International Journal of Theoretical & Applied Computational Intelligence
https://ijtaci.com

Research Article

ECG-based Heart Arrhythmia Recognition using Enhanced Extreme Learning Machine

Ashwaq Neaman Hassan*

Technical Collage of Management, Al-Furat Al-Awsat Technical University, Al-Najaf, Iraq
*Corresponding Author: Ashwaq Neaman Hassan. Email: ashwaqnama@atu.edu.iq
https://orcid.org/0000-0003-0539-6246

Received: 18/02/2026; Accepted: 25/03/2026; Published: 10/04/2026
https://doi.org/10.65278/IJTACI.2026.70

Abstract: Reliable classification of electrocardiogram (ECG) signal patterns is fundamental for enabling early diagnosis and effective clinical evaluation of cardiovascular abnormalities. In this paper, an innovative hybrid system is proposed for accurate classification of cardiac arrhythmias using ECG signals as input features. We present a deep learning-based method which utilizes the CNN model for complex features extraction, then integrates an Enhanced Extreme Learning Machine with Leave-One-Out (LOO) cross validation classifier to obtain both correct and stable models during training. In tests on the MIT–BIH Arrhythmia Database, the accuracy of CNN-EELM-LOO model was 99.81% and it exhibited improvement over classic machine learning and nowadays deep learning processes. The Leave-One-Out cross-validation technique integrated into decision model refinements provides an additional method to increase overall robustness of the model while reducing potential for random initialization effects or bias that can occur even with repeated training. This research demonstrates that the hybrid EELM-LOO architecture functions as an effective tool for automated arrhythmia detection, with strong potential for integration into clinical decision support systems.

Keywords: Hybrid CNN-ELM-LOO; ECG classification; patient-wise validation; Health risks

1. Introduction

The global death toll from cardiovascular diseases persists as the primary cause of death since timely accurate diagnostic tools help prevent associated health dangers [1]. ECG functions as an important non-invasive diagnosis tool by tracking heart electrical signals to discover different heart problems especially arrhythmias [2,3]. Electrocardiography (ECG) signal classification and analysis serves as a necessity to achieve proper patient care through rapid interventions [4]. Substantial improvements in computational methods have made automated ECG classification vital because it helps medical practitioners achieve a combination of better diagnosis and operational excellence [5]. The diagnostic quality depends heavily on medical signal processing because it enables precise reliable and efficient ECG signal interpretation [6]. Automated ECG classification systems gain increased importance because medical facilities now handle growing ECG data volumes that lead to difficulties with manual interpretation [7]. Correctly functioning classification software enables doctors to detect heart rhythm abnormalities swiftly resulting in better clinical results [8]. Complete automated cardiac surveillance depends on these systems which also lead to the invention of portable and real-time heart monitoring technologies critical for early diagnosis of dangerous heart conditions [9,10]. The present ECG classification methods encounter challenges caused by patient-specific variations and signal disturbances and ECG signal complexity in their diagnostic capabilities. Traditionally used machine learning techniques however require extensive feature engineering efforts in order to achieve generalization across diverse patient populations [11]. Recent works have shown that two deep learning models, namely Convolutional Neural Network (CNNs) and Recurrent Neural Network RNNs can be successfully applied to capture complex patterns in the ECG data. Such systems may be operational in hospital settings, but demand a considerable amount of resources and large amounts of labeled data which may not be available to hospitals [12].

Here, the study applies a new architecture through combining CNN deep feature extraction with Enhanced Extreme Learning Machine based classification under Leave-One-Out (LOO). The study aims to construct an appropriate predictive ECG signal classification model that is able to provide high accuracy with prediction stability using the Leave-One-Out method. This study tackles the constraints in this field to provide a novel solution that advances prevailing methods for identification and categorization of arrhythmias. This study provides the following key contributions:

- Hybrid CNN-ELM-LOO architecture
- Uncertainty-based two-stage classification
- Optimized ensemble learning strategy
- Improved generalization using patient-wise validation.

The proposed framework uses a CNN network with EELM-LOO for both improved feature extraction and accurate arrhythmia detection: This research presents an innovative hybrid structure which merges a CNN feature extraction component for ECG signal analysis with an Enhanced Extreme Learning Machine escorted by Leave-One-Out classification block. The EELM-LOO classifier achieves better test data accuracy by processing features extracted through this method.

The implementation of Leave-One-Out within Extreme Learning Machine (ELM) improves both system stability and minimizes sensitivity to random weight initializations: The proposed method strengthens ELM by applying Leave-One-Out (LOO) to its training phase. The efficiency of performance becomes more stable and model simplicity increases while classification accuracy enhances and reliability increases during processing of non-stationary ECG signals when the Leave-One-Out strategy is integrated with Extreme Learning Machines.

The proposed approach uses a two-stage cascading classification structure that handles uncertain samples: The proposed system implements a two-stage cascading classification model that splits uncertain cases from normal processing to conduct additional analysis with another decision unit. The hierarchical structure enables re-evaluation processes which produce more accurate decisions to improve model accuracy and stability and reliability in complex and ambiguous condition.

2. Related Works

Alsayat et al., [13] investigates deep learning models as a tool for ECG image classification to assist medical diagnosis of cardiac conditions. The authors executed cross-validation tests on four ECG categories through transfer learning of eight CNN architectures from ImageNet. They measured performance using F1-score and balanced accuracy metrics. The analysis in stage two involved evaluating 120 combined versions of these models. The Inception structure along with MobileNet and NASNetLarge produced the highest scoring ensemble which delivered 0.9651 F1-score and 0.9640 balanced accuracy results. ROC analysis showed near-perfect AUCs. The research demonstrates that ensemble deep learning techniques show promise for achieving better performance in cardiac diagnostics through ECG analysis. Selvam et al., [14] implemented a CNN-VAE hybrid approach for improving CVD detection using ECG data in their work. The CNN-VAE model acquired training from PTB-XL database which contains more than 21 thousand 12-lead ECG recordings to perform assessments across five super-classes and 23 sub-classes of cardiovascular diseases. The developed technique demonstrated superior performance by achieving 98.51% accuracy together with 97.95% F1-score and extremely low false positive and false negative rate measurements. Research reveals that the designed hybrid model outperformed its counterparts among different deep learning methods. The architecture enables ECG interpretation at a strong level and presents potential to assist clinical professionals with timely and precise disease identification.

ResNet architects created a 34-layer variant for cardiovascular disease (CVD) type classification tasks through the application of scalogram-based time-frequency features in [15]. Transfer learning forms a core element of this model which boosts its classification capabilities and the research compares the deep learning architecture against three different neural network approaches consisting of two CNNs and RNNs. Performance evaluation compared this model against SNMF when employed with its sparse variant. ResNet-34 consistently reaches better accuracy with higher sensitivity and robustness than other approaches in detecting CVDs thus indicating its potential for clinical CVD diagnosis. Singh et al., [16] evaluated ECG signal classification by applying Recurrent Convolutional Neural Networks (RCNN) coupled with Grey Wolf Optimization (GWO). For training and evaluation, the MIT-BIH Arrhythmia databases and PTB Diagnostic ECG are utilized as part of this investigation. The proposed combination of RCNN and GWO generated higher performance outcomes with 98% accuracy surpassing K-Nearest Neighbors, Logistic Regression, Decision Tree, Support Vector Machines and Random Forest. The benchmark models surpassed 90% accuracy but the RCNN-GWO method outperformed them through efficient classification that demonstrates its effectiveness in ECG cardiac diagnosis systems.

