3.3. ECG Waveform Analysis – Feature Extraction

After preprocessing, the proposed model employs ECG waveform analysis through feature extraction techniques such as HSDI for capturing non-linear characteristics, Wavelet transforms for analyzing the transient features, and R-R intervals analysis to calculate intervals between R-wave peaks. This provides a comprehensive and richer representation of ECG signals, leading to better detection of arrhythmias.

3.3.1. Higher Order Spectral Energy Distributed Image

HSDI is utilized to capture the non-linear characteristics of ECG signals by employing 3D-CNN. Although ECG signals are 1D, the HSDI transforms them into a high-dimensional representation that captures temporal and spectral patterns. This transformation allows the use of 3D-CNN to effectively extract complex patterns and inter-feature relationships from these representations, enhancing the model's ability to detect arrhythmias.

Initially, we process an ECG signal segment into an HSDI matrix of size $s \times d \times e$ using 3D-CNN. Equation (4) expresses the computational approach using eight convolutional kernels of size $3 \times 3 \times 3$.

$$\mathbf{H}^{3d}=\mathrm{ReLU}\left(\mathrm{N}\left(3\mathrm{dConvo}\left(\mathbf{H}_{\mathrm{in}}^{\mathrm{patch}}\right)\right)\right) \tag{4}$$

where 3dConvo represents the 3D convolutional, N represents the normalization, ReLU is the rectified linear activation function, and $\mathbf{H}_{\mathrm{in}}^{\mathrm{patch}}$ input patch of the data [25].

3D-CNN outputs a high-dimensional feature block $\mathbf{H}^{3\mathrm{d}} \in \mathbf{A}^{(s-2)\times(d-2)\times(e-2)\times 8}$ contains the spectral-temporal information extracted from the ECG signal.

Now the dimensionality of $\mathbf{H}^{3\mathrm{d}} \in \mathbf{A}^{(s-2)\times(d-2)\times(e-2)\times 8}$ is reshaped to $\mathbf{H}^{\mathrm{reshape3d}} \in \mathbf{A}^{(s-2)\times(d-2)\times((e-2)\times 8)}$ as the 2D-CNN input. It is composed of several branches b each with distinct convolutional kernel magnitudes to obtain rich contextual data. The mathematical procedure is described in (5).

$$\mathbf{H}^{2d}=\mathrm{ReLU}\left(\mathrm{N}\left(2\mathrm{dConvo}\left(\mathbf{H}^{\mathrm{reshape3d}}\right)\right)\right). \tag{5}$$

The 2D-CNN adjusts the size of the convolutional kernel to capture the spectral-temporal feature effectively outputting $\mathbf{H}^{2\mathrm{d}} \in \mathbf{A}^{ck\times ck\times b}$, where ck varies depending on the convolutional kernel. We flatten $\mathbf{H}^{2\mathrm{d}}$ to $\mathbf{H}^{\mathrm{flatten2d}} \in \mathbf{A}^{(ck\times ck)\times b}$ and apply the weighted mapping $\mathbf{H}^{\mathrm{weighted2d}} \in \mathbf{A}^{(ck\times ck)\times b}$ to refine the features.

For representing key spectral information, we use 1D-CNN to minimize the loss produced by spectral characteristics during the patch convolutional as shown in (6):

$$\mathbf{H}^{\text{out2}} = \text{ReLU}\left(\text{N}\left(\text{1dConvo}\left(\mathbf{H}_{\text{in}}^{\text{patch}}\right)\right)\right). \tag{6}$$

Finally, we concatenate the output of 2D-CNN $\mathbf{H}^{\text{out1}}$ with 1D-CNN $\mathbf{H}^{\text{out2}}$ to generate the final output, which extracts the non-linear characteristics of ECG signals.

3.3.2. Wavelet Transform

Wavelet transform is used for feature extraction to analyze transient features by decomposing the ECG signal into time-frequency components. This builds translation-invariant, constant, and instructive signal depictions. The model is implemented with the deep convolutional network; the convolution generates a locally invariant feature S as shown in (7) and consequences in high-frequency of loss data [26]. This data is improved by the wavelet modulus transform as described in (8).

$$M_0 S(c) = S * \partial_d(c), \tag{7}$$

$$P_1 S = \left[M_0 S(c), S * \varphi_d(c)\right]_{d_i \in r_i} \tag{8}$$

where $M_0S(c)$ is the zero-order scattering coefficient obtained by convolving S with time c, $\partial_d(c)$ is the wavelet function at the scale d with time c, $\varphi_d(c)$ is the scaling function at the scale d with time c, P_1S denotes first-order wavelet modulus transform, and d_i is the scale at the index i and r_i is the set of scales at the index i where i = 1,2,3,…,n.

