**International Journal of Informatics and Communication Technology (IJ-ICT)**
Vol. 15, No. 4, December 2026, pp. 1851~1860
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i4.pp1851-1860 

# Classification of cardiac ECG signals using quadratic discriminant analysis for atrial fibrillation

**Wameedh Raad Fathel[1], Marwah Nabeel Ibrahim Al-Siraj[2], Maysaloon Abed Qasim[3], Marwa Mawfaq Mohamedsheet Al-Hatab[2]**
[1]Ministry of Education, General Directorate of Education, Nineveh, Iraq
[2]Technical Engineering College, Northern Technical University, Mosul, Iraq
[3]Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University, Mosul, Iraq

**Article Info**

***Article history:***

Received Jun 25, 2025
Revised Jun 8, 2026
Accepted Jul 5, 2026

***Keywords:***

Atrial fibrillation
ECG classification
Ensemble learning
Machine learning
Spectral features

**ABSTRACT**

Accurate and timely detection of atrial fibrillation (AF) is a crucial issue in avoiding the risk of serious cardiovascular complications such as stroke and heart failure. This study proposed a machine learning-based system used for automated electrocardiogram (ECG) signal classification using features derived from spectral and temporal analyses. The study used a dataset obtained from the PhysioNet 2017 challenge database with 5000 samples labeled as normal rhythm (N), AF, and noisy rhythm (NR). Five primary features: average spectral energy, variation in spectral power, root mean square fluctuation, overall signal energy, and power variance were extracted, expressing both amplitude and frequency-domain characteristics of cardiac dynamics. 10-fold stratified cross-validation approach was used to evaluate the performance of the model using multiple classifiers, which are: quadratic discriminant analysis (QDA), support vector machine (SVM), random forest (RF), K-nearest neighbors (KNN), and multi-layer perceptron (MLP). Additionally, to improve model stability and accuracy, a soft voting-based ensemble model was developed. Although RF achieved the highest accuracy (94.15%), the ensemble model (93.21%) outperformed QDA and SVM. The ensemble approach achieved sensitivities ranging from 0.89 to 0.96 and specificities between 0.95 and 0.97 across all rhythm categories, demonstrating strong generalization.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Maysaloon Abed Qasim
Technical Engineering College for Computer and Artificial Intelligence
Northern Technical University
Mosul, Iraq
Email: maysloon.alhashim@ntu.edu.iq

## 1. INTRODUCTION

One of the leading causes of mortality worldwide is cardiovascular diseases (CVDs), which cause approximately 18 million deaths annually. The electrocardiogram (ECG) is the standard noninvasive tool used for monitoring cardiac electrical activity and detecting rhythm abnormalities such as arrhythmias and myocardial infarctions [1]. The recent advancements in machine learning (ML) and deep learning (DL) have enabled automatic classification of ECG signals, which facilitates earlier diagnosis and scalable monitoring of cardiac conditions.

Despite its clinical importance, manual interpretation of ECG signals is time-consuming and highly dependent on expert cardiologists, making it challenging in large-scale screening and continuous monitoring scenarios. ECG recordings are often affected by noise, baseline wander, and inter-patient variability, which

further complicate accurate diagnosis. Traditional signal processing and rule-based methods rely on handcrafted features and threshold-based decision rules, limiting their robustness and generalization across diverse patient populations [2].

Recent research has demonstrated that deep learning architectures can automatically learn discriminative features directly from raw or minimally processed ECG signals. These approaches have shown superior performance in arrhythmia detection, myocardial infarction classification, and multi-class ECG diagnosis when compared to conventional machine learning methods. Furthermore, the integration of transfer learning and ensemble strategies has improved model generalization, making AI-based ECG analysis a promising tool for real-time clinical decision support and remote healthcare applications [2].

