International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

PulseGuard: Intelligent Arrhythmia Detection and Classification through ECG Signal Analysis

Sharmila P. Zope[*1], Dr. Nandini Chaudhari[2], Dr. Nita Patil[3], Dr. Umesh Pawar[4]

Submitted: 05/05/2023 **Revised**: 17/07/2023 **Accepted**: 06/08/2023

Abstract: Arrhythmias are abnormal heart rates. The heart condition Ventricular Fibrillation (VF) and Ventricular Tachycardia (VT), included in Arrhythmias, are the leading causes of sudden cardiac arrest. It is essential for successful defibrillation treatment to identify potentially serious arrhythmias as early as possible. Heart disorders were studied in several ways. Among them, an ECG (electrocardiography) test is regarded as the most effective noninvasive type of inquiry. Most widely utilized arrhythmia detection, as well as classification methods, rely only on surface Electrocardiogram analysis. So, an algorithm corresponding to shape statistical features and spectral kurtosis features analysis using supervised machine learning algorithms to increase the efficiency of heart diagnostics. The proposed prediction framework considers the feature selection technique with a Support Vector Machine, Naïve Bayes, Linear Discriminant Analysis, K-nearest neighbor, and Decision Tree for early arrhythmia diagnosis. The empirical results on the publicly available MIT-BIH Arrhythmia database with supervised classification achieve an efficient prediction accuracy of 93.75% for the decision tree. This study also seeks to develop a predictive model with a feature selection technique for detecting arrhythmias that improves heart diagnostics' accuracy.

Keywords: *Classification, ECG Signal, Machine Learning, QRST Detection, Sequential Feature Selection.*

1. Introduction

The modern lifestyle of humans is a fundamental cause of health issues, including heart disease. Several people were leading to death due to sudden cardiac attacks [1]. As a result, heart disease has emerged as a significant public health issue. Using a new technique for monitoring heart health in its early stages may save lives and improve people's life quality [2].

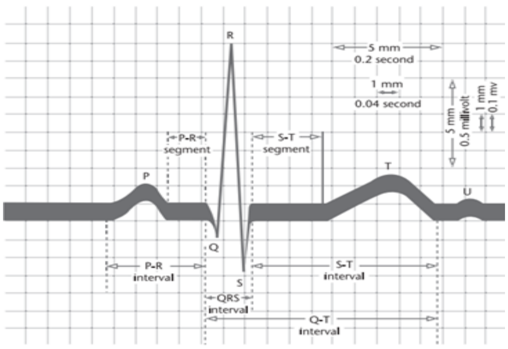

Fig. 1. Components of ECG signal

[1] *School of Computer Sci., KBCs North Maharashtra University, Jalgaon – 425001, India*
ORCID ID: 0009-0000-0330-8199
[2] *DRS Kiran and Pallavi Patel Global University, Vadodara – 391760, India*
ORCID ID: 0000-0002-6178-1328
[3] *School of Computer Sci., KBCs North Maharashtra University, Jalgaon – 425001, India*
ORCID ID: 0000-0001-9616-0356
[4] *SND College of Eng., Nashik – 423401, India*
ORCID ID: 0000-0001-8766-013X
*[*1]sharmilazope@gmail.com, [2]nandini113@gmail.com, [3]nitaapatil@gmail.com, [4]umeshpawar.snd@gmail.com*

Figure 1. displays the ECG signal's components [3]. The electrocardiogram (ECG), which shows the heart's electrical activity caused by contraction and reflection, is frequently used for monitoring the functioning of the heart. QRS complex, P wave, T wave, and the small U wave are the four standard components of an electrocardiogram. A heart arrhythmia is indicated by any anomaly in such a component. The time between two consecutive regular ECG beats is the RR interval, a standard method of determining a person's heart rate [3].

$$\text{Heart Rate} = [60 / \text{RR interval}] \text{ beats per min.}$$

60-100 beats per minute is the average heart rate. Heart arrhythmia is characterized by a heart rate that varies from the usual speed and other symptoms.

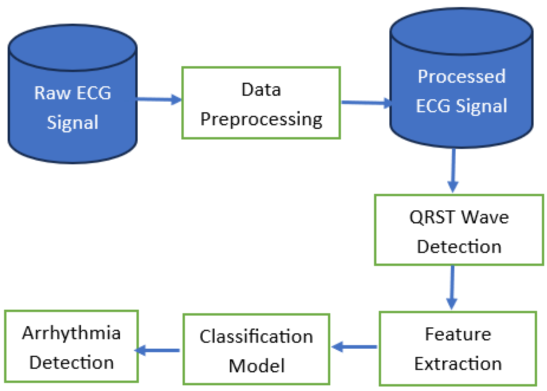

Fig. 2. General steps of Arrhythmia Detection using ECG.

Figure 2 depicts that ECG features are used to generate an automated approach for classifying arrhythmias. Since automatic detection is a computer-aided task, providing the most significant parts of ECG is very important in making accurate diagnosis possible. The earlier approaches focused on various aspects, like preprocessing [4], feature extraction [5], and learning techniques [6]. This study aims to create an ECG-signal-analyzing algorithm for automated arrhythmia identification and categorization. Examining an ECG signal is a straightforward, repeatable, and inexpensive technique. The investigation concerns four arrhythmias databases from MIT-BIH: Standard Arrhythmia ECG Data, CU Ventricular Tachyarrhythmia Database, Ventricular Tachyarrhythmia Database, and Supraventricular Arrhythmia Database. The performance is evaluated with five supervised learning classifiers for efficient diagnosis.

Following is the outline for this paper: Related Work on identifying and categorizing arrhythmias using ECG signal and different approaches for heart disease prediction discussed in Section II. Information regarding the database is provided in Section III. In Section IV, we discuss our methodology. The results and interpretations of the research are discussed in Section V. In Section VI, a summary of the conclusion is presented.