The scattering coefficients for the first-order M_1S are obtained by averaging the wavelet modulus with ∂_d as shown in (9).

$$M_1 S(c) = \left[\left[S * \varphi_{d_1}\right] * \partial_d(c)\right]_{d_1 \in r_1}. \tag{9}$$

Now, the low-frequency component can be considered as $S * \varphi_{d_i}$, which is used to restore the information lost by averaging $M_1S(c)$. The complementary high-frequency coefficient can be extracted by (10):

$$P_2\left[S * \partial_{d_1}\right] = \left[M_1 S(c), \left[S * \partial_{d_1}\right] * \partial_{d_2}(c)\right]_{d_2 \in r_2} \tag{10}$$

where P_2 is the second-order wavelet modulus transform. It can be further defined in (11) for second-order scattering coefficients $M_2S(c)$.

$$M_2 S(c) = \left[\left[S * \varphi_{d_1}\right] * \varphi_{d_2} * \partial_d(c)\right]_{d_i \in r_i}, i = 1, 2, 3, \ldots \tag{11}$$

Iterating the above (5) the wavelet modulus convolution is defined in (12).

$$V_n S(c) = \left[\left[S * \varphi_{d_1}\right] * \ldots \left[* \varphi_{d_n}\right]\right]_{d_i \in r_i}, i = 1, 2, 3, \ldots, n \tag{12}$$

where $V_nS(c)$ is the wavelet modulus convolution of order n. The n-th order scattering coefficients $M_nS(c)$ is shown in (13).

$$M_n S(c) = \left[\left[S * \varphi_{d_1}\right] * \ldots \left[* \varphi_{d_n}\right] * \partial_d(c)\right]_{d_i \in r_i}, i = 1, 2, 3, \ldots, n \tag{13}$$

The final scattering matrix $\mathbf{MS}(c)$ is defined in (14):

$$\mathbf{M\,S}(c) = \left[M_n S(c)\right]_{0 \le n \le e} \tag{14}$$

where d is the maximum decomposition order. To further enhance the scattering coefficients, renormalize the first-order $\tilde{M}_1S(c)$ and second-order $\tilde{M}_2S(c)$ coefficients in (15) and (16).

$$\tilde{M}_1 S(c) = \frac{M_1 S(c)}{S * \partial_d(c)}, \tag{15}$$

$$\tilde{M}_2 S(c) = \frac{M_2 S(c)}{M_1 S(c)}. \tag{16}$$

This process reduces coefficients redundancy by decorrelating different-order coefficients and extracts the transient features from the ECG signals.

3.3.3. R-R Intervals

The ECG signals are processed by removing abnormal R-R intervals, such as outliers and ectopic beats [27]. Any value that falls out of the series of 280 to 1500 milliseconds is considered an outlier, and any value that differs more than 20% from the previous RR interval is considered an ectopic beat. Removing these abnormal intervals ensures the analysis focuses on relevant data, leading to more accurate heart rate variability (HRV) feature extraction. The linear interpolation is used to fill in these abnormal values. Additionally, examining the HRV features from different periods, each R-R interval was split into 5-minute epochs deprived of overlying,

3.4. Enhanced Feature Selection using Recursive Feature Elimination

The proposed model explicitly focuses on selecting the most relevant features for classification to enhance model efficiency and performance by reducing dimensionality and focusing on the most informative features. After signal preprocessing and feature extraction, each feature is analyzed to determine the correlation between the percentage of positive and negative correlations. The proposed model uses the RFE, a widely used algorithm for selecting the features in the predictive model that are most essential for predicting the desired parameter. This technique uses a backward selection procedure to recognize possible features by removing unnecessary features. It determines the significance of each feature and constructs a predictive model with all these features. Then, it ranks in an order and recognizes the unnecessary features. Finally, it iteratively eliminates the least significant features from the model using metrics for model evaluation until the chosen features remain [28]. The selected feature set trains the model, improving interpretability and reducing computational complexity.

3.5. Rectified Adam Optimizer

The proposed system optimizes the objective function at a precise time based on the possible variance deviation by using the RAdam optimizer. After feature selection, the reduced and informative feature set is used to train the LSTM model. It provides a more effective learning model, particularly during early training phases, which improves overall model performance. The model incorporates a rectified term that enables the adaptive momentum to be gradually and steadily stated as a potential square difference function. This helps enhance the training model constancy and achieve better generalization and faster convergence, especially when working with complex data [29]. The objective function is represented in (17):

$$O\left(T,\hat{T}\right)=\frac{1}{n}\sum_{i=1}^{n}\left(T^{i}-\hat{T}^{i}\right)^{2} \tag{17}$$

where $O(T,\hat{T})$ denotes the target function that quantifies the mistakes between the actual T and predicted $\hat{T}$ values, n denotes the sample sequence, T^i denotes the definite value for the i-th sample, and $\hat{T}^i$ is the projected value for the i-th sample is expressed in Algorithm 1.

Algorithm 1: The functional process of RAdam optimizer

Objective: To fine-tune the model parameters
Input: step size δ_s, decay rate (ρ_1,ρ_2), $s = 0$
Output: ω_s

1. Initialize the first and second moment moving, and calculate the maximum length β^{∞}

$$\beta^{\infty}=\frac{2}{\left(1-\rho_2\right)}-1$$

2. $s = s + 1$
3. Determine the gradient G_s of the objective function, maximum length ρ_s of the approximated SMA, the first-moment v_s, and the second-moment w_s moved. Then reverse the first moment $\hat{v}_s$, and calculate $G_s = \nabla_{\omega}(\omega_{s-1})$

$$v_s=\frac{v_s}{\left(1-\rho_1^{s}\right)}$$

$$\rho_s=\rho_{\infty}-\frac{2s\rho_2^{s}}{\left(1-\rho_2^{s}\right)}$$

4. Calculate ω_s based on ρ_s.
5. If $\rho_s > 4$ then
6. Adopt the Adam optimizer, the second moment moving reverse, and construct a rectifier rect_s.