## 2. LITERATURE REVIEW

In 2018, Zhao *et al.* [3] introduced a Bayesian spectro-temporal method used for atrial fibrillation (AF) detection from ECG. This method outperformed standard time-frequency methods such as STFT, CWT, and BurgAR, achieving an overall F1 score of 80.17, with the lowest cross-validation standard deviation (StdF1 = 1.06), which indicates robust and reliable minority-class detection [3]. In 2019, Yee *et al.* [4] proposed an algorithm for AF classification using i) discrete wavelet transform for baseline correction, ii) Pan-Tompkins algorithm for peak detection, and iii) feature selection via genetic algorithm and relief. The performance of bagged tree, decision tree, and k-NN classifiers was perfect, achieving nearly 100% accuracy, sensitivity, and specificity after feature selection [4]. In the same year, Dang *et al.* [5] improved an automated AF detection system based on a deep CNN-BLSTM network, using RR intervals and heartbeat sequences. Their model obtained training accuracy reaching 99.94% and testing accuracy equal to 96.59%, and specificity equal to 97.03%, confirming the robustness of the system in complex ECG classification [5]. Also in 2019, Ansari *et al.* [6] provided a comprehensive review of using DL architectures such as CNNs, RNNs, and Transformers in arrhythmia detection, demonstrating the transfer from handcrafted features to end-to-end learning systems [6]. In 2022, Ziccardi *et al*. [7] conducted a decision tree-based AF detection system using interference variables including Rényi permutation entropy, Fisher information, statistical complexity, Lyapunov exponent, and fractal dimension. Optimal classification quality based on ROC, F1-score, sensitivity, and specificity was achieved using Monte Carlo simulations with a 70%/30% training-testing split [7]. In 2022, Khriji [8], proposed an artificial neural network (ANN) based system for AF detection; they used information taken directly from the ECG signal without explicit feature extraction. They achieved high accuracy and specificity, overcoming prior approaches [8]. In 2023, Setiadi *et al.* [9] used a 1D convolutional neural network with the MIT-BIH database for AF classification. Their method obtained perfect results with all metrics (accuracy, precision, recall, and F1) reaching 100%, highlighting the effectiveness of deep learning in ECG signal analysis [9]. In 2023, Gupta *et al.* [10] introduced a hybrid approach for AF classification integrating local mean decomposition (LMD) and ensemble classifiers. Their method surpassed the traditional ML models with accuracies of 92.33%, 90.33%, and 90% for different AF types [10]. In 2023, Suguna *et al.* [11] introduced a framework for AF, normal, and other rhythms classification based on a random forest and using the PhysioNet 2017 dataset. The morphological and heart rate variability features were extracted. They achieved F1 scores of 0.86, 0.97, and 0.83 for AF, normal, and other rhythms, respectively, with an overall accuracy of 0.91 [11]. In 2024, Dsouza *et al.* [12] developed a hybrid system using a support vector machine (SVM) and feature selection to reduce computational complexity. Using 5822 ECG signals, the results achieved was 0.95 F1 score which was validated through 10-fold cross-validation [12].

Despite all these advances, many studies suffer from challenges such as class imbalance, inter-patient variability, and a lack of standardized evaluation protocols. This study introduces a framework consisting of different ML algorithms that vary from classical discriminant analysis to ensemble and neural models for classifying ECG signals into normal, AF, and noisy rhythm (NR) categories. The study combines feature extraction with advanced ML models. It addresses class imbalance, ensures generalization through cross-validation, and provides a clinically reliable framework for automated atrial fibrillation detection [13]-[15].

## 3. RESEARCH METHOD

### 3.1. Dataset description

The dataset used consists of 5,000 ECG signal samples collected from the PhysioNet 2017 challenge database [16]. The dataset is labeled into three diagnostic categories: 83% of samples are classified as normal, 12% as atrial fibrillation, and 5% as noisy rhythm. Each one represents a heart rhythm segment with annotations validated by expert cardiologists.

In order to ensure an accurate and reliable classification process, the dataset was preprocessed and split into training and testing subsets using a 10-fold stratified cross-validation strategy, maintaining class balance in each fold to avoid overfitting or class bias.

### 3.2. Feature extraction

Discriminative features extracted from the ECG signals captured both temporal and spectral dynamics of cardiac activity. These features were derived from the power spectrum and amplitude variations of the signals because they are very effective in describing rhythm irregularities associated with atrial fibrillation. Table 1 summarizes the extracted features with their mathematical formulations and physiological interpretations.

Table 1. Mathematical expressions and physiological interpretations of ECG feature extraction

| Feature name | Mathematical expression | Physiological interpretation |
|---|---|---|
| Average signal spectral power | $P_{avg} = \frac{1}{N}\sum_{i=1}^{N} X(f_i)$ | Represents the mean spectral power of ECG signal, which reflect the overall energy distribution over frequencies. |
| Spectral power from signal variation | $P_{var} = \frac{1}{N-1}\sum_{i=1}^{N}(X(f_i) - P_{avg})^2$ | Offers insight into rhythm variability by reflecting variations in spectral power that arise from the dynamic changes in the signal amplitude. |
| Integrated and detrended RMS fluctuation | $F_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$ | It measures long-range temporal correlations and detects complex signal dynamics in pathological rhythms. It calculated after integrating and detrending the ECG signal (yi). |
| RMS variation of time series | $RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ | Captures changes in temporal variability, and measures strength of amplitude variations in the ECG waveform. |
| Average total power of time series | $P_{total} = \frac{1}{N}\sum_{i=1}^{N} {X_i}^2$ | Represents the overall signal intensity, providing a global assessment of cardiac activity. |
| Variance of time-series power | $\sigma_p^2 = \frac{1}{N-1}\sum_{i=1}^{N}({X_i}^2 - P_{total})^2$ | Quantifies the variability of power across the ECG signal, giving insight into rhythm irregularity and instability. |

Where Xi represents the original ECG signal amplitude at sample i, Yi is the integrated and removed signal used for long-range correlation analysis. $X(f_i)$ is the fourier transform of the ECG signal, while N is the total number of samples in the time series [17].