2. Related Work

The component of a basic ECG frequency range is between 0.5 Hz to 100 Hz. The diagnosis of various conditions has become more challenging due to the presence of artifacts in the ECG data. A high amount of noise in the ECG record distorts the critical features of the ECG signal. Such data should be filtered or need to be discarded [4]. A notch filter is typically used to reduce power line interference caused by poor ECG equipment grounding and surrounding equipment interference. The other artifact has less impact than the interference noise, which results in the ECG signal to unreadable records [[7]. Several techniques have been tried to eliminate those tendencies to allow for further analysis of the ECG data. The ECG signals were primarily extracted using those filtering techniques in sluggish, non-stationary orientations. It has been explored how to eliminate power line interference and ECG signal noise using several methods [8]. Implementing direct methods of QRS detection allows for the QRS complex to be detected directly using samples of the ECG signal. Those QRS complex detection techniques depend heavily on accurate filtering and thresholding techniques. It is possible to develop a filter that efficiently separates the QRS complex from the ECG signal by applying the power spectrum analysis of various frequency signal components in the ECG signal [9]. The Wavelet Transform (WT) and Integer Moving Average (IMV) are among the most potential techniques in this group. There may need to be more than just knowing the pitch of the R wave to identify many unusual QRS complexes with high amplitudes and prolonged durations (not very steep slopes). Thus, extracting more information from the signal is fundamental to seeing a QRS event [10]. Pre-processing is used to "learn" some of the features used in the identification stage. Some methods in this group are linear and nonlinear prediction, differentiation, syntactic, and neural nets. Many QRS detection techniques are built on differentiation. The median, a nonlinear digital filter, eliminates the baseline drift noise in the ECG signal [11].

ECG signals are subjected to sophisticated feature extraction techniques, and features can be derived in either the time domain or the frequency domain [5]. It has also been proposed to solve the lack of spectral features and non-stationary patterns of ECG signals using Wigner-Ville analysis in a two-dimensional frequency domain [12]. The Wavelet Transform (WT), in comparison to other wavelet families and with conventional descriptors, is one of the various methods used [13]. One was the Karhunen Loeve Transform, Hermitian Basis, promising [14], and many other techniques [6]. To decompose the ECG signal into simple components that can be located in time to detect the QRS complex, the paper [15] applies a wavelet technique. The detection accuracy is 99.8 % despite most ECG signals' noise, baseline drift, and other artifacts. The paper [14] presented that The Harmit basis function coefficients were calculated from a high pass filter to avoid low-frequency components' impact on ECG signals. For accurate heartbeat detection, a novel feature extraction method using clustering results in an accuracy of 99.7%. In [6], the comparative study of feature extraction techniques like trace segmentation, polygonal approximation, and wavelet transformation (WT) with k-mean clustering for ECG signals. In [16], the QRS detection algorithm's performance is affected by the processed ECG signal's frequency range. Therefore, a bandpass filter was applied to assess the algorithm's sensitivity. The findings suggest that a bandpass frequency range of 8-20 Hz offers the optimal signal-to-noise ratio for QRS detection.

P. Sasikala et al. [17] The author suggested that the physiological and geometrical variations of the heart in various people show specific uniqueness in their ECG readings. ECG is a biometric instrument that can be used to identify and validate individuals. The Discrete Wavelet Transform (DWT) is commonly used to extract the geometrical variation within ECG signals. The computational complexity increases due to matrix multiplication used in DWT [18]. One of the ways for efficient image processing can be a different technique to identify geometrical variations in ECGs.

According to V. S. Chauhan et al. [19], an altered interpretation of the slope of the ECG signal can potentially be used as a new feature in the classification set for feature extraction and classification. Dallali et al. [20] considered Heart Rate Variability (HRV) as one of the features taken into account while classifying ECG signals using Wavelet Transformation (WT) and artificial neural networks. ECG classification performance can be improved by including new characteristics that were retrieved from the ECG signal.

Mehta et al. [21] used the SVM technique to extract the QRS complex from ECG signals and the Electrocardiography (CSE) database to assess the approach standards. Apart from the small duration features of ECG, the proposed method of SVM performs better in many other cases. Hence, there is scope to apply different techniques for efficiently handling the small-duration ECG features for image processing. Victor Dan Moga et al. found [22] that although more computationally challenging than any other technique, WT is adequate for identifying physiological signals like ECG signals. C. Saritha et al. [23] reported Wavelet transformation is a valuable technique for categorizing ECG signals. However, several methodological aspects need more research. Because matrix multiplication is such a challenging process, wavelet transformation is also challenging. Pan, J. et al. [24] designed a real-time algorithm to recognize the QRS complex from ECG signals by considering slope, amplitude, and width parameters.

A novel quantized diagnostic technique for heart valve disease analysis, based on the analysis of 4 clinical heart valve sounds CSCW “Cardiac Sound Characteristic Waveform,” was suggested in [25]. The signal was acquired using a BIOPAC acquiring machine. The data is captured and then sent through an Ethernet connection to a computer, where it may be stored, analyzed, and shown in real time. An analytical model with a single degree of freedom (SDOF) was developed to isolate the distinctive waveform. In addition, diagnostic criteria were determined to differentiate between normal and diseased heart sounds through an intuitive graphical depiction, making it possible for even novice users to track their pathology's development. Furthermore, a case study of a patient with heart valve disease is presented, both before and after surgery, to prove the efficacy of the suggested approach.

