FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

New Cardiovascular Signal Processing and Feature Engineering Methods for Disease Detection Based on Machine Learning

Haider Abdulkarim[1], Hussein Khaleel[2], Shaimaa Hameed Abd[3], and Qussay S. Tawfeeq[4*]

[1]*College of Communications Engineering, University of Technology, Baghdad, Iraq; Email: haider.a.abdulkarim@uotechnology.edu.iq*
[2]*College of Communications Engineering, University of Technology, Baghdad, Iraq; Email: hussein.j.khaleel@uotechnology.edu.iq*
[3]*College of Communications Engineering, University of Technology, Baghdad, Iraq; Email: shaymaa.h.abed@uotechnology.edu.iq*
[4]*College of Communications Engineering, University of Technology, Baghdad, Iraq; Email: qussay.s.tawfeeq@uotechnology.edu.iq*

***Correspondence:** *qussay.s.tawfeeq@uotechnology.edu.iq; Tel.: (+9647707603879)*

ABSTRACT- Identifying and diagnosing cardiac problems is heavily based on the data analyzed from an ECG. The utilization of artificial intelligence for automated analysis of ECG's is becoming more common. The way in which the data is processed and which model is selected significantly influences how accurately the signal is classified. This paper presents a methodology for ECG signal processing for creating feature sets in the time domain, in the frequency domain, and of statistical metrics that are used to develop machine learning classification models. All experiments with the proposed methods were conducted using the PTB-XL data set. The results indicate that by designing appropriate feature sets, a wide variety of machine-learning models can be created, which can categorize complex types of cardiac disease more accurately than using random feature sets. Binary and multi-class classification accuracies were achieved at 86.52% and 73.78%, respectively, using the feature sets developed from the proposed methods. Analysis of the features set shows that the feature set adequately represents all identified dynamic characteristics present in the signals collected. As a result, this feature set can also be utilized with any form of method that utilizes an approach other than deep learning. Therefore, there is enough evidence from this research study to conclude that utilizing features-based methods will yield satisfactory results.

Keywords: ECG Classification, Deep Learning, LSTM, CNN, Arrhythmia Detection, Data Augmentation.

ARTICLE INFORMATION

Author(s): Haider Abdulkarim, Hussein Khaleel, Shaimaa Hameed Abd, and Qussay S. Tawfeeq;

Received: 24/06/2025; **Accepted**: 20/02/2026; **Published**: 30/03/2026;
E- ISSN: 2347-470X;
Paper Id: IJEER 2406-20
Citation: 10.37391/ijeer.140123
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-14/ijeer-140123.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

In cardiology field, the process of using Electrocardiogram (ECG) signals features to detect many types of cardiovascular conditions is very important. ECG signals are measured via electrodes on the skin. The morphological features of these signals reflect highly informative characteristics about the functions of the heart. According to the World Health Organization (WHO), cardiovascular diseases are the most dominant cause of mortality worldwide. Therefore, the accurate classification of cardiology diseases, such as coronary artery, arrhythmias, and myocardial infarction in timely manner is crucial for saving many lives worldwide. Early detection could also be used in parallel with wearable devices monitoring, making it suitable for all-day cardiac activity supervision.

The importance of systems with automated ECG classification has been more obvious in the last couple of years, especially in early fast and accurate detection of some life-threatening cardiac conditions, such as ventricular tachycardia and atrial fibrillation, even before the patient could observe any symptoms [1]. Those systems play vital rule in wearable devices technology, where early alarm could alert both the doctors and patients with plausible cardiac failure [2]. There is, however, some challenges to those alarming systems, such as the requirement of large training datasets and the disparity of ECG features among individuals [3].

Over the past few decades, more cardiologists have transitioned from manual patient inspection towards fully automated systems, relying on both Machine- and Deep-Learning methods. These algorithms mainly rely on two stages: extraction of features, and the classification of samples. While those methods are accurate in lab-controlled conditions, they were more prone to errors in uncontrolled environments. To enhance classification accuracy of multiple cardiac diseases, [4] proposed a deep learning algorithm combined with a loss function based on Adaptive Focal Cross-Entropy (AFCE).

Algorithms of machine learning, such Neural Networks and Support Vector Machines (SVMs) have been increasingly gaining attention by researchers in the field of ECG classification, since 2000s. The work presented in [5] showed the accuracy of SVM classifier to detect arrhythmias, using features extracted from ECG waveforms, achieving acceptable accuracy with high preprocessing requirements.

On the other side, Genetic Algorithm (GA) was used in [6] for both feature extraction and classification stages, with high

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

complexity requirements. However, the appearance of deep learning algorithms in 2010s helped further advance ECG classification.

The work in [7] introduced an ECG classifier using Recurrent Neural Networks that can capture temporal correlations in time samples. Results showed that the model was able to detect ventricular and supraventricular ectopic beats.

The research introduced by [8] used PSO to hyper tune deep neural network neuron counts, drop rates and learning rate, which ultimately improved ECG classification.

Since 2020, many datasets were publicly available, such as Physikalisch-Technische Bundesanstalt (PTB-XL). They improve the research by offering large and heterogeneous ECG trials, which enabled more generalized models [1][10].