$$w_s=\frac{w_s}{\left(1-\rho_2^{s}\right)},$$

$$\text{rect}_s=\sqrt{\frac{\left(\rho_s-4\right)\left(\rho_s-2\right)\rho_{\infty}}{\left(\rho_{\infty}-4\right)\left(\rho_{\infty}-2\right)\rho_s}},$$

$$\omega_s=\omega_{s-1}-\frac{\delta_s\,\text{rect}_s v_s}{w_s}$$

7. Output the model parameters ω_s.

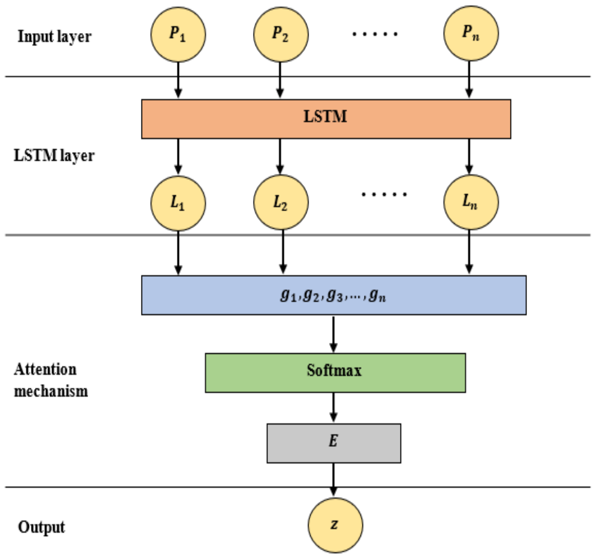

Fig. 2. The architecture of LSTM with an Attention Mechanism.

3.6. Arrhythmia Detection and Classification using LSTM with Attention Mechanism

The proposed model uses an innovative architecture combining LSTM with an attention mechanism. The traditional classifiers, such as CNNs, excel at identifying spatial patterns, whereas LSTM networks are particularly effective at learning temporal relationships by maintaining information across long sequences, a feature not inherently present in CNN. The attention mechanism enhances the model's ability to focus on the most relevant time steps in the ECG signals. Therefore, we use LSTM with an attention mechanism to capture temporal dependencies and focus on significant features, improving accuracy in arrhythmia detection. Based on a basic recursive neural network, a gating mechanism is incorporated into the LSTM, a gated recursive neural network, to regulate the neural network transmission. LSTM can handle intricate patterns and dependencies with an attention mechanism. When making a prediction, the LSTM attention mechanism allows the model to emphasize the significant input at every step. This improves the performance of the model and its constancy when compared to traditional LSTM models. After preprocessing, the input data represents individual features extracted from the ECG signals using three distinct feature extraction techniques. The RFE process identifies the most relevant and significant features, eliminating the irrelevant ones. The remaining selected features are combined into a sequential input fed into the LSTM with an attention mechanism. The input data is split into sequential steps by the LSTM in time series forecasting using an attention mechanism. From the attention mechanism, a fully connected layer receives the weighted average of the LSTM unit's hidden states as input for the final prediction. The following stages are used to implement the attention mechanism.

The attention mechanism's primary concept is to handle the input as a key-value pair. The query is the data that will be output. The associated weights are determined by calculating the similarity between each key and the query once the attention value has been determined. These weights are then normalized using a softmax function. The final attention value is obtained by weighting and adding the weight and the associated key value. To calculate the attention weight matrix the query vector $\mathbf{Q}_T$, key vector $\mathbf{K}_T$, and value vector $\mathbf{V}_T$ are derived from the LSTM's hidden states for each time step. These vectors are created by linearly transforming the hidden state $\mathbf{H}_T$ for a given time step T in the input sequence, as shown in (18).

$$\mathbf{Q}_T = \mathbf{W}_q \cdot \mathbf{H}_T, \mathbf{K}_T = \mathbf{W}_k \cdot \mathbf{H}_T, \mathbf{V}_T = \mathbf{W}_v \cdot \mathbf{H}_T \tag{18}$$

where $\mathbf{W}_q$, $\mathbf{W}_k$, and $\mathbf{W}_v$ are learned weight matrices.

The attention score for each time step is calculated by taking the dot product of the query vector, with each of the sequence's key vectors. The attention weight matrix is then obtained by scaling and applying a softmax function as expressed in (19):

$$\text{Attention score}\left(T, T'\right) = \frac{\mathbf{Q}_T \cdot \mathbf{K}_{T'}^T}{\sqrt{\dim_k}} \tag{19}$$

where $\dim_k$ represents the dimensionality of the key vectors. The softmax function is utilized to ensure that the attention weights sum to 1 across all time steps in (20).

$$\delta_{T,T'} = \text{softmax}\left(\frac{\mathbf{Q}_T \cdot \mathbf{K}_{T'}^T}{\sqrt{\dim_k}}\right) \tag{20}$$

where $\delta_{T,T'}$ represents the attention weight that determines how much focus the model places on the information at the time step T' when processing time step T.

Once the attention weight matrix is computed, they are used to generate a context vector $\mathbf{C}_T$ for each time step T. The context vector is a weighted sum of the value vector $\mathbf{V}_T$ over all time steps as shown in (21):

$$\mathbf{C}_T = \sum_{T'} \delta_{T,T'} \cdot \mathbf{V}_{T'} \cdot \tag{21}$$

This context vector encapsulates the relevant information from the entire sequence that is most important for making predictions at the time step. The final output for that time step is then generated by combining the context vector with the hidden state of the LSTM. This can be done either by concatenation or by a linear transformation followed by a non-linear activation function as expressed in (22):

$$\mathbf{H}_T = \tanh\left(\mathbf{W}_c \cdot \{\mathbf{H}_T; \mathbf{C}_T\}\right) \tag{22}$$

where $\mathbf{W}_c$ represents a learned weight matrix, and $\mathbf{H}_T$ represents the final output of the LSTM at the time step T. This enhanced hidden state $\hat{\mathbf{H}}_T$ is then used for classifying arrhythmias [30], as illustrated in Fig. 2.