In [26], the predictive ability of nine different classification strategies for heart disease and diabetes was assessed. These strategies included the decision tree, KNN, ANN, SVM, and naïve Bayesian neural network. They employed medical profiles, including blood sugar, chest pain, blood pressure, sex, and age. A methodology consisting of the Apriori algorithm and the Support Vector Machine (SVM) to predict heart disease. This method may estimate the probability of a patient developing heart disease. Because of this, the medical community is actively working to diagnose and prevent heart diseases. Analyses have shown that classification-based strategies are more successful and provide better results than traditional approaches. In [27], the authors suggest a standalone system based on DSK6713, which may detect an abnormal heart sound during auscultation. The presented method makes use of sound amplification and analysis strategies. Stochastic algorithms are used to analyze the auscultation recordings; these algorithms are validated by utilizing a library of recorded heart sounds gathered from the Michigan Heart Library. In [28], the authors reviewed the progress in predicting cardiac issues utilizing machine learning (in this case, the ANN). Miss Chaitrali's research report identifies 13 risk factors for heart disease. By analyzing the model's heart rate data, we can predict a person's risk of developing heart disease in the next year. Tools like intelligent mirrors, mice, phones, and even innovative chairs are used to collect this information.

Gawande N. et al. [29] heart diseases like Ischemic heart disease (IHD) and arrhythmia. Additionally, an ECG signal can be used to categorize heart issues. ECG displays the patient's condition and diagnoses and treats various types of heart diseases. The type of heart condition is determined by variations in the P and T wave and QRS complex in the ECG. While some cardiac conditions are mild, others can be life-threatening. Therefore, it is essential to classify cardiac diseases. The CNN algorithm can be used to categorize the different cardiovascular diseases. Khandait V. et al. [30] Digital signal filtering is applied, and SVM is used to classify the ECG data based on its features. The noise detection and feature classification techniques help improve the ECG signal with noise. Designed filters are focused on removing artifacts. Similarly, this paper includes an SVM technique that automatically identifies five distinct conditions. The classifier's accuracy on the PhysioNet ECG Database was 96.60%. Sun C. et al. [31] 12-Lead ECG acquisition equipment that is portable In this study, an innovative technique to transmit Bluetooth-based ECG data collection from the control unit to the user's mobile phone for further processing, storing, and displaying is presented. To develop a compact, wireless, and transportable ECG acquisition system, the three components are merged.

Piotrowski Z. et al. [32] proposed a Center clipping-based algorithm for heart rate detection addon with a short-term Autocorrelation approach. The algorithm analyzes biological signals like ECG, even in noisy environments with different artifacts. The algorithm recognizes different components of the ECG signal, including the PQRST complex and R peak. The paper establishes a system for analysis of heart rate variability. The HRV module uses non-parametric and parametric power spectral density computing techniques. In their work, Wang Y. et al. [33] introduced a novel approach combining short-time multifractal analysis with a fuzzy Kohonen network for arrhythmia-type detection. Heartbeat identification utilizes the short-time generalized dimension (STGD), followed by neural network-based arrhythmia-type detection. Leveraging a new fuzzy Kohonen network for classification, the proposed model surpasses traditional algorithms in accuracy, achieving an impressive 97% and above accuracy while maintaining computational efficiency. Surda, J. et al. [34] ECG signal spectral properties. The ECG signal undergoes digital signal filtering. The filters aimed to eliminate breathing muscle artifacts and supply network 50 Hz frequency. The direct methods for heart rate detection are ECG signal spectral analysis and the Short-Term Autocorrelation method. The Heart Rate detection algorithms are mainly based on the QRS complex as the distance between the R-R interval needed to calculate heart rate. QRS complex can be detected using algorithms from artificial neural networks, genetic algorithms, wavelet transforms, or filter banks.

In the diverse landscape of ECG analysis, the potential for ECG signals to serve as biometric identifiers and aid in arrhythmia classification is becoming increasingly apparent. While Discrete Wavelet Transform offers insights into unique physiological attributes, alternative methods are being explored to balance computational complexity. As the field continues to evolve, pursuing more robust and efficient techniques holds great promise in advancing ECG-based diagnostics and personalized healthcare.

3. Publicly Available Database

The MIT BIH arrhythmia database collected the physiological digitized ECG signals [35]. The database contains 48 records, and the signal is sampled at 250 Hz. Each record containing two-channel ECG signals for 30 min duration was selected from 24-hour recordings of 47 individuals. The database included 116,137 numbers of QRS complexes [36].

The categorization of arrhythmia takes into account four different kinds of databases.

1. Standard Arrhythmia ECG Data
2. CU Ventricular Tachyarrhythmia Database
3. Supraventricular Arrhythmia Database
4. Ventricular Tachyarrhythmia Database

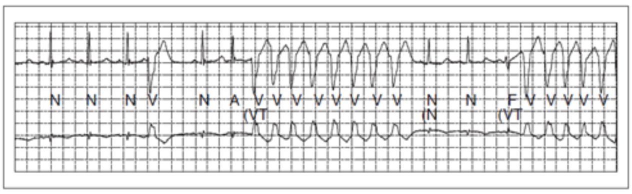

Fig. 3. Ten seconds from record 205 of the MIT-BIH Arrhythmia Database [37].

Figure 3 shows the annotation of ten seconds from a record 205 of the MIT-BIH Arrhythmia Database [37]. Each beat was rigorously observed for annotations as

- A – Atrial premature beat
- F – Ventricular fusion beat
- N – Normal beat
- V – Ventricular premature beat

Along with beat, the Rhythm annotations were also considered as

- (N – Normal sinus rhythm
- (VT – Ventricular Tachycardia

In this record, "(VT" appeared in the center, between the two ECG signals (above MLII, below V1). The online PhysioNet library converts the MIT BIH data into Matlab readable format used in this study [38]. A file in the database for each signal describes its beat, rhythm, and other characteristics. The database is used to test supervised algorithms for detecting arrhythmias and classifications. This study uses full-length ECG signals to categorize them as normal or pathological.