Kolhar et al., [17] conducted ECG signal classification through PTB Diagnostic ECG Database by implementing artificial intelligence models which combined AlexNet with a dual-branch fusion network. AlexNet accomplished 98.64% validation accuracy along with 99% test accuracy after standardization and balancing and model reshaping while the dual-branch model obtained 99% test accuracy. The implemented models obtained superior accuracy results when compared to both Hybrid AlexNet-SVM and DCNN-LSTM methods by generating enhanced precision, sensitivity, and specificity numbers. Deep learning-based ECG analysis proved effective in the study results which backs up its potential for developing automated diagnostic systems that work in resource-restricted healthcare facilities.

Xiao and Wang [18] presented Reinforcement Learning-based Wavelet Base Selection (RLWBS). The new analytical techniques of RLWBS yielded more complete time-frequency data than traditional wavelet selection methods did. The adaptive method enabled better diagnostic accuracy when classifying ECG abnormalities due to its application of reinforcement learning in medical signal preprocessing optimization.

Similarly, Issa et al., [19] produced a deep neural network model with residual blocks (DNN-RB) to categorize ECG cardiac cycles into six distinct beat categories. The algorithm reached a 99.51% accuracy as measured by its average specificity and sensitivity amounts of 99.7% and 98.2% respectively. The DNN-RB model achieved superior performance than other algorithms for the same dataset. This method exists to work in both specialized medical locations and distant locations and has been designed for integration with portable ECG tools using one ECG lead. A web-based application helped process ECG inputs while delivering diagnostic outputs through a system that improved its utilization in practical healthcare settings. Cao et al., [20] developed a deep transfer learning framework that showed effectiveness for ECG classification especially when dealing with limited training data. This method applies ResNet-18 as a general-purpose image classifier while using the MIT-BIH arrhythmia database which meets requirements from AAMI EC57 standards for testing. The research investigates the issue of data leakage that surfaced in previous deep learning models that violated AAMI guidelines. This examination demonstrated that varied splitting procedures appreciably modifies model performance evaluations thus researchers must use AAMI EC57 standards during arrhythmia classification research that leverages the MIT-BIH data set.

Gour, et al., [21] presented an extensive analysis of different machine learning and ensemble learning models which include SVM, CNN, XGBoost and other algorithms for ECG classification using MIT-BIH and PTB-XL datasets. The accuracy average of CNN reached 98.83% across MIT-BIH and PTB-XL while LSTM performed best in sensitivity specifically on the PTB-XL with 88.01% recall average. SVM and k-NN achieved similar levels of success on MIT-BIH ECG recordings with 96.20% accuracy followed by 97.50% success rate. The research analyzed clustering approaches GMM and K-Means because they showed promise for heart disease diagnosis classification. This paper completes essential research by creating a comparative study of tested methods. Zabihi et al., [22] developed a hybrid system to recognize ECG heartbeats as normal or one of three abnormal types: supraventricular, ventricular and fusion. The procedure implements machine learning in conjunction with deep learning capabilities among two interconnected systems. The first subsystem integrates residual network blocks to extract features which are later classified by an LSTM network whereas the second subsystem performs classification using both random forest and different feature extraction methods. Using the Synthetic Minority Over-Sampling Technique (SMOTE) enables the system to handle the problem of unbalanced training data classes. The predictive model demonstrated an accuracy rate of 99.26% thus becoming one of the best systems for detecting arrhythmias.

3. Paper organization

This work aims to create a powerful and rapid technique in ECG signal categorization through deep learning models. Electrocardiogram (ECG) signals are among the most EO important non-invasive methods to monitor and analyze cardiac function, as they embody essential information of the heart's electrical activity. However, during acquisition these signals are corrupted by noise, for example from muscle artifacts, baseline wander, and power-line disturbance which may obscure accurate analysis. So for detecting arrhythmias or any other abnormal patterns the role of proper signal processing algorithms is important not only for denoising purpose but also used to extract good features. The leak method consists of a three-stage pipeline for arrhythmia classification. During the signal processing step, discrete wavelet transform (DWT) is used to filter out noise. Making it superior in noise removal because DWT has the ability to work being time and frequency components, therefore retains the structure of all signal. After denoising, the ECG signals are divided into uniform temporal windows to allow local analysis and prepare for processing steps that will follow. Next, the preprocessed segments of these signals are supplied to the CNN. The CNN, with its layered design and learnable filters can obtain higher level, high-level abstract features from the input data. These features are differentiable descriptors that account not only for temporal dynamics but also for morphological patterns of the heartbeats, enabling discrimination between normal and abnormal beats [23]. At the last stage, feature vectors produced from CNN are fed to the proposed classifier an Enhanced Extreme Learning Machine (EELM) combined with a Leave-One-Out method. The EELM-LOO demonstrates the ability to increase classification accuracy and generalization by optimizing network parameters and explicitly removing single data points for each training iteration. A novel approach is introduced for automated arrhythmia detection and classification system involving preprocessing, deep feature extraction and optimized classification that may assist the clinical decision making. Figure 1 presents an overview of the proposed method and its overall framework.

The first experiment used a 70/30 split for initial training and independent testing. Tenfold cross validation was also executed for the model optimization. Moreover, one-element deletion (LOO) verification in the ELM classifier was used to estimate uncertainty. You can structure the mechanism of evaluation as: Feature extraction using a convolutional neural network (CNN).

- Model training using ten-fold cross-validation.
- Final evaluation using a 70/30 independent split.
- Uncertainty estimation using LOO within the ELM classifier.

3.1 Preprocessing

Electrocardiogram (ECG) signals are often contaminated by various types of noise, which can significantly reduce the signal-to-noise ratio (SNR). The primary objective of the preprocessing stage is to eliminate noise and artifacts from the ECG signal as much as possible, thereby enhancing the SNR. This step is critical to ensuring optimal performance in arrhythmia classification systems. ECG signals can be influenced by various interferences and artifacts, including baseline drift, muscular artifacts, power-line noise, contact noise, electrode movement, and electromyographic interference. In this study, preprocessing is divided into two main stages. Initially, noise reduction is performed using a Discrete Wavelet Transform (DWT). Second, the ECG signal undergoes segmentation, the details of which are discussed in the subsequent sections.

3.1.1 DWT-Based Noise Reduction

The Discrete Wavelet Transform (DWT) is a powerful signal processing method, particularly suitable for the denosing of the ECG signals due to its non-stationary nature. Using wavelet decomposition, a signal can be represented at different scales for localized time and frequency domain analysis. At each level of decomposition, the signal is filtered with low-pass and high-pass filters, after which it is down sampled. The detail coefficients = D1 are the output of high-pass filter and approximation coefficients = A1 are the output of low-pass filter. In the current work, ECG data are decomposed in four levels using Daubechies 6 (db6) wavelet function. The first two levels of detail coefficients frequently contain most of the noise and are, therefore, omitted in communication reconstruction for improved signal quality. We choose the db6 wavelet because of its structural similarity with QRS complex in ECG signals. An illustration in Figure 2. This figure shows an example heartbeat waveform before and after using this denoising method.