In spite of the aforementioned advances, there are many challenges. Those are computational requirements of the classification models, trustable AI models that could gain clinical trust, and the different ECG signal characteristics across different patients. Recent trends in the ECG classification include the compression of data of multi-lead ECG data acquisition, less-hardware requirements, and hybrid classification models that combine physiological and feature-based topologies for more enhanced accuracy, as investigated in [11][12]. The classification of ECG is dynamically changing and evolving, reflecting the need for continuous research. Major advances are achieved using deep learning and AI techniques. However, continuous research is mandatory to prevail over current limitations and to achieve the ultimate goal of fully-automated cardiac-condition diagnosing system.

This work investigates the automated identification of cardiac diseases from raw ECG time-series signals. A comprehensive set of features is proposed, encompassing time-domain, frequency-domain, and statistical characteristics calculated from the ECG records. Multiple machine learning models were employed, which utilize the aforementioned extracted features, to perform classification among cardiac conditions. The adopted models were evaluated and their performances were compared to assess classification effectiveness.

Section 2 presents the methods carried out to process the ECG signals, extract the proposed features, and building the classification models. In *section 3*, the evaluation of the classifiers is discussed and the performances are compared. Finally, *section 4* summarizes the contribution of this work in terms of the approach taken and the achieved performance.

2. METHODOLOGY

This work utilizes the publicly and freely accessible PTB-XL [13][14] clinical dataset. The dataset comprises readings collected using a set of 12 leads that were attached to patients. The dataset includes a total of 21837 records. Each record includes the set of 12 raw ECG signals corresponding to the 12 leads, diagnosis, and patient information, among other metadata.

The diagnoses of the dataset fall into five so-called Superclasses representing the cardiac conditions: Normal cases (NORM), ST/T Change (STTC), Myocardial Infarction (MI), Hypertrophy (HYP), and Conduction Disturbance (CD). *Table 1* lists the Superclasses and their corresponding number of records.

Table 1. Number of records per superclass in PTB-XL dataset

Number of Records	Superclass Symbol	Description
9514	NORM	Normal ECG
5469	MI	Myocardial Infarction
5235	STTC	ST/T Change
4898	CD	Conduction Disturbance
2649	HYP	Hypertrophy

PTB-XL is a multi-label dataset, in which a record can have more than one diagnostic statement that are accompanied by a numerical likelihood. The diagnosis of each record is represented by a standard ECG-SCP code [15].

Each of the ECG signals in a record has a duration of 10 seconds. The dataset includes two sets of records, one having a sampling rate of 500, and having 5000 samples per signal, as well as a down-sampled version at 100 samples per second and having 1000 samples per signal.

The following sections present the work carried out in reading and processing the data, and extracting meaningful information for building the classification models.

2.1. Data Aggregation

This section describes the process by which the readings and diagnosis were collected from the PTB-XL dataset structure. The data was formatted such that each record is represented as a row of samples (raw ECG signal), followed by the corresponding diagnosis.

The records that have a diagnosis with likelihood value less than 80% were discarded. Moreover, for records having multiple classes, the class label having the highest likelihood was selected. By filtering out ambiguous cases, the classifier trains on more definitive records, leading to a more robust model.

The trimmed-down and accumulated data at this stage have a total number of records of 13071, fewer than the original PTB-XL data due to the aforementioned discards. The diagnosis labels for the superclasses are mapped to integer number labels.

Table 2 shows the number of records and the label mapping for the different classes after the aggregation stage. This process is applied for each of the 12 leads, and the result of this stage is the aggregated matrix of signals and diagnosis that is used for processing.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

Table 2. Label mapping and number of records of the aggregated data

Number of Records	Superclass Symbol	Superclass Label
8603	NORM	1
886	MI	2
1988	STTC	3
1431	CD	4
163	HYP	5

2.2. Preprocessing

This section describes the process extracting the proposed features out of the raw ECG signals. These features will be fed into the machine learning models for training and validation. The ECG signals pass through the stages discussed in the next sections.

2.2.1. Band Pass Filter

A dual-stage digital bandpass filter was used on the raw ECG signal. First, a high-pass filter with a cut-off frequency of 0.5 Hz and a Gaussian window was applied to eliminate the baseline wander (slow drift) of the signal. Afterwards, a recursive Gaussian low-pass filter with a cut-off frequency of 15 Hz was applied to suppress high-frequency noise components. *Figure 1* shows the raw versus the filtered ECG signals. It can be seen that the combined bandpass filtering is able to effectively remove both the slow variations and the high-frequency noise components.

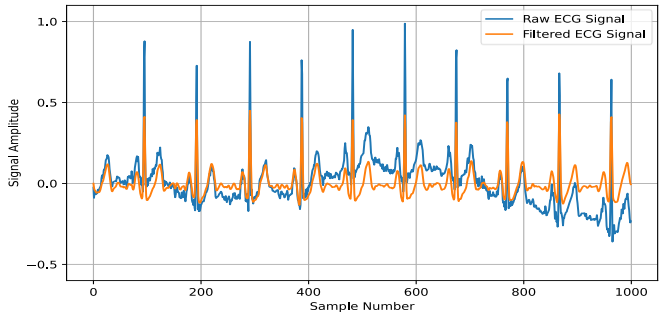

Figure 1. ECG signal before and after the band-pass filter

This filter is necessary for the reliable operation of the feature extraction stage. The Gaussian filter has been chosen because of its linear phase response. This maintains equal delay of frequency components of the ECG signal, and consequently, it does not introduce distortion or filter artifacts. This stage is important for the stability and reliability of the following feature extraction in the time domain.