### 3.3. Data preprocessing

The following preprocessing steps were used in order to ensure data quality and suitability for ML:

- Missing value imputation: Missing entries in the dataset were replaced using the mean value of the corresponding feature:

$$X_{filled} = \begin{cases} x_i & if\ x_i\ is\ present \\ \overline{X},\ if\ x_i\ is\ missing \end{cases} \quad (1)$$

where $\overline{X}$, is the mean of the observed values of the feature.

- Feature standardization: All features were standardized using Z-score normalization to ensure equal contribution across models, particularly for distance- and margin-based classifiers such as SVM and KNN:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (2)$$

where μ and σ are the mean and standard deviation of the feature, respectively.

- Label encoding: Categorical target labels were converted into numerical form using a label encoder, enabling compatibility with scikit-learn classifiers.
- Class imbalance correction: To address class imbalance, the synthetic minority oversampling technique (SMOTE) was applied. SMOTE generates synthetic samples for minority classes by interpolating between existing instances, thereby improving sensitivity and balanced representation:

$$X_{Smote} = Xmajority \cup Xminority \cup Xsynthetic$$

### 3.4. Model development

Discriminant analysis (DA) is a statistical method used particularly in classification tasks, and it has been widely used in the medical diagnostics field. It aims to find a set of features that separate different classes of data in the best way. The work of this method is to construct decision boundaries partitioning the feature space into regions corresponding to each class. These boundaries are established by modelling the data distribution within each class and identifying the features that provide the greatest separation among the classes [18]. This method has been:

Our study proved that this method is very important in differentiating between different classifications of cardiac rhythm, such as normal rhythm (N), AF, and NR. There are two main types of discrimination analysis:

- Linear discriminant analysis (LDA) is commonly used for classification tasks. Clear and effective decision boundaries can be derived by assuming that the data for each class follow a Gaussian (normal) distribution with different mean vectors but share the same covariance matrix [17]. LDA works by determining two main matrices: the within-class scatter matrix ( $S_W$), which reflects the variability of data points within each class, and the between-class scatter matrix ( $S_B$), which represents the variability between the mean vectors of different classes. The optimal projection is achieved by maximizing the ratio of the determinant of the between-class scatter to the within-class scatter, as shown in (3) [19]:

$$\mathrm{j}(w) = \frac{w^T S_B\, w}{w^T s_W w} \tag{3}$$

Where: w is the vector of weights that defines the linear transformation, and SW is the within-class scatter matrix, and SB is the between-class scatter matrix.

- Quadratic discriminant analysis (QDA) is a generalization of LDA that allows for different covariance matrices for each class. This makes QDA more flexible and suitable for data where class distributions have different shapes, variances, or correlations [20]. The probability density function for class i is shown in (4):

$$p\left({}^{x}/_{C_i}\right) = \frac{1}{(2\pi)^{d/2}|\sum i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \sum_i^{-1}(x-\mu_i)\right) \tag{4}$$

Where μ is the mean vector of class i, Σi is the covariance matrix of class i, and d is the dimensionality of the feature space.

The decision rule in QDA assigns a new data point x to the class with the highest posterior probability, which is computed using Bayes' Theorem as in (5) [21]:

$$\delta_i\,(x) = \log P(C_i) - \frac{1}{2}\,(x-\mu_i)^T\ \sum_i^{-1}(x-\mu_i) \tag{5}$$

Where p(Ci) is the prior probability of class i, p(x) is the marginal likelihood of the data, which acts as a normalization factor.

Since p(x) is the same for all classes, the classification rule simplifies by comparing the discriminant function as in (6) [22].

$$\delta_i(x) = \log p(C_i) - \frac{1}{2}\,(x-\mu_i)^T \sum_i^{-1}(x-\mu_i) \tag{6}$$

the data point x is classified into the class Ci with the largest value of δi(x).

### 3.5. Machine learning models

- SVM model is one of the supervised ML techniques that is used to classify data with a large number of features. It depends on creating a boundary between classes by a hyperplane in a high-dimensional feature space. The aim of SVM is to divide the dataset into various classes to locate a maximum margin hyperplane (MMH) [23]. A linear SVM has the (7):

$$f(x) = w.x + b \tag{7}$$

When x: input feature, w: weight vector orthogonal to the hyperplane, and b: bias term.