4. Proposed System

ECG signal analysis is vital in predicting arrhythmia in the medical domain. The ECG signal undergoes segmentation to perform QRST wave detection. Some advanced Machine Learning techniques are used to provide appropriate results and make effective decisions on signal data. Determining the machine learning classifier for the best model fitting for arrhythmia prediction is challenging. This research aims to accurately diagnose different arrhythmia types through a comparative analysis of various machine-learning algorithms. The study follows a structured prediction framework: Initially considered, ECG signal from the MIT BIH arrhythmia database is preprocessed, utilizing a comprehensive feature set for model development. Five well-known supervised machine learning algorithms are then employed to analyze and establish the optimal predictive approach. Subsequently, the experimental results demonstrate the effectiveness of arrhythmia prediction. Figure 4 shows the proposed architecture of the featured classification of arrhythmia using ECG.

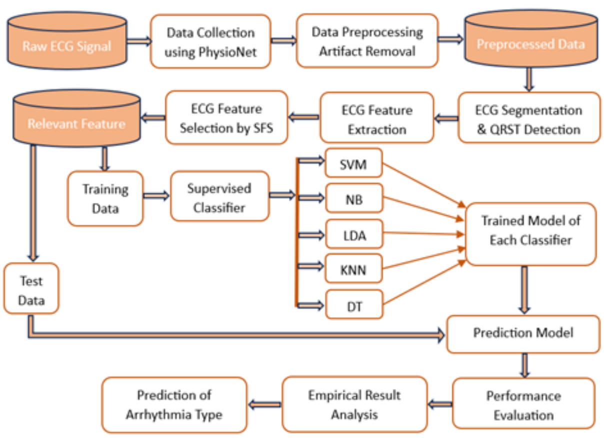

Fig. 4. Proposed Architecture of Featured Classification of Arrhythmia using ECG.

The proposed supervised prediction framework encompasses six primary stages:

1. Database Gathering and Preprocessing.
2. QRST wave Detection.
3. Feature Selection
4. Supervised Learning Classifiers
5. Generation of the Prediction Model.
6. Performance Evaluation.

This holistic approach systematically advances the understanding and accuracy of arrhythmia prediction.

4.1. Database Gathering and Preprocessing

The initial phase of the proposed model involves utilizing the raw ECG signal to detect the QRST waves. For this purpose, preprocessing the original ECG signal is essential, enabling the identification of QRS and T peaks. ECG signal preprocessing steps are as

- PhysioNet ATM is used to obtain the standard Arrhythmia ECG data from the MIT-BIH research laboratory and transform it into a Matlab-accessible format.
- Obtaining Normal Sinus ECG Data from MIT BIH laboratory and converting it into Matlab readable format using PhysioNet ATM.
- Obtaining the types of Arrhythmia ECG Data from MIT-BIH laboratory and converting it into Matlab readable format using PhysioNet ATM.

This processed signal serves as the foundation for subsequent stages. Moving forward, the model extracts features to create an input representation that optimally characterizes the original signal. The final step must divide the processed signal into regular and arrhythmia categories.

4.2. QRST Wave Detection

The state-machine logic in QRST wave detection is applied to isolate the various ECG peaks. It overcomes the noise by high-pass filtering, and low-pass filtering used to handle baseline movement further added threshold criteria to turn off spike detection. The Pan-Tompkin's algorithm was used to detect the QRST wave [24]. QRST wave detection for the sample signal is shown in Figure 5.

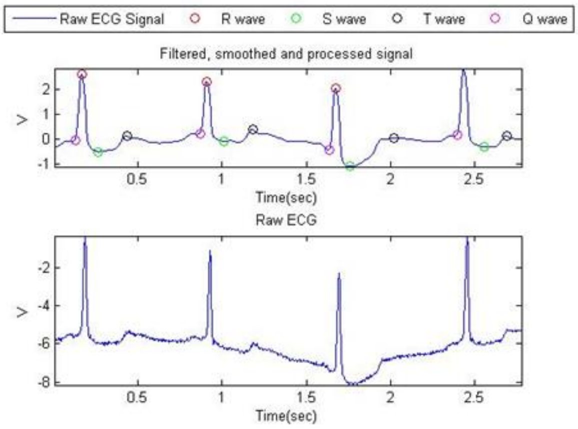

Fig. 5. QRST Wave Detection

The stepwise procedure for detecting QRST waves:

1. Noise Reduction (Filtering)
    - Cutoff low frequency to remove the baseline wander
- Cutoff frequency to wipe out high-frequency noise
    - Cutoff based on fs
    - Bandpass filtering
2. Enter state 1(putative R wave)
3. Locate R by determining the most elevated Peak.
4. See whether the Signal drops below the S wave threshold.
5. Start the S wave recognition state 3 (S detection) procedure.
6. Enter state 4 for potential T wave identification

- Determine whether the signal deviates from the mean.

7. After setting the double threshold using the most recent S wave and baseline of the signal, search for the T wave between these two values.

These steps outline the algorithm for detecting an ECG signal's R, S, and T waves. Further, these ECG components are processed to handle various signal features for arrhythmia detection.

4.3. Features Selection

A crucial step in machine learning is feature selection to identify the most relevant features from ECG signals to enhance the model's performance. Here, we consider the feature set extracted from the ECG signal. Several features are taken into consideration for classification, including.