2.2.2. Standardization

Each ECG signal is processed to have a mean value of zero and variance value of one, according to the following expression:

$$x_{std} = \frac{x - \mu}{\sigma} \quad (1)$$

where x is the original signal, μ is the mean of the signal, and σ is the standard deviation of the signal. Such normalization facilitates inter signal comparison and consistency in the feature extraction stage.

2.2.3. Feature Extraction

This work does not directly utilize the raw ECG signals with the machine learning models, but instead uses features that provide time-domain, frequency domain, and statistical representation of the original signals.

This approach reduces data dimensionality, enhances robustness to noise and inter-subject variability, and facilitates more effective learning by the models. For every one of the 12 leads, each ECG signal (1000 samples over 10 seconds) is transformed into 87 samples, representing the calculated features. *Table 3* to *table 6* below list the features adopted in this study.

Table 3. Statistical and frequency domain features

Index	Feature Group	Description
0	Skewness	Third standardized moment, measures distribution asymmetry around mean $skewness = \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$
1	Kurtosis	Fourth standardized moment, measures the sharpness of a data distribution $\text{kurtosis} = \frac{E[(x-\mu)^4]}{\sigma^4}$
2–12	Percentiles (10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90)	Amplitude percentiles of the ECG signal
13	Shannon Entropy	Signal randomness: $H = -\sum p_i \log p_i$
14	Hjorth Activity	Signal variance σ^2
15	Hjorth Mobility	Measures the frequency components of a signal $\sqrt{\frac{\text{Var}(\dot{x})}{\text{Var}(x)}}$
16	Hjorth Complexity	Ratio of mobility of signal derivative to mobility of signal
17-21	Frequency Band Energy (0–10, 10–20, 20–30, 30–40, 40–50 Hz)	Spectral energy within predefined frequency bands
22–31	Histogram (10 bins)	Normalized histogram of signal amplitude distribution

Table 4. Morphological time domain features

Index	Feature Group	Description
32–33	P-wave duration (mean, std)	Duration statistics of P wave
34–35	Q–R interval (mean, std)	Statistics of time interval between Q and R peaks
36–37	R-wave duration (mean, std)	Duration statistics of R wave
38–39	R–S interval (mean, std)	Statistics of time interval between R and S peaks
40–41	T-wave duration (mean, std)	Duration statistics of T wave
42–43	P–R interval (mean, std)	Statistics of time interval between P and R peaks
44–45	R–T interval (mean, std)	Statistics of time interval between R and T peaks

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

Table 5. Heart rate and amplitude related features

Index	Feature Group	Description
46–47	Heart rate (mean, std)	Statistics of heart rate
48–49	P/R amplitude ratio (mean, std)	Statistics of amplitude ratio of P and R peaks
50–51	T/R amplitude ratio (mean, std)	Statistics of amplitude ratio of T and R peaks
52–53	R amplitude (mean, std)	Amplitude statistics of R peak
54–55	S amplitude (mean, std)	Amplitude statistics of S peak
56–57	P-wave area (mean, std)	Statistics of area under the P wave
58–59	QRS area (mean, std)	Statistics of area under the QRS complex
60–61	T-wave area (mean, std)	Statistics of area under the T wave
62	HRV_MeanNN	Mean of the intervals between successive R peaks (RR intervals)
63	HRV_SDNN	RR intervals standard deviation
64	HRV_RMSSD	RR intervals root mean square
65	HRV_pNN50	Percentage of the difference between successive RR intervals > 50 ms
66	HRV_MedianNN	RR intervals median
67	HRV_IQRNN	RR intervals interquartile range
68	HRV_SDSD	The standard deviation of the successive differences between RR intervals
69	HRV_CVNN	SDNN divided by HRV_MeanNN
70	HRV_CVSD	HRV_RMSSD divided by HRV_MeanNN

Table 6. Dimensionality reduction and time–frequency features

Index	Feature Group	Description
71	PCA First Component	First principal component of the feature set
72–86	Wavelet Features	4-level discrete wavelet decomposition; energy, entropy, and variance per level

The sets of features in *table 4* and *table 5* were extracted with the assistance of the Neurokit2 toolbox [16]. Neurokit2 is a python-based package that provides routines to process neurophysiological signals.

NeuroKit2 is an open-source Python-based library that provides analysis of neurophysiological and biomedical signals. It provides standardized pipelines for multiple modalities, including cardiac activity (ECG), electrodermal activity (EDA/GSR), respiration (RSP), photoplethysmography (PPG/BVP), electromyography (EMG), and related signals, and supports the extraction of a wide range of time-, frequency-, and non-linear features.