4. Results and Discussion

The proposed model is implemented, and the results obtained and the experimental analysis made are discussed in this section. The experiment has been conducted on the following hardware and software configuration GPU: NVIDIA Quadro CPU: Intel® Xeon® CPU E5-1650 v3 @ 3.50 GHz M2000 python 3.10 x64-based processors and the 64bit Windows 10 pro-operating system.

4.1. Performance Analysis of the Proposed Model

In this section, the performance parameters of the proposed model for detecting and classifying the CA are extensively analyzed.

Figure 3 presents the intensity of the ECG signal for the five classes in the dataset, which are labeled as N-normal, L-left bundle branch block, R-right bundle branch block, A-atrial premature beat, and V-ventricular premature beat.

Figure 4 presents the ECG signal's intensity over the duration through three stages: the original signal, the noisy signal, and the denoised signal with moving average and Butterworth filter. The original signal, free of noise, consists of distinct peaks. The Noisy ECG signal appears to be a high-frequency fluctuation covered on the waveform. The noisy signal processed using a moving average filter results in a smoother waveform. The noisy signal is processed using

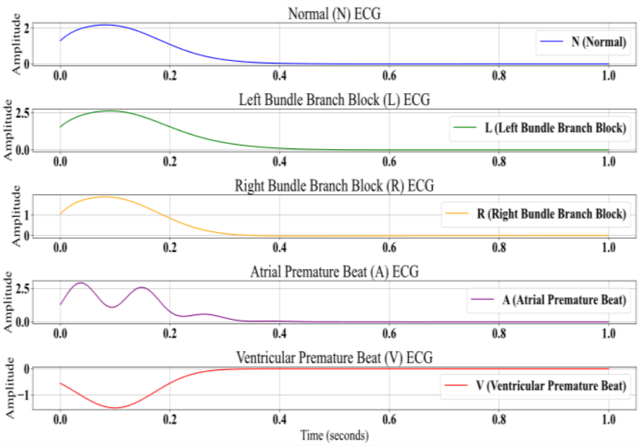

Fig. 3. Waveform of five classes in the MIT-BIH dataset.

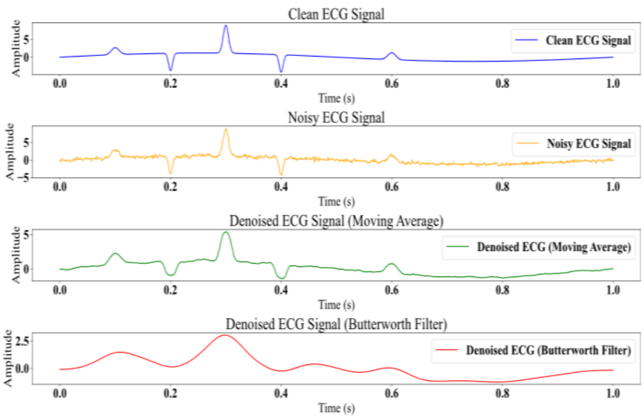

Fig. 4. ECG signal processing and noise reduction.

a Butterworth filter, which balances noise reduction and preservation of the signal's features.

The proposed model demonstrates outstanding performance across several metrics such as accuracy, precision, sensitivity, specificity, F_measure, Matthews Correlation Coefficient (MCC), and Negative Predictive Value (NPV) as 99.40%, 98.52%, 98.52%, 99.63%, 98.52%, 98.15%, and 98.15%, respectively, as illustrated in Fig. 5. This shows the proposed model's effectiveness in detection and classification. Additionally, the model has a low tendency for mistakes, with few false positives and false negatives, as seen by its False Positive Rate (FPR) of 0.37% and False Negative Rate (FNR) of 1.48%.

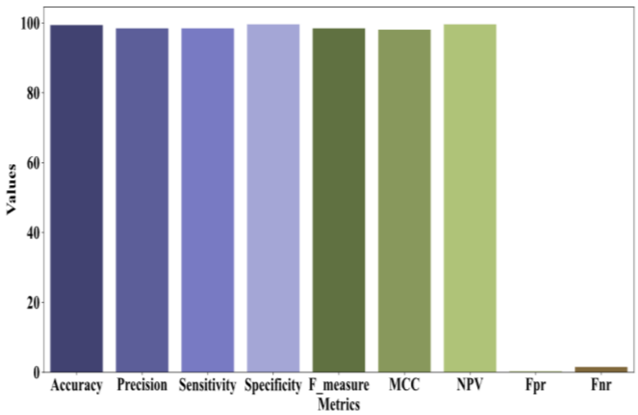

Fig. 5. Performance metrics of the proposed model.

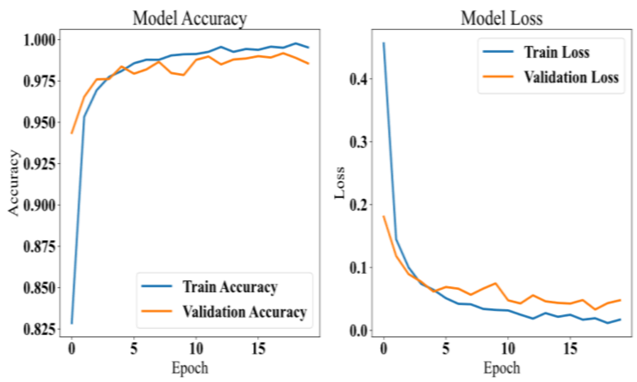

Fig. 6. Accuracy loss evaluation.