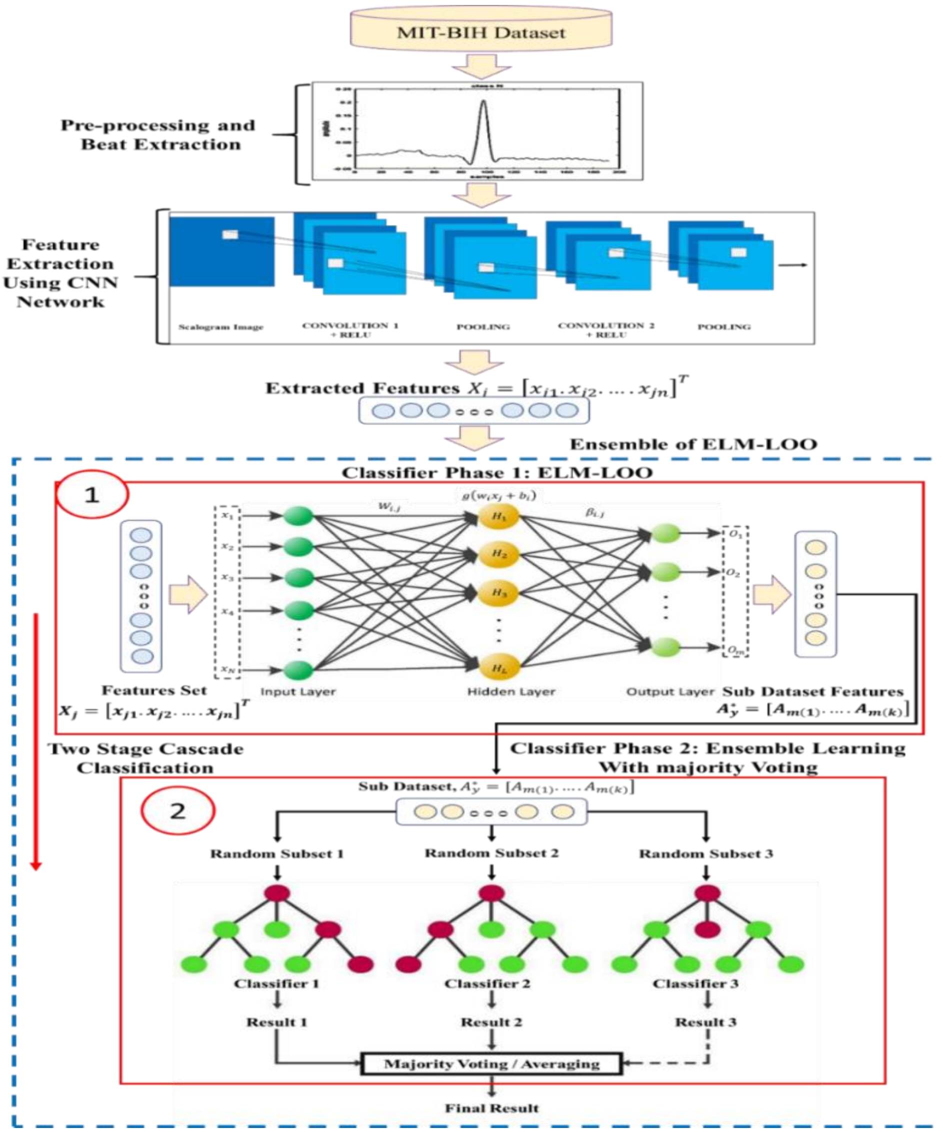

Figure 1: Diagram of the Proposed Method

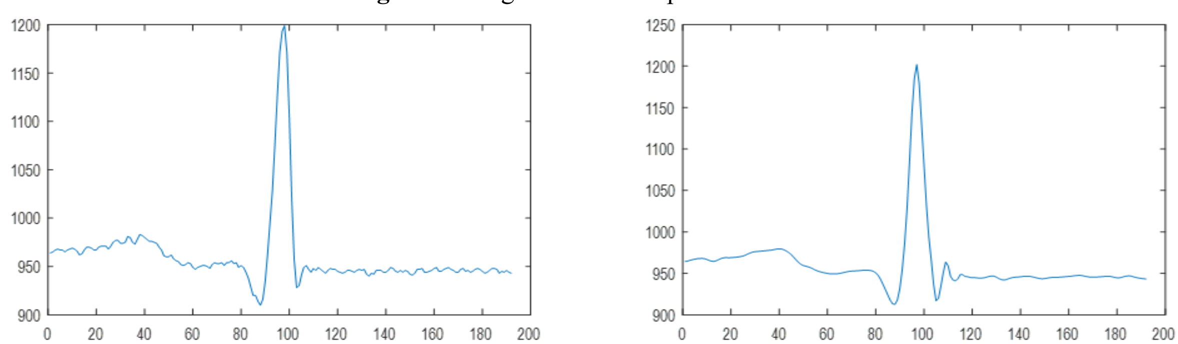

Figure 2: Example of extracted heartbeat. left: noisy signal; right: denoised signal.

3.1.2 ECG Segmentation

ECG recordings generally consist of several heartbeats that need to be detected and segmented to enable accurate classification into predefined categories [24]. Following the denoising process, R-peaks are detected within the ECG signal and used to segment individual heartbeats. Around each detected R-peak, a segment comprising 200 samples is extracted 100 samples preceding and 99 samples following the peak resulting in a window centered on the R-peak. An example of a normal heartbeat obtained through this segmentation technique is illustrated in Figure 3.

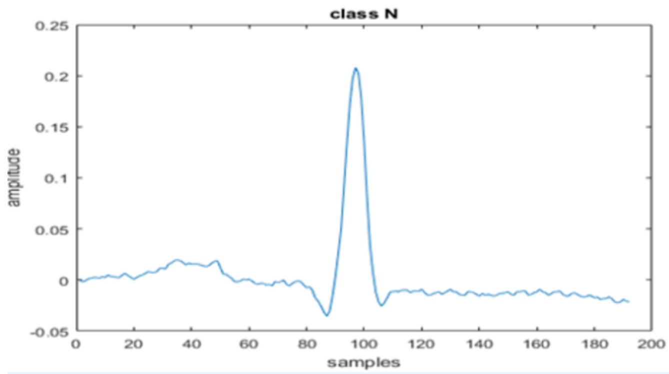

Figure 3: A typical normal heartbeat from an electrocardiogram signal.

3.2 Feature Extraction Using CNNs

After processing the ECG signal segments, they are input into a CNN for feature extraction. The CNN architecture is shown in Figure 4. Additionally, Table 1 displays the construction of the convolutional, pooling, and activation layers and parameters.

Table 1: Size of CNN layers used in the simulation.

Layer	Type	Filters	Kernel Size	Stride	Activation	output
1	Conv1	32	5	1	ReLU	Feature Map
2	Max Poolnig	-	2	2	-	Reduced
3	Conv2	64	3	1	ReLU	Feature Map
4	Max Poolnig	-	2	2	-	Reduced
5	Conv3	128	3	1	ReLU	Feature Map
6	Max Poolnig	-	2	2	-	Reduced
7	Flatten	-	-	-	-	Vector
8	FC1	256	-	-	ReLU	Dense
9	FC2	128	-	-	ReLU	Dense
10	FC3	100	-	-	-	Feature Vector

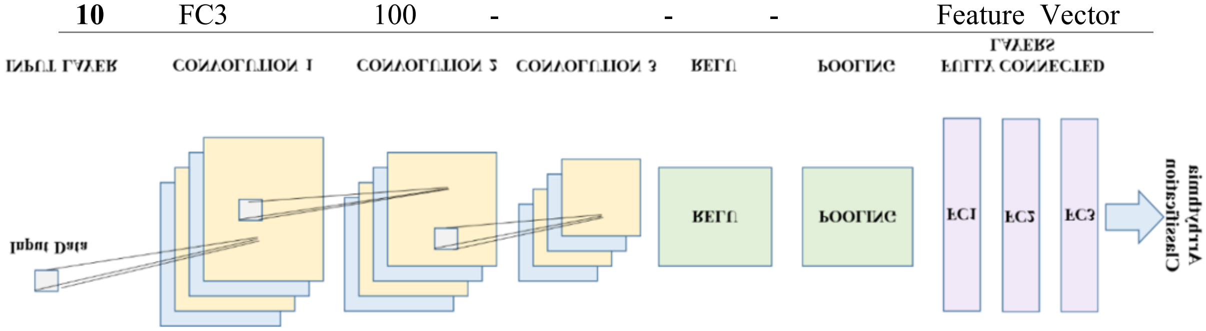

Figure 4: Structure of the proposed CNN.