In this work, NeuroKit2 was employed for analyzing ECG signals. Specifically, its interval-based analysis routines were used to estimate the durations and temporal relationships of ECG waveform components. *Figure 2* shows one complete cardiac cycle of an ECG signal. It is often referred to as a P–QRS–T cycle, because it includes: P wave (atrial depolarization), QRS complex (ventricular depolarization), and T wave (ventricular repolarization). The more advanced analyses of NeuroKit2, such as frequency-domain and non-linear feature extraction, were not applied, as these methods typically require longer continuous recordings, whereas the ECG signals of the PTB-XL considered in this study were limited to a duration of 10 seconds.

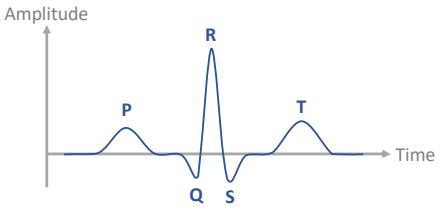

Figure 2. Waveform of a cardiac cycle

As an alternative to the frequency-domain and non-linear analyses provided by NeuroKit2, custom feature extraction methods were implemented in this work. Frequency-domain features were derived by computing the spectral energy within predefined frequency bands, while non-linear characteristics of the ECG signals were quantified using entropy-based measures. These features were directly calculated from the standardized ECG signals and are suitable for short-duration recordings.

While NeuroKit2 provides detection of fiducial points (indices) corresponding to the different ECG waveform components, additional post-processing was performed to extract amplitude-based features such as amplitude ratios, areas, and their corresponding statistical descriptors.

In the case of certain records, the NeuroKit2 library could not identify ECG signal components and, therefore, did not give valid results. To maintain the uniformity of the extracted features, these records were ignored. After eliminating the invalid entries, the final dataset consisted of 2706 records.

Dimensionality reduction methods were explored, in particular, Principal Component Analysis (PCA) [17] and Linear Discriminant Analysis (LDA) [18]. With PCA, the first 20 components were calculated, and were included as features, however, this led to a reduced accuracy of prediction. Additional tests revealed that keeping only the first principal component provided better prediction performance, and therefore only this first component of PCA was included in the final set of features. Features based on LDA were also tested, however a negative impact was observed on classification accuracy, and LDA aspects were therefore omitted from the final feature set.

Characteristics that were obtained with the Discrete Wavelet Transform (DWT) [19] were added to identify the time-frequency properties of the ECG signals. A Daubechies-4 (db4)

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

wavelet with four-levels was used, where the obtained coefficients are associated with different frequency ranges. The wavelet coefficients at each level, were used to calculate the following three features: (i) the wavelet energy at each sub-band, (ii) the wavelet entropy that measures the complexity and irregularity of the component, and (iii) the wavelet variance that describes the variability of the signal activity. The features derived from the wavelet offer extra localization in terms of time that FFT is not able to capture, and provide complementary information.

2.2.4. Combining Features

After the feature extraction, the data representing each lead is organized into a table format, with each row corresponding to a single record of a given lead, and one column represents a feature. For each lead, the resulting number of columns in the features table was 87. The feature tables (a total of 12) were assembled next to each other to achieve a single representative table with 1,044 columns. The final features table constitutes the entire collection of the extracted features of all the 12 leads.

The diagnosis labels that matched the ECG records and linked to the features were stored as an independent single-column vector. They are ordered in the same way as long as rows are concerned. The final unified table of features and the class label vector are shown in *figure 3*.

Lead #1	Lead #2		Lead #12	Class
Extracted features of record #1	Extracted features of record #1		Extracted features of record #1	Label of record #1
Extracted features of record #2	Extracted features of record #2		Extracted features of record #2	Label of record #2
Extracted features of record #N	Extracted features of record #N		Extracted features of record #N	Label of record #N

Figure 3. Structure of the table of the combined features and the class labels

2.2.5. Data Split and Feature Scaling

The table of features was divided into two parts: training and test data, where 80 percent of the records were assigned to model training and 20 percent were for performance evaluation. Before the split, the records were shuffled randomly so that two subsets were statistically representative of the entire dataset. It is mentioned that the distribution of the class labels is no even, with some of the diagnostic classes are more frequent than others. This class imbalance was taken into account at the stage of model building. Feature scaling was applied before the training of the models. This ensures that features that are numerically larger do not incorrectly impact the model. Scaling has been utilized on the basis of the standardization process, transforming every feature (column) to have zero mean and unit variance ($\mu = 0$, $\sigma = 1$). The scaling parameters were derived exclusively from the training set and subsequently applied to both the training and test partitions. This prevents data leakage and maintains the integrity of the evaluation by keeping the training set strictly independent.

2.3. Training of Classification Models

In this work, multiple machine-learning based classification methods were initially evaluated during the development phase. For relevance, only the models that demonstrated the best performance and were ultimately used to in the reported results are presented. The selected algorithms employ different approaches to model and classify data.

The classifiers used in this study account for class imbalance during training by assigning higher weights to the classes having less occurrences, thereby reducing bias toward majority classes.

2.3.1. Support Vector Machines

Support Vector Machines (SVMs) [20] are supervised machine learning models commonly used for both classification and regression. The fundamental working concept of an SVM classifier is to transform the input features into a higher-dimensional space, and then to determine a decision boundary that separates data samples from the different classes while maximizing the margin between them. This makes the classifier more resistant to noise and more generic with respect to new unseen data.