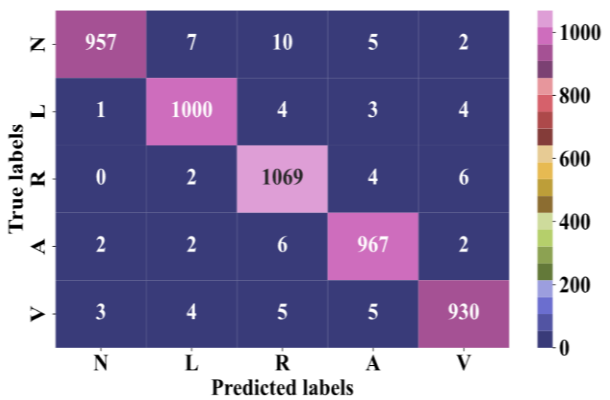

Fig. 7. Confusion matrix.

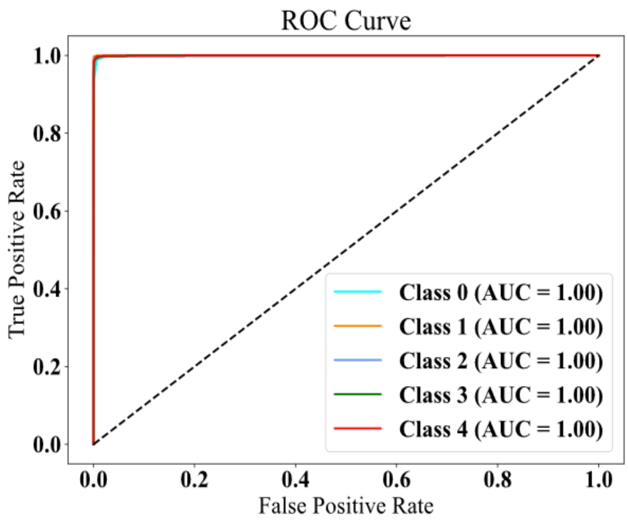

Fig. 8. ROC curves for the proposed model.

Figure 6 presents the accuracy and loss validation curve to determine the proposed model's effectiveness during training, providing valuable instincts about its learning progress over time. This curve is a critical tool for estimating the model's performance and adaptation to the dataset.

Figure 7 displays the confusion matrix used in classification models to evaluate their performance. It contrasts the model predictions with the actual target values. While off-diagonal cells show misclassifications, diagonal cells show accurate predictions. This reveals the reduction in misclassifications achieved by the proposed methodology.

Figure 8 demonstrates the ROC curves for the proposed model for accurately classifying CA conditions. The ROC curve highlights how true and false positive rates change at different threshold standards. The AUC indicates the ability of the model to differentiate positive and negative cases.

4.2. Comparative Analysis of the Proposed and Existing Models

Table 2 and Figure 9 compared the proposed model with existing models such as Chi-Square-D-C PSO, CNN with GCAB and SRB, and Res-BiANet. The proposed model achieved notable accuracy, precision, specificity, Recall, and F1 Score values of 99.4%, 98.52%, 98.52%, and 98.52%, respectively, outperforming the existing models.

The proposed model utilizes the MIT-BIH arrhythmia dataset for classification and achieved 99.40% accuracy. Figure 10 compares recent works such as Chi-Square-D-C PSO, CAT-Net, 1D-CNN, and the automated DL model using the same dataset, achieving 98%, 99.14%, 99%, and 99.2%, respectively, comparatively lower than our proposed model.

Figure 11 illustrates the computational efficiency comparison of the proposed model with existing models like Res-BiANet, CNN with GCAB, and Chi-Square-D-C PSO in terms of training time, memory usage, and inference time.

The proposed model has the shortest training time of 800 seconds, the fastest inference time of 30 milliseconds, and uses the least memory with 800 MB potentially due to optimizing model parameters and integrating feature extraction methods.

4.3. Ablation Study

To further validate and find the performance of the proposed model, we performed an ablation study by comparing

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score (%)
Chi-Square-D-C PSO [9] (2024)	98	98.18	98	98.03
CNN with GCAB + SRB [21] (2024)	95.7	79.3	77.6	77.5
Res-BiANet [11] (2024)	92.38	88.46	85.15	86.88
LSTM with attention mechanism (proposed)	**99.4**	**98.52**	**98.52**	**98.52**

Tab. 2. Comparison of performance metrics of the proposed model with existing models.

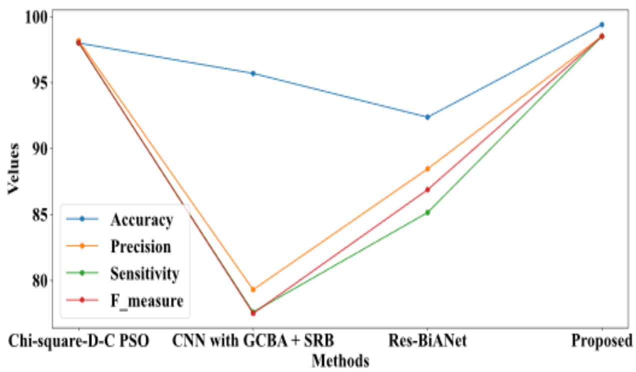

Fig. 9. Comparison of the proposed model with existing models.