Average Heart Rate (HR)

Average Heart Rate Variability (HRV)

Mean RR Interval

Root Mean Square of Successive Distance of R-R Interval (rmssd)

Number of R peaks in ECG that Differ More than 50 milliseconds (nn50)

Percentage NN50 (pnn50)

Standard Deviation of R-R (SD_RR)

Standard Deviation of Heart Rate (SD_HR)

Standard Deviation (std)

Mean

Kurtosis

Skewness

Root Mean Square (RMS)

Peak to Peak

Shape Factor

Crest Factor

Margin Factor

Impulse Factor

Energy

Features related to Spectral Kurtosis (SK).

SK Mean

SK Std

SK Skewness

SK Kurtosis

During this work, we have implemented a sequential feature selection method. The two parts of this strategy are as follows:

- The approach attempts to minimize a criteria function over all practicable subsets of features. Mean squared error considered for regression models and misclassification rate are standard measures for classification models.
- An evaluation procedure that adds or eliminates characteristics from a candidate subset in a sequential search. Because it is often impossible to compare the criterion value overall 2^n subsets of an n-feature data set (where n is the number of features and C is objective calls' cost), sequential searches can only do that in one direction, either expanding or narrowing the candidate set.

SFS (Sequential forward selection) is a method where features are introduced to a candidate set one at a time until doing so no longer lowers the criteria.

4.4. Supervised Classifiers

After feature selection, the arrhythmia dataset's significant features are considered input for machine learning classifiers. Selected relevant features of the arrhythmia dataset are considered to develop a prediction framework using the classification technique. We evaluated the performance using five supervised machine learning classifiers: Support Vector Machine, Naïve Bayes, Linear Discriminant Analysis, K-Nearest Neighbor, and Decision trees. The fundamentals of all these learning classifiers contributing to the development of the prediction model are discussed below:

4.4.1. Support Vector Machine (SVM)

SVM is high-performance and capable of predicting, analyzing, regressing, and classifying datasets. It is used to predict and analyze the dataset using regression and classification techniques. One supervised learning algorithm is the SVM, primarily used in data mining classification. It can perform classification by separating classes by discovering the optimal hyperplane. SVM gives the classification result by processing training data to generate a support vector machine model. Maximizing the combination between the instances of two classes can minimize the error. The SVM is flexible for datasets as linear or non-linear. It also efficiently handles large feature sets. These capabilities contribute to creating a more precise prediction model for arrhythmia detection.

4.4.2. K-Nearest Neighbours (KNN)

KNN is a supervised type of learning algorithm, and by nature, it is nonparametric. It finds from the input k-number of objects close to the exact point query or majority vote. It works based on the nearest class object with the minimum distance from the request to the training example. According to KNN, it is the fastest algorithm in its execution time to build models. After gathering all the neighbors, apply a simple majority vote to the prediction query. For each query Xn to be classified, x1, x2, … xk are the k instances. These instances are training set that is nearest to xn. The Euclidean distance is calculated.

4.4.3. Naive Bayes (NB)

Naïve Bayes has a supervised learning algorithm classification technique, a probabilistic classification learning algorithm based on applying Bayes theorem. Naïve Bayes is used for diagnosis and prediction of the problem. For Naive Bayes classifiers, every pair of classified features is independent. The Naïve Bayes algorithm requires a small amount of training data during classification to predict and evaluate the parameter. The Naïve Bayes classification method predicts an associate of each class. For instance, the probability for the specified record for the target class. The class with maximum probability is expected to be the most likely.

4.4.4. Decision Trees (DT)

Decision Tree is an efficient and commonly used technique for classification and prediction. A Decision tree is a branching-like tree structure where every node represents a record test, the outcome represents the tree branch, and the leaf node (end node) holds the target class label. It is the supervised learning classification algorithm. In the decision tree classification algorithm, the branch of data distribution is easily understandable, and every leaf is pure gain evidence from the branch. The classification algorithm uses top-down greedy search methods for producing a tree. Execution of the decision tree produces a sorting tree whose leaf denotes the ending class, and the internal attributes represent a possible number of outputs of the branch features.

4.4.5. Linear Discriminant Analysis (LDA)

LDA is applied for dimensionality reduction, hence called Normal Discriminant Analysis. It is a commonly used supervised reduction technique for classification problems called Discriminant Function Analysis. LDA performs modeling the differences between various groups in terms of classes. It is used to project the features in a higher-dimension space into a lower-dimension space. For example, we have to divide two classes with multiple features efficiently. First, consider only one feature for classification, which may result in overlapping, as shown in Figure 6.

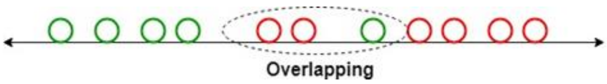

Fig. 6. Features Classification using LDA.

To resolve this, progressively incorporating more features becomes crucial for accurate classification. This iterative process aims to clarify class separation, enhancing classification accuracy for arrhythmia prediction.

4.5. Generation of Prediction Model

The proposed strategy for distinguishing arrhythmia data from standard data. The prediction model was developed with five supervised machine-learning classifiers for arrhythmia diagnosis. The prediction framework starts with the raw ECG data available for QRST wave identification. The raw ECG signal must first be preprocessed to identify the recorded ECG's Q R S T peaks to proceed to the next step. Then, the preprocessed ECG signal undergoes segmentation, which supports the detection of the QRST wave. Further stages are feature extraction or creating the input that most accurately describes the original signal and the feature selection process to establish the most significant features to optimize the performance.

The relevant features database is divided into training and test data. The five supervised classifiers are SVM, NB, LDA, KNN, and DT, which are applied to training data for the prediction model. Each algorithm follows a distinct path to train the prediction model. Each trained prediction model was evaluated using test data for performance analysis. Classifying the processed signal as usual or arrhythmia is the last phase of the prediction framework.

4.6. Performance Evaluation

The performance of the ML classifiers for arrhythmia detection was evaluated with four measures: Accuracy, precision, recall, and f-measure. Table 1 depicts measures for computing the performance parameters.