The CNN was trained using the Adam optimizer with an initial learning rate of 0.001, batch size of 128, and a maximum of 500 epochs. Early stopping with a patience of 20 epochs was used to prevent overfitting.

The last fully connected layer produces 100 deep features, which are used as inputs for the ELM-LOO classifier.

3.3 Two-Stage Classification Using ELM-LOO (CNN-EELM-LOO)

In this study, an enhanced ensemble learning algorithm. The EELM-LOO is used for the classification of electrocardiogram (ECG) signals. This model is a developed version of the ELM, which is known for its simple structure, fast training speed, and reasonable accuracy on relatively low-complexity data. However, a major limitation of ELM is its high sensitivity to the random initialization of hidden layer weights and biases, which can lead to instability and significant fluctuations in performance across different runs. Furthermore, ELM's simple architecture may limit its generalization capability, particularly when dealing with complex or highly noisy datasets. To address these issues, an ensemble structure is introduced in the form of EELM-LOO. This approach not only compensates for the shortcomings of the basic ELM algorithm but also leverages its primary advantages, such as high-speed training and analytical solution formulation.

3.3.1 Extreme Learning Machine (ELM) Algorithm

The ELM is a single-hidden-layer feedforward neural network in which the input weights and hidden biases are randomly assigned and are not updated during training. Output weights are calculated analytically only. This architecture enables extremely fast learning speeds while maintaining acceptable performance. The main advantage of ELM lies in its random selection of hidden layer parameters, reducing the training to a straightforward estimation of output weights. Figure 5 illustrates the overall structure of the ELM.

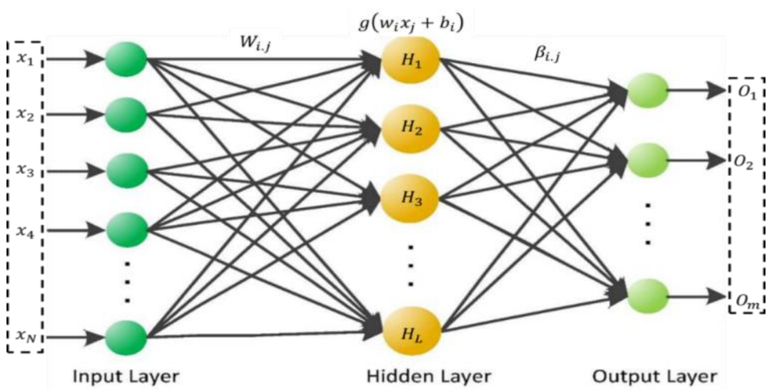

Figure 5: General structure of the ELM.

The activation function g(x), given a training set $\{(\boldsymbol{x_j}.\boldsymbol{t_j})\}_{\boldsymbol{j=1}}^{\boldsymbol{N}}$ with N samples and m classes, defines the ELM output with L hidden neurons as:

$$\sum_{i=1}^{L} \beta_i g_i\left(x_j\right) = \sum_{i=1}^{L} \beta_i g\left(w_i x_j + b_i\right) = o_j. \qquad j = 1.2\ldots\ldots N \tag{1}$$

In this equation, the input vector is denoted by $X_j = [x_{j1}.x_{j2}.\ldots.x_{jn}]^T$, the connection weights between the i-th hidden neuron and the input neurons are $w_i = [w_{i1}.w_{i2}.\ldots.w_{in}]^T$, The desired outcome of the output is $t_j = [t_{j1}.t_{j2}.\ldots.t_{jm}]^T$, and b_i is the bias of the i-th hidden neuron. Additionally, $\beta_j = [\beta_{j1}.\beta_{j2}.\ldots.\beta_{jm}]^T$ denotes the weights connecting the i-th hidden neuron and the output layer, while o_j represents the actual network output for the input X_j. Randomly, the hidden parameters $\{w_i.b_i\}$ can be initialized without being tuned throughout training. The goal of ELM is to solve the following consolidated model that minimizes the error between o_j and t_j:

$$\min_{\beta}\|H\beta - T\| \quad (2)$$

Where:

$$H(w_1.\dots.w_L.b_1.\dots.b_L) = \begin{bmatrix} g(w_1x_1+b_1) & g(w_Lx_1+b_L) \\ g(w_1x_N+b_1) & g(w_Lx_N+b_L) \end{bmatrix}.\beta = \begin{bmatrix} \beta_1^T \\ \dots \\ \beta_L^T \end{bmatrix}.T = \begin{bmatrix} T_1^T \\ \dots \\ T_N^T \end{bmatrix} \quad (3)$$

In this context, β is the output weight matrix and H is the hidden layer output matrix. The problem of minimizing least squares is represented in equation (2), and its solution can be obtained using $\hat{\beta} = H^{\dagger}T$. Here, $H^{\dagger}$represents the pseudoinverse of the matrix H. The actual output is calculated for the input data, then the predicted label for each class is determined based on the maximum value index in the output vector**.** Equation (4) applies a regularized least squares method to optimize the solution:

$$\min_{\beta}\|H\beta - T\|_F + \frac{1}{\lambda}\|\beta\|_F \quad (4)$$

In this equation, λ is the regularization parameter used to balance the regularization term and the training error. Therefore, the optimization solution to Equation (4) is determined as follows:

$$\text{If } L \leq N \qquad \beta = (H^TH + \lambda I)^{-1}H^TT \quad (5)$$

$$\text{If } L \geq N \qquad \beta = H^T(HH^T + \lambda I)^{-1}T \quad (6)$$

In these equations, I denotes the identity matrix.

3.3.2 ELM-LOO Algorithm

Optimal tuning of the regularization parameter in the ELM algorithm is crucial for improving the model's generalization capability. In this paper, to select the most appropriate value for this parameter, the Leave-One-Out Cross-Validation (LOO) method is employed. In the field of machine learning, this approach is regarded as one of the most precise and effective methods for choosing models and optimizing parameters. When the database contains N samples, the core idea of the LOO method is to split the data into N training subsets such that, in each iteration, one sample is left out while the model is trained on the remaining data. The model's efficacy and generalization capacity are then assessed using the excluded sample. To minimize the computational complexity of the LOO validation process, this study utilizes the Predicted Residual Sum of Squares (PRESS) statistic. The PRESS program provides a basis for computing the Mean Squared Error (MSE) within the LOO framework. The PRESS-based MSE, denoted as MSE^{PRESS}, is defined as follows:

$$MSE^{PRESS} = \frac{1}{N}\sum_{J=1}^{N}\left(\frac{t_j - o_j}{1 - HAT_{jj}}\right)^2 \quad (7)$$

In this equation, o_j signifies the actual output of the network for the sample j, t_j is the target output (label). The j-th diagonal element of the HAT matrix, represented as HAT_{jj}, is described as follows:

$$HAT = HH^{\dagger} = H(H^TH)^{-1}H^T \quad (8)$$

As specified in Equation (3), H refers to the output matrix of the hidden layer. To determine the best value for the regularization parameter λ, various candidate values are selected from a search space. For each value, $MSE^{PRESS}$is computed, and the value of λ that results in the lowest error is chosen as the optimal regularization parameter. To further reduce computational overhead during this search, the PRESS statistic is used without the need to repeatedly recompute the pseudoinverse of H for each λ.