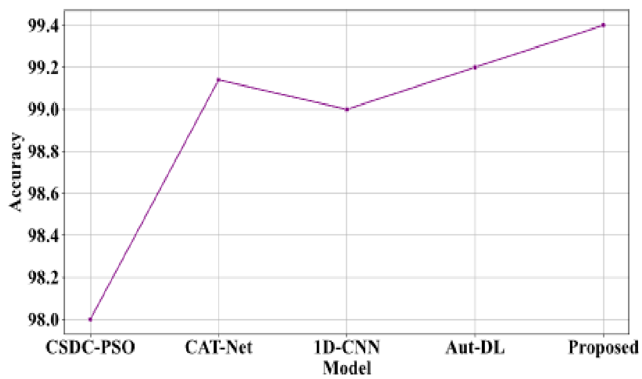

Fig. 10. Comparison with the recent works using the same datasets.

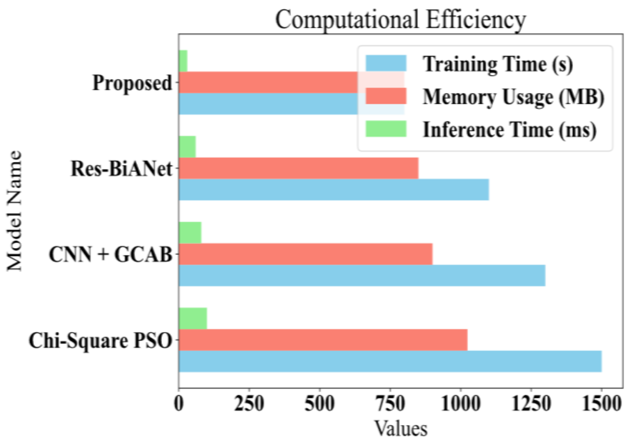

Fig. 11. Comparison of computational efficiency with the existing works.

the Adam optimizer instead of the RAdam optimizer and demonstrated the output in Fig. 12. The result shows the impact of optimization on improving the model performance across several metrics involved, such as accuracy, precision, sensitivity, specificity, F1_measure, MCC, NPV, FPR, and FNR, as 97.51%, 93.78%, 93.78%, 98.44%, 93.78%, 92.22%, 98.44%, 1.55%, and 6.22% respectively.

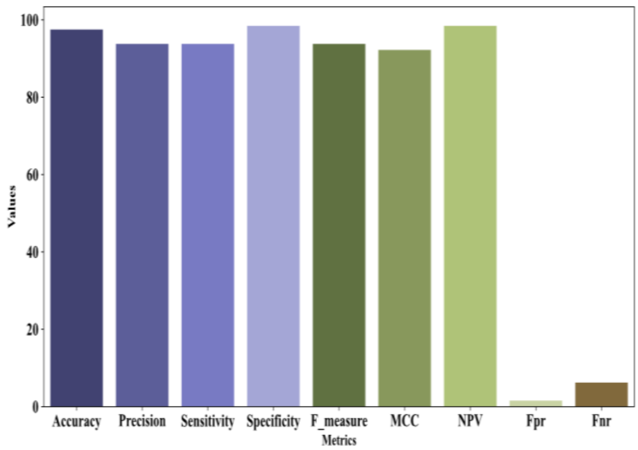

Fig. 12. Performance metrics of the proposed model without RAdam optimizer.

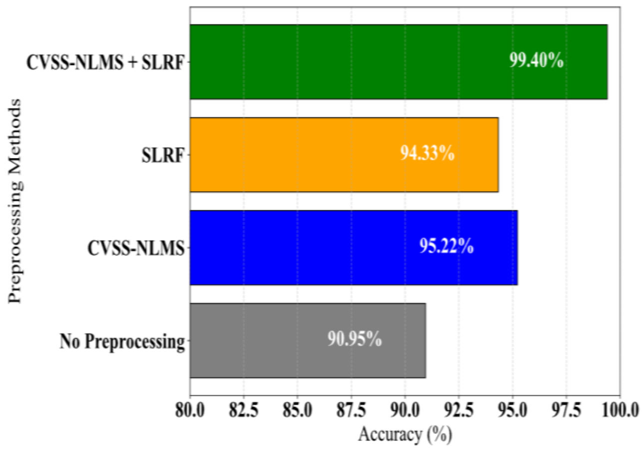

Fig. 13. Performance metrics of the proposed model with and without preprocessing steps.

Reference	Methodology	Accuracy
Al-Shammary et al. [9] (2024)	Chi-Square-D-C PSO	98%
Islam et al. [12] (2024)	CAT-Net	99.14%
Dhyani et al. [16] (2023)	SVM	99.02%
Farag [19](2023)	CNN	98.18%
Kim et al. [21] (2024)	CNN with GCAB and SRB	95.7%
Proposed Model	**LSTM with an attention mechanism**	**99.40%**

Tab. 3. Comparison of related work with the proposed model.

Additionally, we performed an ablation study by comparing the uses of preprocessing in the proposed model and demonstrated the output in Fig. 13. The result shows the impact of using the preprocessor CVSS-NLMS and SLRF on improving the model performance. The proposed work removes the noise baseline wander, powerline interface, and muscle artifacts from the ECG signal. The study also enhances the quality of ECG signals, making them more suitable for further processing.