SVM classifiers are able to make use of various kernel functions that define the boundaries of the classes in the feature space. A linear kernel assumes that the classes can be separated using a linear decision boundary, while non-linear kernels allow SVMs to model more complex class separations. The Radial Basis Function (RBF) was applied in this work as a non-linear kernel.

The Python implementation of the SVM classifier was done through the Scikit-learn library [21], and trained on the collection of features based on the training data.

2.3.2. Random Forest

The Random Forest (RF) [22] is an ensemble machine learning technique, it constructs and evaluates multiple decision trees and collectively use their outputs to improve prediction performance. It allows the model to deal with multifaceted sets of features. The decision trees in the RF are trained independently, and their predictions are aggregated. The final class is determined as the one that is most frequently predicted. Random Forest is able to capture complex and non-linear feature boundaries with minimal parameter tuning, and has been widely adopted in biomedical signal classification.

RF classifier was implemented and trained on the feature set by means of the Scikit-Learn module. The number of trees was set to 65.

2.3.3. K-Nearest Neighbor

The K-Nearest Neighbor (KNN) [23] is a machine learning technique that classifies samples according to their similarity to labeled training data. Rather than learning an explicit model, KNN assigns class labels by comparing a sample to a number of k neighbors that are closest in the feature space based on a distance metric such as Euclidean distance.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

Since the outcome of KNN is based on the local neighborhood of a sample, it can naturally model non-linear class boundaries, making it suitable for complex biomedical feature sets.

The KNN classifier was implemented using the Scikit-Learn library in Python, and trained on the extracted feature set. A number of 40 neighbors was set to determine a point's class, and the weight of each neighbor is weighted inversely with the distance from a given trial. Thus, closer points have more impact on the decision than points that are further away.

2.3.4. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) [24] are a type of deep learning methods that use multiple stacked layers to identify patterns from the input data. By combining convolution, non-linear activation functions, and pooling operations, CNNs capture local patterns and model complex non-linear relationships. The depth of the network plays a critical role in its modelling capacity.

The CNN classifier implementation in MATLAB is depicted in *figure 4*. As described in the previous sections, each element of the ECG record represents an independent feature. Therefore, the 1×1 convolution layers learn transformations of each feature without assuming any spatial relationship between adjacent elements. The CNN consists of three 1×1 convolutional layers with 16, 32, and 64 filters, each using the Rectified Linear Unit (ReLU) activation function. The ReLU function applies a non-linear transformation as it outputs 0 if the input is negative, and outputs the same input value if the latter is positive, and is expressed by:

$$ReLU(x) = max(0, x) \quad (2)$$

A fully connected layer combines all features from the preceding convolutional stages into a 64-dimensional feature space. During training, this layer randomly sets the outputs of 50% of the neurons to zero to reduce model overfitting. The final fully connected layer maps the feature space of the previous stage to a vector with a length that is equal to the number of classes. The softmax layer transforms class scores into a probability distribution where all values sum to one. The probability of class *i* is expressed as:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (3)$$

Finally, the classification layer compares the probabilities to the true class labels.

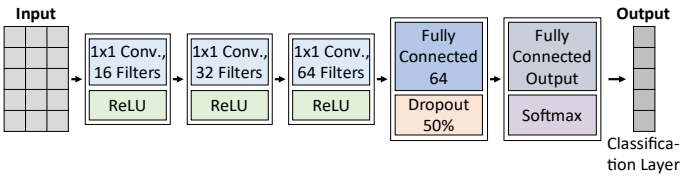

Figure 4. Structure of the CNN classifier implementation

3. RESULTS

This section reports the performance of the classification models described in the previous section, evaluated using the test dataset which includes the features extracted from the 12 leads. Two classification strategies were considered: *(i)* a binary classifier to consider only two classes representing normal and abnormal ECG signals, and *(ii)* a multi-label classifier to distinguish between all five diagnostic superclasses.

The performance each classifier was evaluated in terms of the accuracy, Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) [25], and confusion matrix [26]. The accuracy is calculated as the ratio of the number of correct predictions to the total number of predictions. The AUC indicates the ability of the model to distinguish between different classes across different decision thresholds. AUC values range from 0 to1, where 0.5 represents random guessing, 1 indicates perfect discrimination of classes, while values less than 0.5 suggest that the model is misinterpreting the classes due to an error.

The confusion matrix provides a summary of the classification performance. Rows and columns of the matrix represent the actual and predicted labels respectively. Each element of the matrix displays a count for the number of records at a given location, indicating whether a correct prediction was made.

3.1. Binary Classifier

For this classifier, normal (NORM) class categories were set to 0 and all other labels which depict abnormal diagnoses were set to 1. *Table 7* provides the overview of the classification performance measured using accuracy and AUC. Similar binary classification accuracies of 84.27%, 84.08%, and 85.39% were achieved by SVM, RF, and KNN respectively, and a slightly less value of 82.77% for CNN. These results reflect that the calculated features are indeed informative and can be successfully used for classification tasks.