To analyze the output matrix of the hidden layer H, Single Value Decomposition (SVD) is utilized, thereby optimizing the process. Assuming the matrix H is decomposed as $H = UDV^T$, where D denotes a diagonal matrix comprising the singular values of H, and U and V are orthogonal matrices, the required analytical expressions can be derived as follows. If $L \leq N$, HAT_r can be found as: $HAT_r = H(H^T H + \lambda I)^{-1} H^T$ and then $HAT_r = (VDU\wedge T\ UDV\wedge T + \lambda I)^{-1} H^T$.

$$HAT_r = (D\wedge 2 + \lambda I)^{-1} V^T H^T \quad (9)$$

According to Equations (5) and (9), the output of the ELM can then be described as follows:

$$V(D^2 + \lambda I)^{-1} V^T H^T T \quad (10)$$

Additionally, the diagonal elements of HAT_r can be directly computed by summing the rows of the following matrix:

$$(HV(D^2 + \lambda I)^{-1}) \odot (HV) \quad (11)$$

In this expression, $\odot$ denotes the Hadamard (element-wise) product of two matrices. Since Equations (9) and (3) are independent of U, instead of computing $H = UDV^T$, it is sufficient to decompose $H^T H = VD^2 V^2$ directly. To improve computational efficiency, it is important to note that many matrices used in the computations, such as HV and $V^T H^T T$, are independent of the regularization parameter λ. Therefore, these matrices can be precomputed and stored before the MSE^{PRESS} calculation process begins. This significantly reduces computation time and improves the efficiency of the parameter search.

After determining the optimal value of the regularization parameter, denoted by λ_{opt}, the optimal output weight matrix, denoted by $\hat{\beta}$, is computed using the following relationship:

$$\hat{\beta} = (H^T H + \lambda_{opt} I)^{-1} H^T T, \quad \hat{\beta} = V(D^2 + \lambda_{opt} I)^{-1} V^T H^T T \quad (12)$$

If $L > N$, then HAT_r matrix is calculated as:

$$HAT_r = HH^T (HH^T + \lambda I)^{-1}, HAT_r = (UDV^T VDU^T + \lambda I)^{-1}, \ HAT_r = HH^T U(D^2 + \lambda I)^{-1} U^T \quad (13)$$

Since Equation (12) is independent of V, the matrix HH^T can be decomposed as $HH^T = UD^2 U^T$. In this case, the weight matrix of the output $\hat{\beta}$, corresponding to the optimal regularization parameter λ_{opt} is calculated as:

$$\hat{\beta} = H^T (HH^T + \lambda_{opt} I)^{-1} T, \quad \hat{\beta} = H^T U(D^2 + \lambda_{opt} I)^{-1} U^T T \quad (14)$$

3.3.3 EELM-LOO Algorithm

The hidden layer weights used in the ELM-LOO classifier were initialized randomly, which may lead to significant variability in performance between different runs and results that could lack stability. Moreover, for more intricate or noisy data, ELM-LOO may not possess enough capacity to characterize complex patterns with its simple structure, resulting in weaker generalization ability on test data. Ensuring sufficient robustness and stability for these applications is a prominent limitation, however, indicating the need to develop more optimized forms or hybrid models combining other learning approaches. To address these issues, an enhanced two-stage algorithm EELM-LOO (Ensemble ELM-LOO) is proposed, designed to correct classification uncertainties introduced by the ELM-LOO algorithm. In this approach, classification is initially performed using ELM-LOO. Then, samples for which the model exhibits uncertainty in the output vector are passed to a second phase for classification via an ensemble learning method. Specifically, if the difference between the top two output scores of the ELM-LOO network denoted by T_{diff} is less than a predefined threshold α, the classification is considered uncertain, and the second phase of classification is triggered. To solve a classification issue containing m classes, let the output of the ELM-LOO network be represented as, $O = \{o_1, \ldots, o_m\}$. The predicted class for a given test input y is determined as:

$$ID(y) = \arg\max_i (o_i) \quad (15)$$

In a typical classification process, the class with the highest output value is selected as the predicted class. However, in the EELM-LOO framework, both the maximum and the second-highest output values are considered. The difference between them T_{diff} is calculated. If $T_{diff} > \alpha$, it indicates a confident prediction, and the ELM-LOO output is accepted. Conversely, if $T_{diff} < \alpha$, the prediction is deemed uncertain, and a complementary ensemble learning-based method is employed to enhance classification accuracy. In the proposed method, instead of relying solely on the highest output from the ELM-LOO model, the top k output values are used. Accordingly, training samples from the original dataset that belong to the corresponding top-k classes are extracted to form a sub-dataset $A_y^* = [A_{m(1)}. \ldots. A_{m(k)}]$, where $\mathrm{m(i)} \in \{1.2. \ldots. \mathrm{m}\}\}$ represents all training samples associated with class $\mathrm{m(i)}$. This sub-dataset is then used to train an ensemble learning algorithm in the second phase. For the secondary classification, the Bagging ensemble method is adopted. An ensemble of base classifiers $\{h_1. h_2. \ldots. h_T\}$ is constructed using random sampling (with replacement) from A_y^*. Each sampled subset is used to train a separate base classifier. The final class label for the test sample y is determined through majority voting among the outputs of the base classifiers, expressed as:

$$(y) = mode(\{h_t(y)\}_{t=1}^{T}) \qquad (16)$$

Here, $C(y)$ represents the predicted label for a test sample y, while gives us back the predicted class that occurs most frequently. Bagging simplifies the ensemble framework proposed in this study, and it serves to lower variance error by instilling diversity between the base classifiers and combining their outputs. By focusing on only the cs top k most probable classes corresponding to ELM-LOO, an improvement in decision-making reliability is achieved when it comes to samples that are situated near class boundaries or tend to have ambiguous characteristics.

4. Experimental Results

The experimental results study the efficiency of the approach, inside this section certain well-accepted performance metrics verify both the gained performance over common methods and benchmarking with state-of-the-art techniques. Implementation is done in MATLAB (version 2024a) on NVIDIA GPU with at least 12 GB memory. Experiments use all of the 25,000 ECG signals (5,000 samples per class) 70% of the data corresponding to 17,500 signals are designated for training the CNN and the remaining 7,500 signals for testing. The model will classify each ECG sample corresponding to 100 features that were extracted via CNN using some techniques ELM, ELM-LOO, and the improved version of EELM-LOO proposed in this study. To assess classification performance, we employ 10-fold cross-validation: the data is divided into ten subsets, each subset being used once as a validation fold, while nine folds are used for training. The whole dataset is randomly shuffled before splitting to avoid bias. The results of the experiments are displayed in tables that indicate average values computed over 50 independent executions to enhance reliability and to reduce the effect of random fluctuations. The ELM classifier was trained using 300 hidden neurons and a sigmoid activation function, and the organization parameter was optimized via cross-validation. To train the model we used Leave-One-Out (LOO).