4.4. Comparative Analysis with Recent Models on Classifying CA

Table 3 presents a comparison study of the suggested model against the analysis of related work papers. The current model attains a significantly higher accuracy of 99.40% indicating the proposed model's performance in classifying the CA.

4.5. Discussion

The study presents an improved method for detecting and classifying CA. Our approach employs hybrid noise reduction techniques that improve the quality of the ECG signals, making it more suitable for feature extraction. Utilizing feature extraction and selection techniques, we achieve a comprehensive and richer representation of ECG signals and focus on the most informative features. Integrating LSTM with an attention mechanism enables us to capture the temporal dependencies and focus on significant features, improving accuracy in arrhythmia detection. We analyze the ECG signals and evaluate the performance of the proposed model across various metrics. Further, we conducted an ablation analysis to examine the impact of incorporating optimization and preprocessing steps to improve the proposed model performance. Finally, we compare the proposed model with existing models and demonstrate that it outperforms them across all evaluated metrics

5. Conclusion

The proposed model presented a hybrid noise reduction, feature selection, and LSTM with an attention mechanism to improve CA detection and classification. The model employed enhanced noise reduction techniques such as CVSS-NLMS and SLRF. The proposed model ensures that ECG signals are high-quality for precise feature extraction. This enhancement in preprocessing improves the accuracy of subsequent analysis. The model leverages sophisticated feature extraction methods like HSDI, wavelet transforms, and R-R intervals. This comprehensive feature extraction approach provides a detailed representation of the signals for detecting subtle arrhythmias. Using RFE for feature selection improved the model performance by focusing on the most relevant features and enhanced both efficiency and accuracy in arrhythmia classification. Adopting the RAdam optimizer ensured a stable learning process during the early training phase, leading to faster convergence and improved overall performance. Finally, integrating LSTM with an attention mechanism enables the model's capability to detect arrhythmia with higher precision. The proposed model achieved an accuracy of 99.40%, demonstrating its potential for improving cardiac monitoring and diagnosis.

Availability of data and materials:
The code is available as public in kaggle repository:
https://www.kaggle.com/datasets/taejoongyoon/mitbit-arrhythmia-database/code

Availability of code:
The code is available as public in github repository:
https://github.com/Sivarangani/HT-1145.git

References

[1] KINGMA, J., SIMARD, C., DROLET, B. Overview of cardiac arrhythmias and treatment strategies. *Pharmaceuticals*, 2023, vol. 16, no. 6, p. 1–20. DOI: 10.3390/ph16060844

[2] BHATIA, K., D'SOUZA, R., MALHAMÉ, I., et al. Anaesthetic considerations in pregnant patients with cardiac arrhythmia. *BJA Education*, 2023, vol. 23, no. 5, p. 196–206. DOI: 10.1016/j.bjae.2023.01.008

[3] RAKZA, R., GROUSSIN, P., BENALI, K., et al. Quinidine for ventricular arrhythmias: A comprehensive review. *Trends in Cardiovascular Medicine*, 2025, vol. 35, no. 2, p. 73–81. DOI: 10.1016/j.tcm.2024.07.003

[4] WANG, X., LV, Y., YANG, J., et al. Clinical efficacy and safety of Shensong Yangxin capsule combined with metoprolol for treating cardiac arrhythmia: A meta-analysis of randomized controlled trials. *Phytomedicine Plus*, 2023, vol. 3, no. 4, p. 1–9. DOI: 10.1016/j.phyplu.2023.100478

[5] GUO, Q., HUO, Y., LIU, Q., et al. Ruxolitinib as a CaMKII inhibitor for treatment of cardiac arrhythmias: Applications and prospects. *Heart Rhythm*, 2025, vol. 22, no. 1, p. 231–239. DOI: 10.1016/j.hrthm.2024.07.118

[6] FEDERICO, M., VALVERDE, C. A., GONANO, L. A., et al. CaMKII: A link between metabolic disorders and cardiac arrhythmias. *Aspects of Molecular Medicine*, 2023, vol. 2, p. 1–11. DOI: 10.1016/j.amolm.2023.100022

[7] KÖNEMANN, H., ELLERMANN, C., ZEPPENFELD, K., et al. Management of ventricular arrhythmias worldwide: Comparison of the latest ESC, AHA/ACC/HRS, and CCS/CHRS guidelines. *JACC: Clinical Electrophysiology*, 2023, vol. 9, no. 5, p. 715–728. DOI: 10.1016/j.jacep.2022.12.008

[8] KIM, C. J., LEVER, N., COOPER, J. O. Antiarrhythmic drugs and anaesthesia, part 1: Mechanisms of cardiac arrhythmias. *BJA Education*, 2023, vol. 23, no. 1, p. 8–16. DOI: 10.1016/j.bjae.2022.11.001

[9] AL-SHAMMARY, D., NOAMAN KADHIM, M., MAHDI, A. M., et al. Efficient ECG classification based on Chi-square distance for arrhythmia detection. *Journal of Electronic Science and Technology*, 2024, vol. 22, no. 2, p. 1–13. DOI: 10.1016/j.jnlest.2024.100249

[10] RAI, H. M., YOO, J., DASHKEVYCH, S. GAN-SkipNet: A Solution for data imbalance in cardiac arrhythmia detection using electrocardiogram signals from a benchmark dataset. *Mathematics*, 2024, vol. 12, no. 17, p. 1–31. DOI: 10.3390/math12172693