In order to boost classification performance, a majority vote ensemble was applied, which is a combination of the predictions of the SVM, RF and KNN classifiers. This vote takes advantage of the different categorization mechanisms of each model. The majority vote classification achieved slightly higher accuracy of 86.52%. AUC values of 0.9157, 0.9070, 0.9122, and 0.89 were attained for SVM, RF, KNN and CNN classifiers respectively. Again, AUC values of similar range are observed, with CNN being slightly lower reflecting its lower accuracy value. These AUC values reflect strong discriminative capability for all models. The SVM and KNN achieved higher AUC values, suggesting superior class separation compared to RF and CNN. The high AUC values confirm that the proposed set of extracted features allows for effective discrimination between classes in a binary classification task.

Table 7. Accuracy and AUC values for the classifiers

	SVM	RF	KNN	Majority Vote	CNN
Accuracy (%)	84.27	84.08	85.39	86.52	82.77
AUC	0.9157	0.9070	0.9122	--	0.89

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

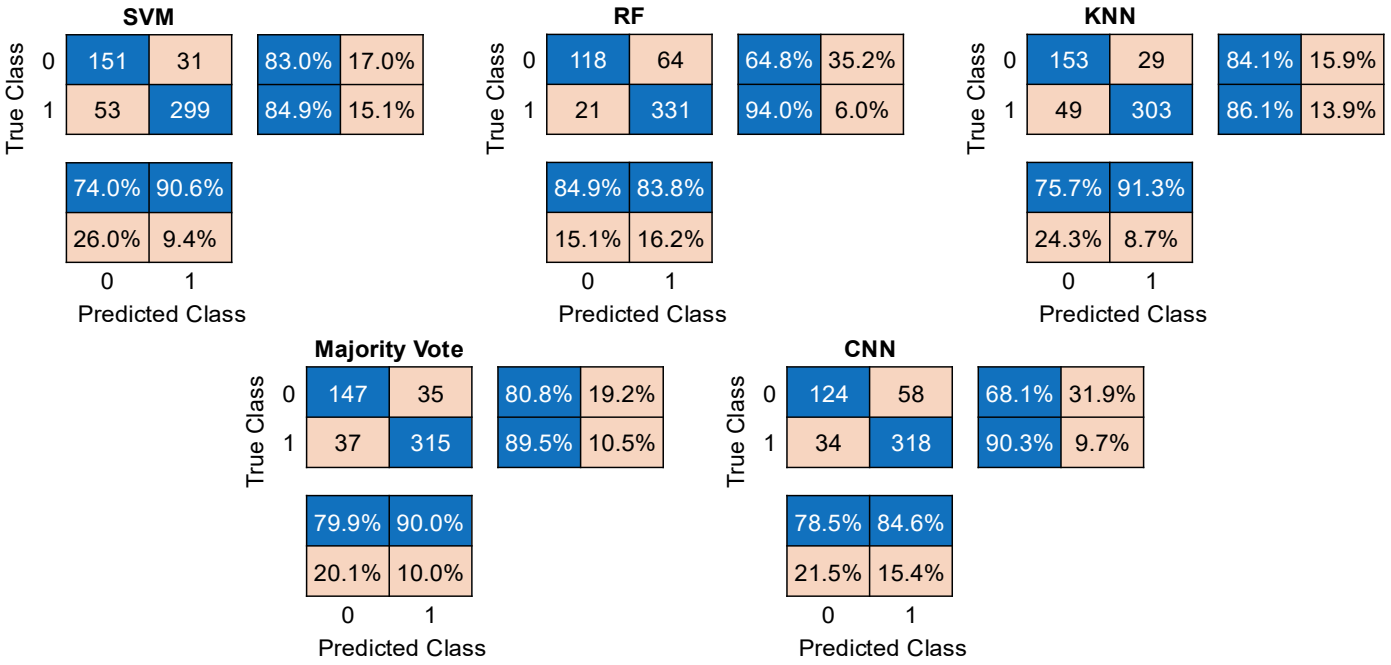

Figure 5. Confusion matrices for the binary classifiers

The confusion matrices for the binary classifiers are shown in *figure 5*. Each cell of the matrix displays the total number of predictions. The rows indicate the true class, and the columns indicate the predicted class. Main diagonal elements represent counts of the correctly predicted classes, while the remaining cells hold counts of misclassified test records. The vertical two-column matrix on the right side of the confusion matrix displays the percentages of correctly and incorrectly classified test records of a true class, in the first and second columns, respectively. The horizontal two-row matrix below the confusion matrix displays the percentages of correctly and incorrectly classified test records of a predicted class, in the first and second rows, respectively. It indicates the prediction precision per class. The confusion matrices show that while all the models achieve comparable performances, they exhibit different trade-offs between detecting class 1 (abnormal) and identifying class 0 (normal) avoiding false alarms.

It can be seen that RF achieves the highest detection performance for class 1 of 94%, meaning that it is able to detect positives most effectively, however, at the cost of higher false positives (35.2%), resulting in misclassification of class 0. CNN shows a similar pattern to RF. SVM and KNN exhibit a balanced behavior achieving similar detection performance for class 0 and class 1 test records.