Table 1. Accuracy Measures.

		Predicted Labels	
		Class= Yes	***Class = No***
Actual Class	Class = Yes	True Positive (TP)	False Negative (FN)
	Class = No	False Positive(FP)	True Negative (TN)”

Accuracy is the ratio of correctly classified records to the total patient records.

Accuracy = (TP+TN)/(TP+FP+FN+TN)

Precision is performed by taking the ratio of correctly classified positive records to the total predicted positive samples.

Precision = TP/TP+FP

Recall (Sensitivity) - The recall is calculated by taking the ratio of indeed classified positive samples to all samples in actual class yes.

Recall = TP/TP+FN

The F1 Score is performed by taking the weighted average of precision and recall.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

5. Results and Discussion

The empirical findings of our study introduce a machine-learning approach to predicting heart disease. The framework's effectiveness is demonstrated using the MIT BIH arrhythmia database. The initial segment focuses on assessing the performance of diverse supervised machine learning algorithms, considering all features. Specifically, we employed a Support Vector Machine (SVM), Naïve Bayes (NB), Liner Discriminant Analysis (LDA), K Nearest Neighbor (KNN), and Decision Tree (DT). We evaluated the performance with sequential feature selection algorithms in the subsequent part. The framework ensures robust model assessment and minimizes potential biases in classifier performance estimation. This machine learning approach offers insights into the comparative analysis of different classifiers with all features and for selected relevant features. Hence, we consider two cases as

Case 1: Arrhythmia Prediction with All Feature (AP-All)

Case 2: Arrhythmia Prediction with Feature Selection Technique (AP-FS)

The performance is evaluated based on the prediction accuracy, precision, recall, and f-measure of arrhythmia diagnosis. Table 2. depicts the Accuracy of the classification of Arrhythmia for different supervised machine-learning algorithms considering all features compared with the following column accuracy of the learning classifier with sequential feature selection considering the most relevant features only.

Table 2. Classification Accuracy.

ML Algorithms	AP-All	AP-FS
SVM	0.892	0.8295
NB	0.9294	0.9176
LDA	0.9091	**0.9261**
KNN	0.858	**0.9091**
DT	**0.9375**	0.9091

The prediction model has an accuracy of around 93.75 percent, as shown by the obtained value of 0.937 for the Decision tree. The classification accuracy is better if we consider decision tree classification without feature selection. The performance of the LDA and KNN algorithms increases with feature selection techniques. The features selected using the sequential forward selection technique are Average HRV, nn50, pnn50, SD_RR, PSE, Kurtosis, Margin Factor, SK Mean that improved accuracies of LDA and KNN are 92.61% and 90.91%, respectively.

Figure 7. displays the average classification accuracy for supervised machine learning algorithms under consideration. As shown in Figure 7, the decision tree classifier gives the best accuracy for the arrhythmia classification, followed by the Naïve Bayes classifier with an accuracy of 92.94%.

Fig. 7. The Classification Accuracy for Arrhythmia Detection.

The performance of the LDA and KNN algorithms increases with feature selection techniques. The LDA classifier performed well for prediction with feature selection, achieving 92.61% accuracy, 81.72% precision, 79.29% recall, and 80.05% F1-score. It showed high diagnosis powers for the types of arrhythmia identification. The KNN classifier with Sequential feature selection, which had an accuracy of 90.91%, came in second place to LDA. It demonstrated a balanced performance with a precision of 76.33%, a recall of 76.37%, and an F1-score of 75.57%.

Precision is the ratio of correctly predicted ECG records to the framework's total optimistic predictions. DT also achieved a respectable precision of 82.01%, proving its ability to distinguish real positive cases from all correctly predicted positive ones. Table 3 shows the precision of the classification of Arrhythmia for different supervised machine-learning algorithms.

Table 3. Precision Measure.

ML Algorithms	AP-All	AP-FS
SVM	0.7503	0.6500
NB	0.8071	0.7794
LDA	0.7749	**0.8172**
KNN	0.6127	**0.7633**
DT	**0.8201**	0.7681

Recall is the ratio of correctly predicted ECG records by classifier relative to the total number of positive ECG records. It is also considered as the sensitivity of the prediction model. A significant fraction of the dataset's real positive cases could also be successfully identified by DT, as seen by its high recall rate of 85.44%. Table 4. shows the recall of the classification of Arrhythmia for different supervised machine-learning algorithms.

Table 4. Recall Measure.

ML Algorithms	AP-All	AP-FS
SVM	0.7882	0.6875
NB	0.8392	0.7955
LDA	0.7775	**0.7929**
KNN	0.6561	**0.7637**
DT	**0.8544**	0.7711

The harmonic mean of the precision and recall numbers is the F1 score. F1 is generally more valuable than accuracy if target classes are not equally distributed, though it is computationally challenging. The significant F1 score achieved by DT is 0.830.

Table 5. shows the F Score of the classification of Arrhythmia for different supervised machine-learning algorithms.

Table 5. F Score Measure.

ML Algorithms	AP-All	AP-FS
SVM	0.763	0.6323
NB	0.8147	0.7823
LDA	0.7662	**0.8005**
KNN	0.624	**0.7557**
DT	**0.8308**	0.7631

The proposed framework was practical for predicting arrhythmia using supervised learning approaches with feature selection, and several efficacy measures assessed the overall performance. The result analysis suggests that DT, NB, and LDA classifiers for the complete feature set, each exhibiting strengths in specific performance areas, are particularly good at recognizing arrhythmia. For the feature selection-based approach, the results of LDA and KNN offer insightful analysis for developing and applying the prediction framework of arrhythmia diagnosis.