4.1 Database

The Data utilized for the study is obtained from MIT-BIH Cardiac Arrhythmia Database [25], which consists of ECG recordings from 47 subjects, consisting of 25 males (age range: (32- 89) years), and also 22 women (age range: (23 - 89) years), nearly 60% of whom were hospitalized. Since anatomical difference always exists among subjects, ECG signals were recorded using two different leads; namely Lead II and Lead V which were sampled at 360 Hz over 24 hours monitoring session. It contains 48 segments transient for 30 minutes. The beat types are also classified, while the masses of beats have interpreters for most beads along diagnostic cues and R-peaks positions. In total, the database contains110 ,000 annotated heartbeats. In the original MIT-BIH Arrhythmia Database, classes are highly imbalanced and classification models may be biased to majority classes. To remedy this situation, we built a balanced subset of 25,000 beats in ECG (5,000 samples per class). Stratified sampling technique was used as balancing technique where we under sampled subclasses according to their frequency in the majority class while avoiding biased samples of the minority class. Furthermore, sampling was performed at the patient level to avoid data leakage. The different beat categories are represented in the dataset, as outlined in Table 2.

Table 2: Summary of 5 Different Classes In MIT-BIH

ANSI/AAMI classes	**Non-ectopic (N)**	**Supra-ventricular (S)**	**Ventricular (V)**	**Fusion (F)**	**Unknown (U)**
	Normal (N)	Aberrated atrial premature (A)	Ventricular escape (V)	Fusion of ventricular and normal (F)	Unclassifiable (U)
	Left bundle branch block (LBBB)	Atrial premature (a)	Premature ventricular contraction (E)		Paced (p)
MIT-BIH classes	Right bundle branch block (RBBB)	Supraventricular premature (S)			Fusion of paced and normal (f)
	Nodal (junctional) escape (j)	Nodal (junctional) premature (J)			
	Atrial escape beat (e)				

Per AAMI recs we adopted the patient data segregation mechanism. Patient model was used to build training and testing datasets. In addition, patients in the training and testing arms were fully segregated. This change dramatically improves the generalization performance while also reducing the potential for data leakage.

4.2 Evaluation Metrics style

In this paper, Accuracy, Precise, Recall, and F1-score are used to evaluate the classification performance and are calculated as follows [26-28]:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{17}$$

Precision (P) quantifies the proportion of correctly identified relevant instances among all instances predicted as positive. It is calculated using the following equation:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

Recall (R) represents the fraction of actual relevant instances that are successfully identified. It is calculated as below:

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

The F1 score is a commonly used metric for evaluation, especially when precision (P) and recall (R) offer potentially conflicting insights. It is computed by calculating the harmonic mean of precision and recall, as shown in the following equation:

$$F1 - score = \frac{2PR}{(P+R)} \tag{20}$$

Where: True Positive (TP): Number of correctly predicted positive instances, False Positive (FP):

Number of incorrectly predicted positive instances, False Negative (FN): Number of incorrectly predicted negative instances, and True Negative (TN): Number of correctly predicted negative instances.

4.3 Evaluating the Training Process

The loss function when training a neural network is illustrated here in Figure 6. On this curve, the y (vertical) axis shows the value of the loss for each step in training, and the x (horizontal) axis indicates number of iterations. As shown in the plot, loss function is relatively high at earlier time step and starts coming down gradually with increasing iterations. This gradual and consistent decrease shows that the model is converging correctly, and its capability of fitting training data is increasing progressively. From the beginning of the training up to about 400 iterations, we observe a steep negative slope on the loss curve due to faster learning capacity of model for more generalized and basic features. Past this point, the rate in losses dropped reduced appears slower and by about iteration 1500, we can see the curve stabilizes possibly indicating that the model converged to a sweet spot. This behavior suggests that the parameters controlling the network have gradually converged to their optimum values which minimize the prediction error on training data. In addition, the small variation of the loss curve as learning progresses suggests stable training characterized by an absence of oscillatory or divergent behaviour.

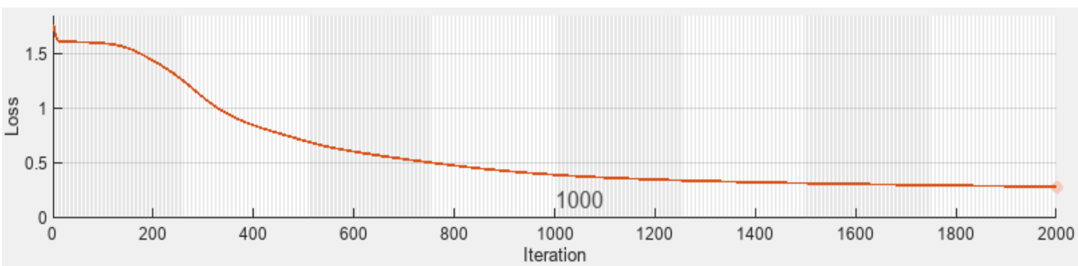

Figure 6: Loss curve of the neural network during training.

This stability results from the good design of the network architecture, a well-chosen learning rate, and appropriate preprocessing and ordering of training data. According to Figure 6, the training of proposed neural network had been successful and model has learned meaningful features from ECG signals which leaded to a proper classification of arrhythmias.

4.4 Evaluation of the Proposed Approach and Comparative Analysis

The confusion matrix in Figure 7 presents the performance of the proposed model for ECG signals based cardiac arrhythmias classification. In this matrix, rows contain the actual classes and columns represent the classes that were predicted by the model. The classification comprises five unique categories of arrhythmias termed as F, N, S, U and V, respectively. A detailed disaggregation from the confusion matrix demonstrates that the suggested architecture provides an exceptional rate of classification on separate classes. For class 159, there are in total two samples (both correctly classified as fatigue), thus has an accuracy of 100% Calculation U: All samples, 1527/1527 correct. For class S we have 1508 samples out of 1509 are predicted correctly, so the accuracy for that class is 99.9%. V class is also classified correctly in 1483 out of 1484 having 99.9 % accuracy. The classification accuracy is 99.6% for classes F and N. These results unambiguously show the capability of the proposed model about accurately detecting different classes of arrhythmia incorrectly. Moreover, the strong performance in each of the classes is indicative of a uniformly well-balanced learning process whereby all relevant features that were able to distinguish between classes were captured during training on ECG signals. The Receiver Operating Characteristic (ROC) curve obtained by assessing the performance of the proposed method in diagnosing cardiac arrhythmias is shown in Figure 8. This curve shows a plot of the True Positive Rate (TPR) vs False Positive Rate (FPR) based on multiple decision thresholds, thereby evaluating the model's ability to separate different classes. The area under the curve (AUC) is one of the most important measures obtained from the ROC curve and determines the overall accuracy of a model used for classification problem. The method showed excellent capability in discriminating between positive and negative instances with an AUC value of 99.93% associated with a very low error of prediction. Setting the ROC knee point is one of important things for ROC analysis. This point usually indicates the best compromise between sensitivity (also referred to as True Positive Rate, TPR) and specificity; therefore a decision threshold located at this position maximizes model overall accuracy.