The majority vote ensemble provides a compromise between true positive detection of class 1, and false positive misclassifications of class 0. A similar trend is observed in terms of prediction precision (bottom matrices). The majority vote and RF ensembles show a more balanced performance in predicting classes 0 and 1, while the other single models show a higher precision in detecting class 1 compared to class 0.

Comparison of the performance of the proposed work, specifically the majority vote, against previous related works, is show in *table 8*, in terms of accuracy and AUC. The proposed approach achieves a higher accuracy of 86.52% as well as AUC of 0.89, surpassing prior clinical studies. This demonstrates improved binary ECG classification performance.

Table 8. Performance comparison with previous works

Reference	Accuracy (%)	AUC
This work	86.52	0.89
[27]	81.77	0.81
[28]	73.80	0.79
[29]	73.50	--
[30]	70.50	0.79

3.2. Multi-Label Classifier

In this classification task, the classifiers implemented in Section 2.3 were trained on the records with multiple labels representing the 5 superclasses. SVM and CNN classifiers were considered in this part, as they outperformed the other classifiers discussed previously, according to the experiments conducted during this study.

Table 9 lists the accuracy and AUC values of SVM and CNN multi-label classifiers. As expected, the classification performance is inferior compared to the binary classifier, due to the elevated complexity of discriminating among multiple classes (cardiac conditions) rather than separating normal and abnormal cases.

In terms of accuracy, SVM has slightly higher value of 73.78% compared to the CNN of 70.51%. Similarly, the SVM attains a marginally higher AUC of 0.8695 compared to 0.8493 for the CNN. This indicates a better overall classification performance and discriminative ability.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

Table 9. Accuracy and AUC for the multi-label classifiers

	SVM	CNN
Accuracy (%)	73.78	70.51
AUC	0.8695	0.8493

The confusion matrices of the multi-label classifiers are shown in *figure 6*. Class label 1 represents the normal case (NORM), and labels 2, 3, 4, and 5 represent the cardiac conditions Myocardial Infarction (MI), ST/T Change (STTC), Conduction Disturbance (CD), and Hypertrophy (HYP), respectively. The confusion matrices show that both SVM and CNN are effective at detecting the normal class (NORM) or disease presence, but less effective at differentiating between the cardiac conditions 2-5. Class 5 is the most challenging for both classifiers, with extensive confusion across other classes.

The right-side matrices show that both models achieve higher detection ability for the normal class (≈81%), moderate for classes 2–4, and substantially lower for class 5. The bottom matrices display class-wise precisions, showing relatively stable precision for classes 1–4 and poor precision for class 5.

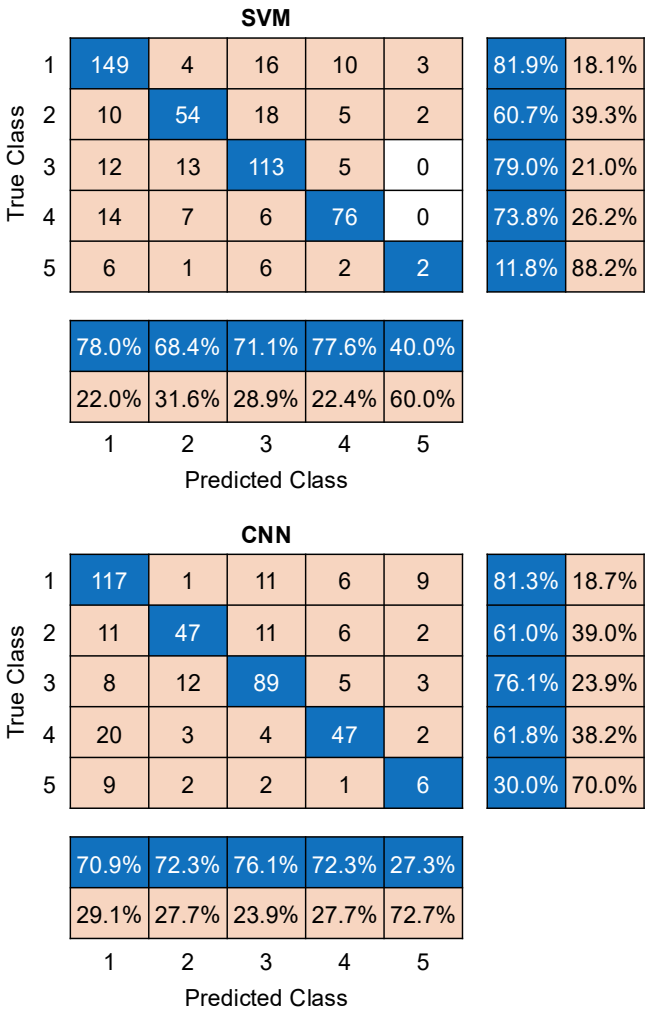

Figure 6. Confusion matrices of the multi-label classifiers

4. CONCLUSION

This study proposed an automated ECG signal classification leveraging a set of extracted features from raw signals. The features encompass a comprehensive set of time-domain, frequency-domain, and statistical characteristics that efficiently capture signal dynamics.

The validation of this work was done using the PTB-XL dataset, and using multiple machine learning classifiers. The performance of the suggested method was evaluated achieving an accuracy of 86.52% for the binary classifier, and 73.78% for the multi-label (five labels) classifier. It has also been demonstrated that the classic models of machine learning performed slightly better than the CNN using the engineered features. This underscores the fact that features design is of utmost importance to attain high performance classification with lower model complexity.