- KNN: This algorithm is the most important techniques in ML used to classify objects by finding the closest data points in a multi-dimensional feature space. In the training process, KNN retains the feature

vectors and labels from the training data. For classification, it calculates the distance between a test point and all training points, then finds the K nearest points. The test point is assigned to the class that is most common among these neighbors [24] (8).

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (8)$$

Where: x, y: The feature vectors of two data points, n: The number of features.

- Random forest builds an ensemble of decision trees and predicts the final class via majority voting [25], [26]. For each tree t in the forest (number of trees T=100) (9):

$$h_t(x) = DecisionTree_t(x) \quad (9)$$

the final prediction is obtained by majority vote across all trees (10):

$$\hat{y} = mode\{h_1(x), h_2(x), h_3(x) \ldots\ldots h_T(x)\} \quad (10)$$

- A multilayer perceptron (MLP) is a feedforward neural network composed of multiple hidden layers, where each layer applies a linear transformation followed by a nonlinear activation function. For an input vector, the first hidden layer computes (11):

$$a^{(1)} = f(W^{(1)}x + b^{(1)}) \quad (11)$$

and the second hidden layer outputs (12):

$$a^{(2)} = f(W^{(2)}a^{(1)} + b^{(2)}) \quad (12)$$

Where f(·) is a nonlinear activation function such as ReLU or tanh, and W and b denote the weights and biases of each layer. Finally, for multi-class classification, the output layer applies a softmax function to generate predicted class probabilities (13):

$$\hat{y} = softmax(W^{(0)}a^{(2)} + b^{(0)} \quad (13)$$

where $\hat{y}$ is the vector of class probabilities. This architecture enables the network to model complex nonlinear relationships between input features and output classes [27].

- Soft voting ensemble: It combines different classifiers, and it works by averaging the classifiers' predicted probabilities and optionally identifying different weights to each classifier. For an m-base classifier, each classifier m outputs a probability vector as in (14):

$$\hat{p}^{(m)} = P_{m,1}, P_{m,2}, \ldots P_{m,C} \quad (14)$$

C represents the number of classes. The ensemble prediction is then obtained using (15):

$$\hat{p} = \frac{\sum_{m=1}^{M} w_m \hat{p}^{(m)}}{\sum_{m=1}^{M} w_m} \quad (15)$$

and the final predicted class corresponds to the one with the highest averaged probability (16):

$$\hat{y} = arg \max_{c} \hat{p}_c \quad (16)$$

Where $w_m$ is the weight assigned to classifier mmm and $\hat{p}_c$is the ensemble probability for class c. This approach effectively leverages the strengths of multiple models, improving classification robustness and accuracy [25].

### 3.6. Cross-validation and model evaluation

In order to ensure robust and unbiased evaluation of the classifiers, 10-fold stratified cross-validation was applied; this technique divides the dataset into ten equal subsets (folds) while preserving the original class proportions in each fold. During each iteration, 90% of the folds are used for training and the remaining 10% for testing; this process is repeated ten times. Then the performance metrics are determined across all folds to provide reliable estimates of classifier performance [28].

### 3.7. Performance evaluation metrics

Different metrics were used to evaluate the performance of the model; these metrics are as follows [29]: Accuracy (ACC) is the total proportion of correctly classified samples; it can be calculated using (17).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

Sensitivity (SEN) or true positive rate: it is also called recall, and it represents the classifier's ability to identify positive cases correctly, as in (18).

$$SEN = \frac{TP}{TP+FN} \quad (18)$$

Specificity (SPE) or true negative rate: Is the measurement of the classifier's ability to identify negative cases correctly; it can be calculated using (19).

$$SPE = \frac{TN}{TN+FP} \quad (19)$$

Receiver operating characteristic (ROC) – area under the curve (AUC): Both micro-averaged and macro-averaged AUC values were computed to evaluate the discriminative power of classifiers across different decision thresholds, providing insight into overall and per-class performance [30], [31]. These metrics together provide a comprehensive evaluation of model performance and facilitate the selection of optimal thresholds for practical and clinical applications.

## 4. RESULTS AND DISCUSSION

### 4.1. Performance metrics

Table 2 shows the results achieved after evaluating the performance of ECG classification models.