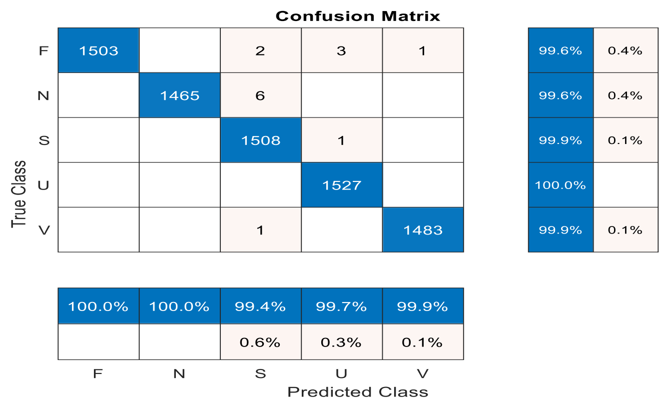

Figure 7: Confusion matrix of the proposed method.

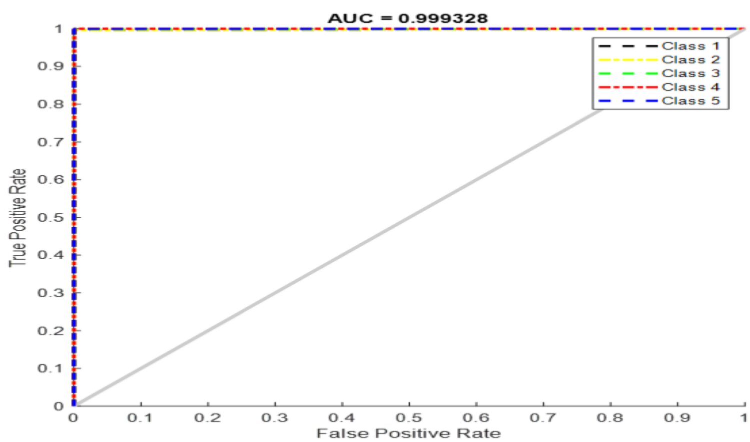

Figure 8: Loss The ROC curve of the proposed method.

As we can see from the accessed ROC curve, this location of knee point is prominent at the up left side of plot for maximum true positive rate with least false positives. The placement of this point in such a prime area points out the considerable strength and solid generalization power of the applied model when tested on data. These characteristics confirmed the potential of proposed method for achieving not only low false positive and negative rates but also maintaining even more precise trade-off between recall and accuracy when desired, making it reliable solution for numerous sensitive domains like medical diagnostics.

By definition of the F-score as the harmonic mean of precision and recall, analyzing these two metrics will give you an idea of how well your model balances correct detections with false positives. Unlike accuracy, which can lead to misleading interpretations when datasets are unbalanced, the F-score is an indicator of the reliability of actual model performance in complex scenarios such as arrhythmia classification. As depicted in Figure 9, different methods demonstrate varying levels of performance. For instance, the CNN-based approach achieves an F-score of 95.19%, while Forest and SVM methods reach 93.89% and 95.11%, respectively. Other machine learning and deep learning techniques also perform reasonably well, with F-scores of 84.71% and 88.21%, respectively. However, the hybrid method combining machine learning and deep learning (HMLDL) improves performance to an F-score of 97.03%. Remarkably, the proposed CNN-EELM-LOO method achieves an F-score of 99.81%, surpassing all previously mentioned approaches. Moreover, examining the intermediate stages of the proposed model's development (i.e., CNN-ELM with an F-score of 90.82% and CNN-ELM-LOO with 94.50%) reflects a consistent and effective improvement in network design.

This increasing trend indicates that the integration of the improved EELM structure with the Leave-One-Out strategy has notably enhanced the model's capacity to differentiate between classes with greater precision and recall. In conclusion, the proposed method not only excels in accurately identifying arrhythmias but also achieves superior balance in classification performance, making it highly suitable for clinical and diagnostic applications.

In Table 3, we provide an extensive performance comparison of our proposed framework with existing studies in the field of rhythmic and arrhythmic detection based on Accuracy metric. As stated in the table, traditional Machin Learning and Deep Learning methods have shown high performance.

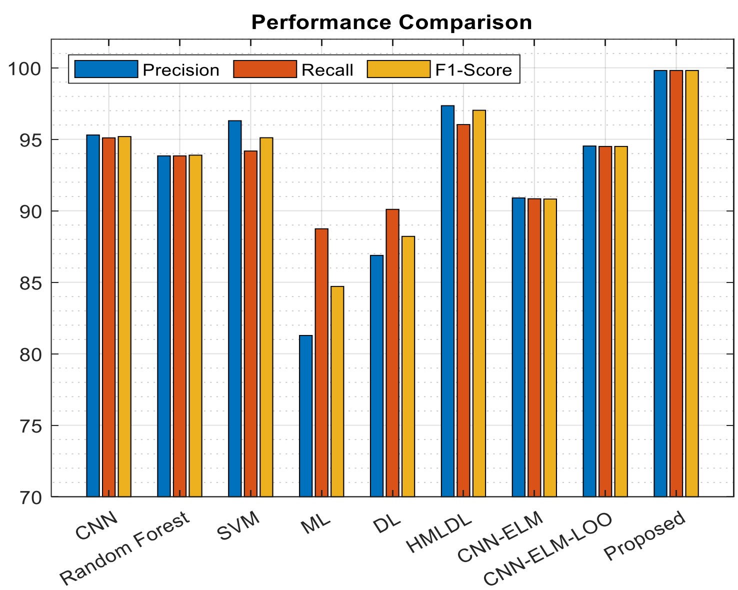

Figure 9: An evaluation of the proposed method in relation to previous studies is presented, using Precision, Recall, and F1-score.

The SVM-based method achieved an accuracy of 96.20% and DL reported an accuracy of 98.43%, both presenting competitive results. Additionally, the HMLDL discussed in the work of Zabihi et al. was shown to have an excellent accuracy value of 99.26%, which is also one of the highest values that can be found within the current literature. The proposed method, with the improvement of applying EELM-LOO architecture, achieved an accuracy rate of 99.81%, which outperforms all the methods shown in the table. Stage intermediates of the proposed model, the CNN-ELM and CNN-ELM-LOO yielded accuracies of 90.84%, and 94.50% respectively in a clear stage progression towards effective model structural enhancement. CNN-ELM, CNN-ELM-LOO, and proposed CNN-EELM-LOO are responsible for improving performance step by step through LOO validation, multiple EELM cooperation or uncertainty decision strategy.

Table 3: Performance of the proposed method compared with previous studies based on accuracy.

Author	Methodology	Accuracy
Cao M et.al. [20]	CNN	**90.80**
Gour A et.al. [21]	Random forest	**93.80**
Gour A et.al. [21]	SVM	**96.20**
Zabihi F. et.al. [22]	Machin Learning (ML)	**97.46**
Zabihi F. et.al. [22]	Deep Learning (DL)	**98.43**
Zabihi F. et.al. [22]	Hybrid Machin-Learning Deep-Learning (HMLDL)	**99.26**
	CNN-ELM	**90.84**
Proposed Method	CNN-ELM-LOO	**94.50**
	CNN-EELM-LOO	**99.81**

The drastic increase in the accuracy of classification results demonstrates the power of using a deep learning mechanism integrated with an optimized ELM structure and placing a Leave-One-Out approach during training. The improved learning capability of the model over complex arrhythmia for ECG patterns has greatly increased the overall performance in terms of accuracy. Thus, the presented method can be regarded as one of the most precise and trustworthy methods that have been proposed in this area until now.