6. Conclusion

Utilizing a sequential feature selection algorithm and QRST wave detection, this study analyzed the efficiency of a supervised learning classifier for arrhythmia diagnosis using an ECG signal. The study concerns four types of arrhythmias databases from MIT BIH data: Standard Arrhythmia ECG Data, CU Ventricular Tachyarrhythmia Database, Supraventricular Arrhythmia Database, and Ventricular Tachyarrhythmia Database. The five supervised learning classifiers, including SVM, Naive Bayes (NB), Linear Discriminant Analysis (LDA), k-nearest Neighbors (KNN), and Decision Tree (DT), are utilized in generating an arrhythmia prediction model. The QRST wave detection algorithm was implemented to extract different statistical features from the ECG signal. Shape statistical features and spectral kurtosis features are fed to the classifier for classification. The model's performance was evaluated using accuracy, precision, recall, and f measures. The accuracy of the decision tree approach is 93.75%. Using features selection, the accuracy of LDA and KNN is increased to 92.61% and 90.91 %. This paper highlights the role of machine learning in predicting arrhythmias and provides valuable information for developing a framework for efficient arrhythmia diagnostics. The findings suggest that DT is a promising candidate for accurate and reliable arrhythmia diagnosis. LDA and KNN, however, are effective when using feature selection strategies. These successes will assist researchers of the healthcare domain in selecting the optimal algorithms for early arrhythmia detection using ECG data and intervention, eventually enhancing patients' lives and overall wellness.

References

[1] T. O. Aje, “Cardiovascular disease: A global problem extending into the developing world,” World J Cardiol, vol. 1, no. 1, p. 3, 2009, doi: 10.4330/wjc.v1.i1.3.

[2] H. C. McGill, C. A. McMahan, and S. S. Gidding, “Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study.,” Circulation, vol. 117, no. 9, pp. 1216–27, Mar. 2008, doi: 10.1161/CIRCULATIONAHA.107.717033.

[3] A. T. Reisner, G. D. Clifford, and R. G. Mark, “The physiological basis of the electrocardiogram.,” in Advanced methods and tools for ECG data analysis, 2006, pp. 1–25.

[4] V. Viknesh and P. R. Prashanth, “Matlab implementation of ECG signal processing,” IOSR Journal of VLSI and Signal Processing, vol. 3, no. 1, pp. 40–47, 2013, doi: 10.9790/4200-0314047.

[5] I. Romero and L. Serrano, “ECG frequency domain features extraction: a new characteristic for arrhythmias classification,” in 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2001, pp. 2006–2008. doi: 10.1109/IEMBS.2001.1020624.

[6] D. Cuesta-Frau, J. C. Perez-Cortes, G. Andreu-Garcia, and D. Novak, “Feature extraction methods applied to the clustering of electrocardiographic signals. A comparative study,” in Object recognition supported by user interaction for service robots, IEEE Comput. Soc, 2002, pp. 961–964. doi: 10.1109/ICPR.2002.1048197.

[7] C. Levkov, G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky, “Removal of power-line interference from the ECG: a review of the subtraction procedure,” Biomed Eng Online, vol. 4, no. 1, p. 50, Dec. 2005, doi: 10.1186/1475-925X-4-50.

[8] Soo-Chang Pei and Chien-Cheng Tseng, “Elimination of AC interference in electrocardiogram using IIR notch filter with transient suppression,” IEEE Trans Biomed Eng, vol. 42, no. 11, pp. 1128–1132, 1995,

doi: 10.1109/10.469385.

[9] P. E. Trahanias, “An approach to QRS complex detection using mathematical morphology,” IEEE Trans Biomed Eng, vol. 40, no. 2, pp. 201–205, 1993, doi: 10.1109/10.212060.

[10] V. X. Afonso, “ECG QRS detection,” in Biomedical digital signal processing: C-language examples and laboratory experiments for the IBM PC, 1993, pp. 237–264.

[11] P. Valluraiah and B. Biswal, “ECG signal analysis using Hilbert transform,” in 2015 IEEE Power, Communication and Information Technology Conference (PCITC), IEEE, Oct. 2015, pp. 465–469. doi: 10.1109/PCITC.2015.7438211.

[12] R. M. S. S. Abeysekera and B. Boashash, “Time-frequency domain features of ECG signals: their application in P wave detection using the cross Wigner-Ville distribution,” in International Conference on Acoustics, Speech, and Signal Processing, IEEE, May 1989, pp. 1524–1527. doi: 10.1109/ICASSP.1989.266731.

[13] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, “Comparing wavelet transforms for recognizing cardiac patterns,” IEEE Engineering in Medicine and Biology Magazine, vol. 14, no. 2, pp. 167–173, 1995, doi: 10.1109/51.376755.

[14] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, “Clustering ECG complexes using Hermite functions and self-organizing maps,” IEEE Trans Biomed Eng, vol. 47, no. 7, pp. 838–848, Jul. 2000, doi: 10.1109/10.846677.

[15] Cuiwei Li, Chongxun Zheng, and Changfeng Tai, “Detection of ECG characteristic points using wavelet transforms,” IEEE Trans Biomed Eng, vol. 42, no. 1, pp. 21–28, 1995, doi: 10.1109/10.362922.

[16] M. Elgendi, M. Jonkman, and F. De Boer, “FREQUENCY BANDS EFFECTS ON QRS DETECTION,” in Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing, SciTePress - Science and and Technology Publications, 2010, pp. 428–431. doi: 10.5220/0002742704280431.

[17] P. Sasikala and R. Wahidabanu, “Robust R Peak and QRS detection in Electrocardiogram using Wavelet Transform,” International Journal of Advanced Computer Science and Applications, vol. 1, no. 6, 2010, doi: 10.14569/IJACSA.2010.010608.