Table 2. Performance comparison of ECG classification models

| Model | Accuracy (%) | Sensitivity | Specificity | AUC (Macro) |
|---|---|---|---|---|
| QDA | 64.30 | 0.64 | 0.82 | 0.849 |
| SVM | 82.29 | 0.82 | 0.91 | 0.939 |
| RF | 94.15 | 0.93 | 0.96 | 1.000 |
| KNN | 91.39 | 0.90 | 0.95 | 0.995 |
| MLP | 91.21 | 0.89 | 0.94 | 0.985 |
| Ensemble | 93.21 | 0.92 | 0.95 | 0.997 |

### 4.2. Model-wise evaluation

Among the different models utilized in this study (Figure 1), the lowest performance was recorded using QDA with 64.3% accuracy and 0.849 AUC. Its confusion matrix is shown in Figure 1(a), exhibiting overlap between AF and NR classes, which indicates that QDA is limited in modeling the non-Gaussian and complex distributions typical of ECG features. Nonetheless, the moderate specificity suggests reasonable identification of normal rhythms. The performance was enhanced when SVM was used, obtaining 82.3% accuracy and balanced sensitivity and specificity. The RBF kernel enabled effective modeling of nonlinear relationships, which improved the separation between rhythm classes. Some misclassification between AF and NR still appeared, most likely because of overlapping temporal and frequency-domain characteristics. Figure 1(b) illustrated the SVM confusion matrix, showing that RF achieved the highest performance with 94.15% accuracy and a perfect AUC of 1.000. Minimal classification errors appeared in the confusion matrix shown in Figure 1(c), which demonstrates the model's ability to reflect complex nonlinear interactions. The variance was reduced, and the generalization was improved when ensemble averaging was used within the RF. RF showed excellent detection across both majority and minority classes. KNN model achieved 91.39% accuracy and 0.995 AUC. Its high specificity indicates effective recognition of normal rhythms, although slight overlap between AF and NR persisted, as shown in Figure 1(d). Although the computational demand increases with dataset size, distance-based classification was improved using feature normalization. MLP algorithm also obtained strong performance, achieving 91.21% accuracy and 0.985 AUC. Figure 1(e) shows the confusion matrix of MLP. The ensemble model combined predictions from all classifiers, resulting in stable and high-level performance. It achieved 93.21% accuracy and 0.99 AUC. The hybrid probabilistic and margin-based voting improved class balance and mitigated individual model biases. The confusion matrix of the ensemble model is shown in Figure 1(f).

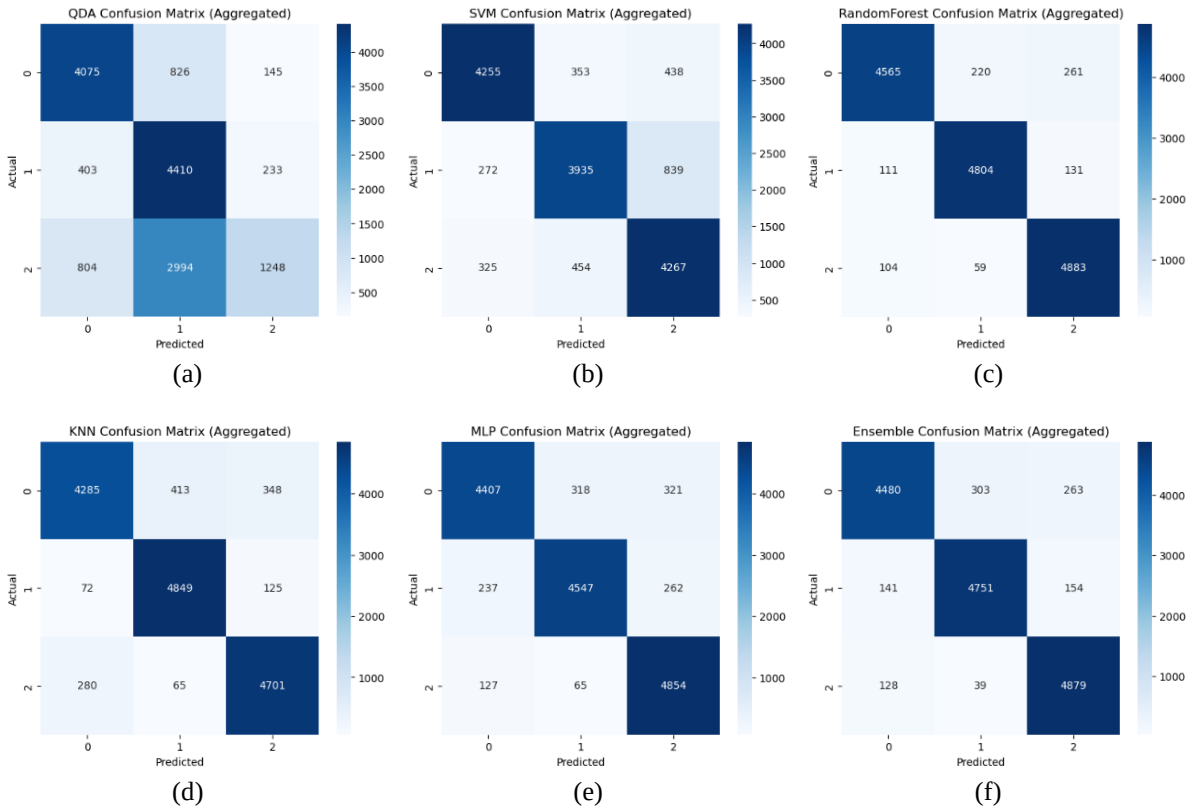



Figure 1. Confusion matrices of ECG classification: (a) QDA model, (b) SVM model, (c) RF model, (d) KNN model, (e) MLP model, and (f) ensemble model