5. Conclusion

A hybrid ECG-based cardiac arrhythmia detection system was developed in the proposed research. Here, the CNN is used for feature extraction purpose and it overlaps to EELM-LOO which conducts classification with both objectives of achieving high accuracy and maintaining stable results. In the preprocessing step, DWT is implemented to denoise and enhance the ECG signal quality, hence promoting feature extraction. On the MIT-BIH Arrhythmia Database, The result of our proposed CNN-EELM-LOO model provides a promising overall classification accuracy which reaches to 99.81%. It outperforms current state-of-the-art such as existing traditional machine learning based approaches and those of deep learning. Moreover, the model showed strong arrhythmia class separation, with 100%, 99.9% and 99.9% accuracy for U, S, and V classes respectively. The robustness of Leave-One-Out cross-validation offers important improvement in the sensitivity of the model to parameter initialization, allowing a consistent and repeatable outcome for tissue classification with each run. This is a key step towards an automated ECG diagnosis, able to deliver more accurate interpretations and analysis in less time. Future studies may be aimed to examine either the extensibility of this solution or its implementation in real-time, time-critical cardiac monitoring systems.

Funding Statement*:* The author received no specific funding for this study.

Data Availability: The data that support the findings of this study is MIT-BIH Arrhythmia Database v1.0.0. available online publicly https://www.physionet.org/content/mitdb/1.0.0/

Conflicts of Interest: The authors declare no conflicts of interest regarding this study.

Authors contributions. Conceptualization; data curation and methodology; validation and visualization; writing—original draft preparation; writing—review and editing; supervision and project administration: ANH. The author had approved the final version.

References

[1] Pandey, V., Lilhore, U. K., and Walia, R. (2025). "A systematic review on cardiovascular disease detection and classification". *Biomedical Signal Processing and Control*, 102, 107329.

[2] Strasburger, J. F., Eckstein, G., Butler, M. et al., (2022). "Fetal arrhythmia diagnosis and pharmacologic management". *The Journal of Clinical Pharmacology*, 62, S53-S66.

[3] Ardeti, V. A., Kolluru, V. R., Varghese, G. T. et al., (2023). "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to AI-based approaches". *Expert Systems with Applications*, 217, 119561.

[4] Hassan, A. N., Mohsin, M. J., and Khwayyir, A. H. K. (2020). "Discrimination of ECG signal based on S-interpolation and Quantum Neural Network". *Annals of Tropical Medicine and Public Health*, 23(10).

[5] Wu, Z., and Guo, C. (2025). "Deep learning and electrocardiography: systematic review of current techniques in cardiovascular disease diagnosis and management". *Biomedical Engineering Online*, 24(1).

[6] Darroudi, A., Parchami, J., Sarbishaei, G. et al., "Removing ECG noise from surface EMG based on information theory". *In proceeding of IEE Iranian Conference on Electrical Engineering (ICEE)*, 1403-1408. 2018.

[7] Xiao, Q., Lee, K., Mokhtar, S. A. et al., (2023). "Deep learning-based ECG arrhythmia classification: a systematic Review". *Applied Sciences*, 13(8), 4964.

[8] Hussein, E. A. R., Abdulridha, H. M., and Hassan, A. N. (2015). "Feature extraction of ECG signal using cubic spline technique". *European Journal of Scientific Research*, 128(1), 97-107.

[9] D, H., Vijay Anand, L. D., J, J. et al., "Cardiac Arrhythmia Diagnosis Using Deep Learning: A 1D CNN-GRU Approach with Multiclass SVM and DWT Analysis". *in 2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII): IEEE*, 1-7. 2024.

[10] Gayathri, M., Lavanya, V., Gobi Shankar, S. et al., "Machine learning-based cardiac arrhythmia prediction using DWT and combined morpho-statistical features". *in Computer Science Engineering*, London, CRC Press, 81-90, 2024.

[11] Koley, B. L. and Ankura, D., "Automatic detection of cardiac arrhythmias using discrete wavelet transform and support vector machine". *in 2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS): IEEE*, 1-5. 2024.

[12] Alamatsaz, N., Tabatabaei, L., Yazdchi, M. et al., (2024). "A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection". *Biomedical Signal Processing and Control*, 90, 105884.

[13] Alsayat, A., Mahmoud, A. A., Alanazi, S. et al., (2025). "Enhancing cardiac diagnostics: a deep learning ensemble approach for precise ECG image classification". *Journal of Big Data*, 12(1).

[14] Selvam, I. J., Madhavan, M., and Kumarasamy, S. K. (2025). "Detection and classification of electrocardiography using hybrid deep learning models". *Hellenic Journal of Cardiology*, 81, 75-84.

[15] Mavaddati, S. (2025). "ECG arrhythmias classification based on deep learning methods and transfer learning technique". *Biomedical Signal Processing and Control*, 101, 107236.

[16] Singh, P. N., and Mahapatra, R. P. (2024). "A novel deep learning approach for arrhythmia prediction on ECG classification using recurrent CNN with GWO". International Journal of Information Technology, 16(1), 577-585.

[17] Kolhar, M., and Al Rajeh, A. M. (2024). "Deep learning hybrid model ECG classification using AlexNet and parallel dual branch fusion network model". *Scientific Reports*, 14(1).

[18] Xiao, Q., and Wang, C. (2025). "Adaptive wavelet base selection for deep learning-based ECG diagnosis: A reinforcement learning approach". *PLOS ONE*, 20(2), e0318070.

[19] Issa, M. F., Yousry, A., Tuboly, G. et al., (2023). "Heartbeat classification based on single lead-II ECG using deep learning". *Heliyon*, 9(7), e17974.

[20] Cao, M., Zhao, T., Li, Y. et al., (2023). "ECG Heartbeat classification using deep transfer learning with Convolutional Neural Network and STFT technique". *Journal of Physics: Conference Series*, 2547(1), 012031.

[21] Gour, A., Gupta, M., Wadhvani, R. et al., (2024). "ECG Based Heart Disease Classification: Advancement and Review of Techniques". *Procedia Computer Science*, 235, 1634-1648.

[22] Zabihi, F., Safara, F., and Ahadzadeh, B. (2024). "An electrocardiogram signal classification using a hybrid machine learning and deep learning approach". *Healthcare Analytics*, 6, 100366.

[23] Hassan, A. N., Mohassel Feghhi, M., and Esmaeili, V. (2021). "A fast automatic modulation classification based on stft using hybrid deep neural network". *Journal of Communication Engineering*, 10(2), 328-343.

[24] Alinsaif, S. (2024). "Unraveling arrhythmias with graph-based analysis: a survey of the MIT-BIH database". *Computation*, 12(2), 21.

[25] Physionet 2005. "MIT-BIH Arrhythmia Database v1.0.0". physionet.org. Retrieved 2026, April 06 from https://www.physionet.org/content/mitdb/1.0.0/.

[26] Elaraby, A., Saad, A., Karamti, H. et al., (2023). "An optimized deep learning approach for robust image quality classification". *Traitement Du Signal*, 40(4), 1573-1579.

[27] Saad A. "Bird species identification using spectrograms and convolutional neural networks", Ph.D. dissertation, Dept. Computer and Microelectronic Syst., University Teknologi Malaysia, Johor, Malaysia, 2020.

[28] Saad, A., S. Kamil, I., Alsayat, A. et al., (2022). "Classification COVID-19 based on enhancement X-Ray images and low complexity model". *Computers, Materials & Continua*, 72(1), 561-576.