[18] R. J. Martis, U. R. Acharya, and L. C. Min, “ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform,” Biomed Signal Process Control, vol. 8, no. 5, pp. 437–448, 2013, doi: 10.1016/j.bspc.2013.01.005.

[19] V. Chouhan, S. S. Mehta, V. S. Chouhan, and N. S. Lingayat, “Delineation of QRS-complex, P and T-wave in 12-lead ECG,” 2008. [Online]. Available: https://www.researchgate.net/publication/255646948

[20] A. Dallali, A. Kachouri, M. Samet, A. Dallali, A. Kachouri, and M. Samet, “Integration of HRV, WT and neural networks for ECG arrhythmias classification,” vol. 6, no. 5, 2011, [Online]. Available: www.arpnjournals.com

[21] S. S. Mehta and N. S. Lingayat, “Comparative study of QRS detection in single lead and 12-lead ECG based on entropy and combined entropy criteria using support vector machine,” 2007. [Online]. Available: www.jatit.org

[22] M. Moga, V. Dan Moga, L. Ceccherini-Nelli, G. Mihalas, C. Luca, and G. Mihalas, “Wavelets as methods for ECG signal processing History of Medical Informatics View project Autonomic imbalance and entropy assessment in syncope View project Wavelets as Methods For ECG Signal Processing,” 2004.

[23] C. Saritha, V. Sukanya, and Y. N. Murthy, “ECG Signal Analysis Using Wavelet Transforms,” 2008.

[24] W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” 1985.

[25] H. Yuliang, Q. Junxuan, W. Haibin, and W. Xiubo, “Quantized analysis for heart valve disease based on cardiac sound characteristic waveform method,” in 2010 2nd International Conference on Signal Processing Systems, IEEE, Jul. 2010, pp. V1-258-V1-262. doi: 10.1109/ICSPS.2010.5555596.

[26] C. Sowmiya and P. Sumitra, “Analytical study of heart disease diagnosis using classification techniques,” in 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), IEEE, Mar. 2017, pp. 1–5. doi: 10.1109/ITCOSP.2017.8303115.

[27] P. Majety and V. Umamaheshwari, “An electronic system to recognize heart diseases based on heart sounds: A stochasatic alogorithm implemented on DSK6713,” in 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, May 2016, pp. 1617–1621. doi: 10.1109/RTEICT.2016.7808106.

[28] R. Wijaya, A. S. Prihatmanto, and Kuspriyanto, “Preliminary design of estimation heart disease by using machine learning ANN within one year,” in 2013 Joint International Conference on Rural

Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T), IEEE, Nov. 2013, pp. 1–4. doi: 10.1109/rICT-ICeVT.2013.6741541.

[29] N. Gawande and A. Barhatte, “Heart diseases classification using convolutional neural network,” in 2017 2nd International Conference on Communication and Electronics Systems (ICCES), IEEE, Oct. 2017, pp. 17–20. doi: 10.1109/CESYS.2017.8321264.

[30] N. V. Khandait and A. A. Shirolkar, “ECG signal processing using classifier to analyses cardiovascular disease,” in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, Mar. 2019, pp. 855–859. doi: 10.1109/ICCMC.2019.8819777.

[31] C. Sun, J. Liao, G. Wang, B. Li, and M. Q.-H. Meng, “A portable 12-lead ECG acquisition system,” in 2013 IEEE International Conference on Information and Automation (ICIA), IEEE, Aug. 2013, pp. 368–373. doi: 10.1109/ICInfA.2013.6720325.

[32] Z. Piotrowski and K. Różanowski, “Robust Algorithm for Heart Rate (HR) Detection and Heart Rate Variability (HRV) Estimation,” Acta Phys Pol A, vol. 118, no. 1, pp. 131–135, Jul. 2010, doi: 10.12693/APhysPolA.118.131.

[33] Yang Wang, Yi-Sheng Zhu, N. V. Thakor, and Yu-Hong Xu, “A short-time multifractal approach for arrhythmia detection based on fuzzy neural network,” IEEE Trans Biomed Eng, vol. 48, no. 9, pp. 989–995, 2001, doi: 10.1109/10.942588.

[34] J. Surda, S. Lovas, J. Pucik, and M. Jus, “Spectral Properties of ECG Signal,” in 2007 17th International Conference Radioelektronika, IEEE, Apr. 2007, pp. 1–5. doi: 10.1109/RADIOELEK.2007.371653.

[35] R. G. Mark, P. S. Schluter, G. Moody, P. Devlin, and D. Chernoff, “An annotated ECG database for evaluating arrhythmia detectors,” IEEE Trans Biomed Eng, vol. 29, no. 8, pp. 600–600, Jan. 1982.

[36] M. K. Sarkaleh, “Classification Of Ecg Arrhythmias Using Discrete Wavelet Transform and Neural Networks,” International Journal of Computer Science, Engineering and Applications, vol. 2, no. 1, pp. 1–13, Feb. 2012, doi: 10.5121/ijcsea.2012.2101.

[37] G. B. Moody and R. G. Mark, “The impact of the MIT-BIH arrhythmia database.,” IEEE Eng Med Biol Mag, vol. 20, no. 3, pp. 45–50, 2001, doi: 10.1109/51.932724.

[38] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet,” Circulation, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[39] M, A. ., M, D. ., M, A. ., M, V. ., I, S. T. ., & P, K. . (2023). COVID -19 Predictions using Transfer Learning based Deep Learning Model with Medical Internet of Things . International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 43–50. https://doi.org/10.17762/ijritcc.v11i3.6200

[40] Ólafur, S., Nieminen, J., Bakker, J., Mayer, M., & Schmid, P. Enhancing Engineering Project Management through Machine Learning Techniques. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/112