### 4.3. ROC–AUC performance comparison

ROC analysis was used to evaluate the ability of each model to distinguish between the three rhythm classes. AUC scores are summarized in Table 3. ROC curves are shown in Figure 2, which highlights that the RF and Ensemble models achieved strong class separation, with a perfect AUC value of 1 across all rhythm categories.

Table 3. ROC–AUC scores for each rhythm class

| Model | AUC (N) | AUC (AF) | AUC (NR) | Macro-AUC |
|---|---|---|---|---|
| QDA | 0.912 | 0.844 | 0.791 | 0.849 |
| SVM | 0.952 | 0.934 | 0.932 | 0.939 |
| RF | 1.000 | 1.000 | 1.000 | 1.000 |
| KNN | 0.997 | 0.994 | 0.995 | 0.995 |
| MLP | 0.988 | 0.983 | 0.983 | 0.985 |
| Ensemble | 0.999 | 0.997 | 0.996 | 0.997 |

### 4.4. Model performance comparison and discussion

The best overall performance was achieved using the RF algorithm, which demonstrated complete class separation and the highest classification accuracy. The Ensemble model achieved slightly lower accuracy but exhibited greater robustness and improved class balance. The SVM provided reliable nonlinear discrimination, whereas QDA was limited by its sensitivity to covariance variations among features. Both MLP and KNN models delivered strong and consistent performance, highlighting the effectiveness of neural network-based and instance-based learning approaches.

These findings confirm the suitability of ensemble and tree-based methods for ECG signal classification, particularly when combined with balanced preprocessing strategies and comprehensive multi-domain feature extraction.

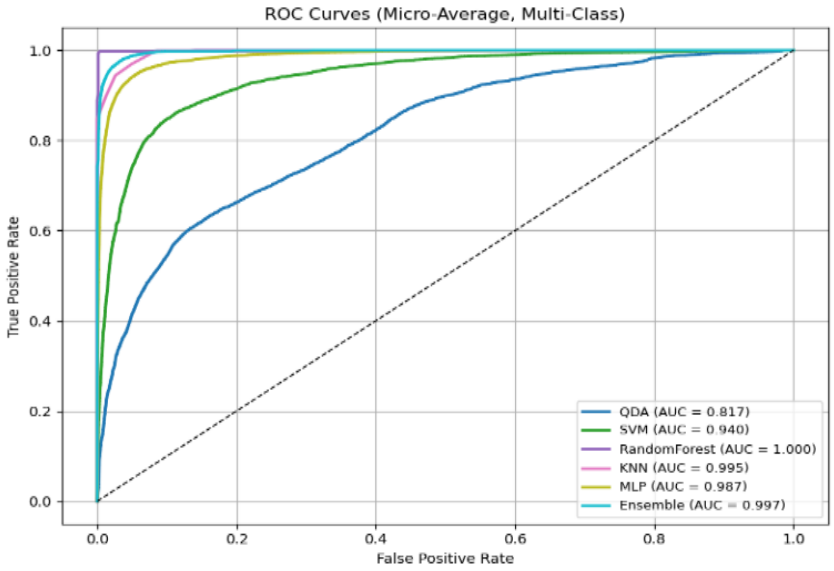



Figure 2. ROC curves for ECG rhythm classification using QDA, SVM, RF, KNN, MLP, and ensemble models

## 5. CONCLUSION

In this study, a framework is proposed for automated ECG signal classification into Normal, Atrial Fibrillation, and Noisy Rhythm categories. The system used six classifiers: QDA, SVM, RF, KNN, MLP, and a soft voting Ensemble model. The proposed system combined feature standardization, SMOTE-based class imbalance correction, and 10-fold stratified cross-validation.

The results achieved by the system showed that the RF achieved the highest classification accuracy among all classifiers, recording 94.15% accuracy and demonstrating superior discriminative capability, whereas the Ensemble model exhibited robust and stable predictive performance. These results ensure the importance of addressing class imbalance and jointly evaluating sensitivity and specificity to ensure clinical applicability.

## FUNDING INFORMATION

No funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wameedh Raad Fathel | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Marwah Nabeel Ibrahim Al-Siraj | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Maysaloon Abed Qasim | ✓ | | ✓ | ✓ | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Marwa Mawfaq Mohamedsheet Al-Hatab | | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | | | ✓ | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

There is no Conflict of Interest.