[11] WU, Y., TANG, Q., ZHAN, W., et al. Res-BiANet: A hybrid deep learning model for arrhythmia detection based on PPG signal. *Electronics*, 2024, vol. 13, no. 3, p. 1–15. DOI: 10.3390/electronics13030665

[12] ISLAM, M. R., QARAQE, M., QARAQE, K., et al. CAT-Net: Convolution, attention, and transformer-based network for single-lead ECG arrhythmia classification. *Biomedical Signal Processing and Control*, 2024, vol. 93, p. 1–15. DOI: 10.1016/j.bspc.2024.106211

[13] AHMED, A. A., ALI, W., ABDULLAH, T. A. A., et al. Classifying cardiac arrhythmia from ECG signal using CNN deep learning model. *Mathematics*, 2023, vol. 11, no. 3, p. 1–16. DOI: 10.3390/math11030562

[14] SUN, J. Automatic cardiac arrhythmias classification using CNN and attention-based RNN network. *Healthcare Technology Letters*, 2023, vol. 10, no. 3, p. 53–61. DOI: 10.1049/htl2.12045

[15] CAÑÓN-CLAVIJO, R. E., MONTENEGRO-MARIN, C. E., GAONA-GARCIA, P. A., et al. IoT based system for heart monitoring and arrhythmia detection using machine learning. *Journal of Healthcare Engineering*, 2023, 6401673, p. 1–13. DOI: 10.1155/2023/6401673

[16] DHYANI, S., KUMAR, A., CHOUDHURY, S. Analysis of ECG-based arrhythmia detection system using machine learning. *Methods X*, 2023, vol. 10, p. 1–15. DOI: 10.1016/j.mex.2023.102195

[17] YANG, X., ZHONG, J. Automatic classification method of arrhythmias based on 12-lead electrocardiogram. *Sensors*, 2023, vol. 23, no. 9, p. 1–16. DOI: 10.3390/s23094372

[18] DAYDULO, Y. D., THAMINENI, B. L., DAWUD, A. A. Cardiac arrhythmia detection using deep learning approach and time frequency representation of ECG signals. *BMC Medical Informatics and Decision Making*, 2023, vol. 23, p. 1–14. DOI: 10.1186/s12911-023-02326-w

[19] FARAG, M. M. A tiny matched filter-based CNN for inter-patient ECG classification and arrhythmia detection at the edge. *Sensors*, 2023, vol. 23, no. 3, p. 1–23. DOI: 10.3390/s23031365

[20] SANTANDER BAÑOS, F., HERNÁNDEZ ROMERO, N., MORA, J. C. S. T., et al. A novel hybrid model based on convolutional neural network with particle swarm optimization algorithm for classification of cardiac arrhythmias. *IEEE Access*, 2023, vol. 11, p. 55515–55532. DOI: 10.1109/ACCESS.2023.3282315

[21] KIM, H. K., SUNWOO, M. H. An automated cardiac arrhythmia classification network for 45 arrhythmia classes using 12-lead electrocardiogram. *IEEE Access*, 2024, vol. 12, p. 44527–44538, DOI: 10.1109/ACCESS.2024.3380892

[22] DIN, S., QARAQE, M., MOURAD, O., et al. ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features. *Artificial Intelligence in Medicine*, 2024, vol. 150, p. 1–10. DOI: 10.1016/j.artmed.2024.102818

[23] MANICKAM, C., GOVINDASAMY, M., MUTHUSAMY, S., et al. A novel method for design and implementation of systolic associative cascaded variable leaky least mean square adaptive filter for denoising of ECG signals. *Wireless Personal Communications*, 2024, vol. 137, p. 1029–1043. DOI: 10.1007/s11277-024-11450-3

[24] SU, H., ZHANG, H., WU, Z., et al. Relaxed collaborative representation with low-rank and sparse matrix decomposition for hyperspectral anomaly detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, vol. 15, p. 6826–6842. DOI: 10.1109/JSTARS.2022.3193315

[25] MA, Y., LAN, Y., XIE, Y., et al. A spatial–spectral transformer for hyperspectral image classification based on global dependencies of multi-scale features. *Remote Sensing*, 2024, vol. 16, no. 2, p. 1–20. DOI: 10.3390/rs16020404

[26] LIU, Z., YAO, G., ZHANG, Q., et al. Wavelet scattering transform for ECG beat classification. *Computational and Mathematical Methods in Medicine*, 2020, p. 1–11. DOI: 10.1155/2020/3215681

[27] LIU, K., JIAO, Y., DU, C., et al. Driver stress detection using ultra-short-term HRV analysis under real world driving conditions. *Entropy*, 2023, vol. 25, no. 2, p. 1–13. DOI: 10.3390/e25020194

[28] KILMEN, S., BULUT, O. Scale abbreviation with recursive feature elimination and genetic algorithms: An illustration with the test emotions questionnaire. *Information*, 2023, vol. 14, no. 2, p. 1–13. DOI: 10.3390/info14020063

[29] LIU, B., CHEN, W., WANG, Z., et al. RAdam-DA-NLSTM: A nested LSTM-based time series prediction method for human–computer intelligent systems. *Electronics*, 2023, vol. 12, no. 14, p. 1–20. DOI: 10.3390/electronics12143084

[30] KANG, Q., CHEN, E. J., LI, Z. C., et al. Attention-based LSTM predictive model for the attitude and position of shield machine in tunnelling. *Underground Space*, 2023, vol. 13, p. 335–350. DOI: 10.1016/j.undsp.2023.05.006