In the case of multi-label classifier, observing the confusion matrices indicated that the largest portion of errors was experienced when discriminating between the abnormal classes, which point to the difficulty of fine-grained diagnosis.

In summary, the findings of this work highlight that ML-based classification can be improved by using carefully engineered features.

Research efforts will continue in order to enhance the accuracy of classification, especially in the multi-label case. Multiple paths will be explored, including the use of a portion of the available leads, defining features that can better distinguish the abnormal states, and further investigating deep learning methods.

REFERENCES

[1] J. Rabcan, V. Levashenko, E. Zaitseva, and M. Kvassay, “Advancing ECG Signal Classification with a Fuzzy Classifier Approach,” IEEE Access, 2025

[2] S. Palermi, et al., “Artificial intelligence and the electrocardiogram: A modern renaissance”, European Journal of Internal Medicine, Volume 140, 2025, 106329

[3] B. T. Pham, et al., “Electrocardiogram Heartbeat Classification for Arrhythmias and Myocardial Infarction,” Sensors, vol. 23, no. 6, Mar. 2023

[4] H. N. Monday et al., "Enhancing ECG Classification in Cardiac Diagnostics: A Novel Approach Using Adaptive Focal Cross-Entropy Loss Function," in IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 10, pp. 7161-7174, Oct. 2025

[5] Saini, Indu & Kumar, V. & Khosla, Arun. (2013). Feature extraction of ECG signal using a support vector machine. Journal of Electrocardiology. Vol. 46, issue 4, 2013

[6] Gupta, V., Saxena, N.K., Kanungo, A. et al. “A review of different ECG classification/detection techniques for improved medical applications”. Int J Syst Assur Eng Manag 13, 1037–1051 (2022)

[7] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin and H. Yang, "Patient-specific ECG classification based on recurrent neural networks and clustering technique", 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), Innsbruck, Austria, 2017, pp. 63-67

[8] Kahlessenane, Y., Bouaziz, F., & Siarry, P. (2025). “A new particle swarm optimization-enhanced deep neural network for automatic ECG arrhythmias classification”. Computer Methods in Biomechanics and Biomedical Engineering, 1–15

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 1 | Pages 216-224 | e-ISSN: 2347-470X

[9] Immaculate Joy Selvam, Moorthi Madhavan, Senthil Kumar Kumarasamy, "Detection and classification of electrocardiography using hybrid deep learning models", Hellenic Journal of Cardiology, Volume 81, 2025, Pages 75-84

[10] Oke, O.A., Cavus, N., "Electrocardiogram image classification for six classes of heart diseases". Iran J Comput Sci 8, 419–439 (2025)

[11] Patra, N., Banerjee, S. & Bhadra, S. "On-device compression of multilead electrocardiogram using tunable-Q wavelet transform and MLPNN trained using multi-optima optimization based PSO". SIViP 19, 628 (2025)

[12] Teng Chen, Yumei Ma, Zhenkuan Pan, Weining Wang, Jinpeng Yu, "Fusion of multi-scale feature extraction and adaptive multi-channel graph neural network for 12-lead ECG classification", Computer Methods and Programs in Biomedicine, Volume 265, 2025

[13] Wagner, P., Strodthoff, N., Bousseljot, RD. et al. PTB-XL, a large publicly available electrocardiography dataset. Sci Data 7, 154 (2020)

[14] Wagner, Patrick, et al. "PTB-XL, a large publicly available electrocardiography dataset" (version 1.0.3). PhysioNet (2022)

[15] International Organization for Standardization. ISO 41064:2023: Health Informatics — Standard Communication Protocol — Computer-Assisted Electrocardiography. ISO, 2023

[16] Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods, 53(4), 1689-1696

[17] Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). New York: Springer-Verlag

[18] FISHER, R.A. (1936), THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. Annals of Eugenics, 7: 179-188

[19] Daubechies, I. (1992) Ten Lectures on Wavelets. SIAM, Philadelphia

[20] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995)

[21] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

[22] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001)

[23] Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13 (1967), pp.21-27

[24] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324

[25] Fawcett, T. An introduction to roc analysis. Pattern Recogn. Lett. 27(8), 861–874 (2006)

[26] Ting, K.M. (2011). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA

[27] Safdar, M.F., Nowak, R.M., Pałka, P. et al. An integrated algorithm for single lead electrocardiogram signal analysis using deep learning with 12-lead data. Sci Rep 15, 34955 (2025)

[28] K. Lee, et al. Mobile single-lead ecg atrial fibrillation prediction enhancement integrated by standard ecg algorithm with deep learning model. Eur. Heart J. 45, 3431 (2024)

[29] G. Luongo et al. Machine learning using a single-lead ECG to identify patients with atrial fibrillation-induced heart failure. Front. Cardiovasc. Med. 9, 812719 (2022)

[30] Kim, J. et al. Identification of atrial fibrillation with single-lead mobile ECG during normal sinus rhythm using deep learning. J. Kor. Med. Sci. 39(5), e56 (2024)