## DATA AVAILABILITY

The data that support the results of this study are openly available in [ AF classification from a short single-lead ECG recording: The PhysioNet Computing in Cardiology Challenge 2017," in Computing in Cardiology, Rennes, France, 2017] at https://www.physionet.org, reference number [17].

## REFERENCES


[1] D. S. Park and G. I. Fishman, "The cardiac conduction system: development, function and therapeutic targets," *Nat. Rev. Cardiol*, vol. 23, pp. 303–323, 2026, doi: 10.1038/s41569-025-01227-x.

[2] Q. Xiao et al., "Deep learning-based ECG arrhythmia classification: A systematic review," *Applied Sciences*, vol. 13, no. 8, p. 4964, Apr. 2023, doi: 10.3390/app13084964.

[3] Z. Zhao, S. Särkkä, and A. Bahrami Rad, "Spectro-temporal ECG analysis for atrial fibrillation detection," *in Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*, 2018, doi: 10.1109/MLSP.2018.8517085.

[4] N. J. Yee, V. Vijean, S. Awang, *et al.*, "Atrial fibrillation identification through ECG signals," *J. Phys.: Conf. Ser.*, vol. 1372, no. 1, Art. no. 012058, 2019, doi: 10.1088/1742-6596/1372/1/012058.

[5] H. Dang, M. Sun, G. Zhang, et al., "A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals," *IEEE Access*, vol. 7, pp. 75577–75588, 2019, doi: 10.1109/ACCESS.2019.2918792.

[6] Y. Ansari, O. Mourad, K. A. Qaraqe, and E. Serpedin, "Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017–2023," *Frontiers in Physiology*, vol. 14, p. 1246746, Sept. 2023, doi: 10.3389/fphys.2023.1246746.

[7] I. G. Ziccardi, A. Rey, and W. Legnani, "ECG signals classification using overlapping variables to detect atrial fibrillation," *Trends in Computational and Applied Mathematics*, vol. 23, no. 3, 2022, doi: 10.5540/tcam.2022.023.03.00569.

[8] L. Khriji, "ECG pattern recognition technique for atrial fibrillation detection," *Revue d'Intelligence Artificielle,* vol. 36, no. 2, pp. 205–211, 2022, doi: 10.18280/ria.360205.

[9] R. M. Setiadi, M. Fachrurrozi, and M. N. Rachmatullah, "Classification of atrial fibrillation in ECG signal using deep learning," *SJIA*, vol. 4, no. 1, 2023, doi: 10.36706/sjia.v4i1.53.

[10] K. Gupta, V. Bajaj, and I. A. Ansari, "Atrial fibrillation detection using electrocardiogram signal input to LMD and ensemble classifier," *IEEE Sensors Letters*, vol. 7, no. 8, 2023, doi: 10.1109/LSENS.2023.3281129.

[11] G. C. Suguna, S. Shirahatti, S. R. Bangari, et al., "Classification of atrial fibrillation using random forest algorithm," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 10S, 2023, doi: 10.17762/ijritcc.v11i10s.7599.

[12] V. Dsouza, K. V. Keerthana, and V. F. Densil Raj, "Atrial fibrillation classification: A hybrid approach integrating support vector machines and feature selection for reduced computational complexity," *in Proc. ICA-AIC*, 2024, doi: 10.1109/ICA-AIC60222.2024.10575256.

[13] R. R. O. Al-Nima, M. M. M. Al-Hatab, and M. A. Qasim, "An artificial intelligence approach for verifying persons by employing the deoxyribonucleic acid (DNA) nucleotides," *Journal of Electrical and Computer Engineering*, vol. 2023, Art. no. 6678837, 2023.

[14] M. A. Al-Hashim, W. R. Fathel, H. D. Ali, and M. M. M. Al-Hatab, "Enhanced non-invasive blood glucose monitoring system employing wearable optical technology," *Fusion: Practice and Applications*, vol. 19, no. 1, pp. 1–9, 2025, doi: 10.54216/FPA.190101.

[15] M. M. M. Al-Hatab, R. R. O. Al-Nima, and M. A. Qasim, "Classifying healthy and infected Covid-19 cases by employing CT scan images," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3279–3287, 2022, doi: 10.11591/eei.v11i6.4344.

[16] G. D. Clifford, *et al.*, "AF classification from a short single-lead ECG recording: The PhysioNet computing in cardiology challenge 2017," *in Computing in Cardiology*, Rennes, France, 2017. [Online]. Available: https://www.physionet.org

[17] S. Guo and H. Tracey, "Discriminant analysis for radar signal classification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 4, pp. 3134–3148, 2020, doi: 10.1109/TAES.2020.2965787.

[18] K. T. Le, C. Chaux, F. J. P. Richard, and E. Guedj, "An adapted linear discriminant analysis with variable selection for the classification in high-dimension, and an application to medical data," *Computational Statistics & Data Analysis*, vol. 152, p. 107031, 2020, doi: 10.1016/j.csda.2020.107031.

[19] J. Zhao and Y. Sun, "A comparative study of linear and quadratic discriminant analysis for big data classification," *Journal of Data Science*, vol. 20, no. 4, pp. 725–738, 2022, doi: 10.6339/jdsci.2022.200418.

[20] S. Das and S. Gupta, "A robust approach for classification using quadratic discriminant analysis in biomedical applications," *Soft Computing*, vol. 24, no. 7, pp. 5091–5102, 2020, doi: 10.1007/s00542-020-06091-x.

[21] R. Singh, N. Rajpal, and R. Mehta, "Dynamic ECG classification using shift-invariant DTCWT and discriminant analysis," *Lecture Notes in Electrical Engineering*, 2020, pp. 490–500, doi: 10.1007/978-3-030-30577-2_43.

[22] J. Wang and J. Li, "Application of discriminant analysis in identifying medical diseases using ECG data," *Biomedical Signal Processing and Control*, vol. 67, Art. no. 102651, 2021, doi: 10.1016/j.bspc.2021.102651.

[23] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An overview on the advancements of support vector machines (SVMs) in the medical field," *Information*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.

[24] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, Art. no. 113, 2024, doi: 10.1186/s40537-024-00973-y.

[25] A. Manconi, G. Armano, M. Gnocchi, and L. Milanesi, "A soft voting ensemble classifier for detecting patients affected by COVID-19," *Applied Sciences*, vol. 12, no. 15, Art. no. 7554, 2022, doi: 10.3390/app12157554.

[26] A. B. Shaik, "A brief survey on random forest ensembles in classification," in *Proc. Int. Conf. on Machine Learning and Applications*, Springer, 2018, pp. 411–420, doi: 10.1007/978-981-13-2354-6_27.

[27] N. F. Sopelsa Neto *et al.*, "A study of multilayer perceptron networks applied to classification of ceramic insulators using ultrasound," *Applied Sciences*, vol. 11, no. 4, Art. no. 1592, 2021, doi: 10.3390/app11041592.

[28] T. Leinonen, *et al.*, "Empirical investigation of multi-source cross-validation in clinical ECG classification," *Computers in Biology and Medicine*, vol. 183, Art. no. 109271, 2024, doi: 10.1016/j.compbiomed.2024.109271.
[29] R. D. Mahdi, M. A. Qasim, N. M. Allayla, and R. A. Jaleel, "A customized IoMT-cloud based healthcare system for analyzing of brain signals via supervised mining algorithms," *Journal of Engineering Science and Technology*, vol. 17, pp. 76–83, 2022, doi: 10.31838/jcr.07.05.430.
[30] M. Al-Siraj, Mesut Çevik, and S. M. Ibrahim, "Suggestion new monitoring system by depending on the human activity recognition videos," in *Proceedings of the 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, 2022, doi: 10.1109/ISMSIT56059.2022.9932751.
[31] O. Rainio *et al.*, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, Art. no. 56706, 2024, doi: 10.1038/s41598-024-56706-x.

## BIOGRAPHIES OF AUTHORS



**Wameedh Raad Fathel** is a computer engineer with experience in industry and education. He holds a Bachelor's degree from Northern Technical University, Iraq, and a Master's degree from Near East University, Turkey. His expertise includes programming, network setup, and AI applications. His research fields include AI in healthcare, Image processing, signal processing, and IoT applications. He can be contacted at email: wameed81@gmail.com.



**Marwah Nabeel Ibrahim Al-Siraj** received the B.Sc. degree in Computer Engineering from Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq, and the M.Sc. degree in Electrical and Computer Engineering from Universita Altınbaş Istanbul, Turkey. She is a lecturer at the Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq. She can be contacted at email: marwa.nabil@ntu.edu.iq.



**Dr. Maysaloon Abed Qasim** received the B.Sc. degree in Electrical Engineering from Mosul University, Iraq; the M.Sc. degree in Electronic and Communication Engineering from Mosul University, Iraq; and the Ph.D. degree in Electronic and Computer Engineering from Istanbul, Turkey. She is an Assistant Professor at the Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University (NTU), Mosul, Iraq. She has published multiple research papers. She can be contacted at email: maysloon.alhashim@ntu.edu.iq.



**Marwa Mawfaq Mohamedsheet Al-Hatab** received the B.Sc. degree in Medical Instruments Engineering from Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq, and the M.Sc. degree in Biomedical Engineering from Universita Politecnica delle Marche (UNIVPM). She is a lecturer at Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq. She has published multiple research papers. She can be contacted at email: marwa.alhatab@ntu.edu.iq.