IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 March 2024, accepted 5 April 2024, date of publication 10 April 2024, date of current version 29 April 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3387041

RESEARCH ARTICLE

Comparative Analysis of Machine Learning Algorithms With Advanced Feature Extraction for ECG Signal Classification

TANUJA SUBBA AND TEJBANTA CHINGTHAM
Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University (SMU), Gangtok, Sikkim 737132, India
Corresponding author: Tanuja Subba (tanujasubba69@gmail.com)

ABSTRACT Electrocardiogram is a heartbeat signal that can be used for the application of Human-computer interaction. Electrocardiography (ECG) is a prominent way to analyze heart rate and to diagnose cardiovascular disease. However, its availability has been restricted, especially in contexts with limited resources, due to the cost associated with conventional ECG signal processing equipment. The importance of ECG signal processing classification for improving early diagnoses in clinical and remote monitoring contexts is highlighted here. The dataset considered for this work is MIT-BIH arrhythmia which has 15 categories and summarized in 5 classes Normal (N), Superventricular ectopic beats (SVEB), Ventricular ectopic beat (VEB), Fusion beats (F), and Unknown beats (Q). The work discusses the importance of automated classification techniques that make it possible to analyze vast amounts of ECG data effectively and objectively. This research presents an investigation into the classification of ECG signals using various Machine Learning (ML) methods. Specifically, the performance of Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), K Nearest Neighbor (KNN), and Support Vector Machine (SVM) algorithms are examined. Among these classifiers, RF exhibits a remarkable accuracy of 98%. The results demonstrate the superior performance of the proposed approach for heartbeat classification systems.

INDEX TERMS Electrocardiogram, signal processing, feature extraction, machine learning, random forest.

I. INTRODUCTION

With the increase in health problems and the aging population, the increase in cardiovascular disease is on the brick, therefore, acquisition and classifying ECG signals are very important. Processing and classification of ECG signals are crucial for identifying cardiovascular disorders and keeping track of heart health. The cost of typical ECG signal processing methods, however, may limit their application, especially in settings with limited resources. In 1901, Willem Einthoven used a string galvanometer to build an ECG machine. He labeled the numerous deflections with the letters P, Q, R, S, and T to produce the ECG signal, which is shown in Figure 1. Today's medical research continues to produce accurate diagnostic results. For an accurate diagnosis of cardiac disease, real-time processing is important. The ECG waveform can be described by a high energy concentration in the QRS complex and a low energy concentration in the T wave and U wave. In 50% to 75% of ECGs, the T and U waves are often not discernible [1].

One of the main causes of mortality worldwide is cardiac illness. Cardiology specialists insist on early detection of cardiac problems, frequently using an ECG signal, to treat heart patients promptly and effectively. To appropriately define this form of heart disease, numerous researchers are contributing in this area. In [3] their suggested system includes a four-step gathering of the ECG signal data (from the Physionet ST-T and MIT/BIH databases), data preprocessing and denoising, feature extraction, and classification of the signal utilizing an enhanced random forest technique. Nowadays, direct ECG

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

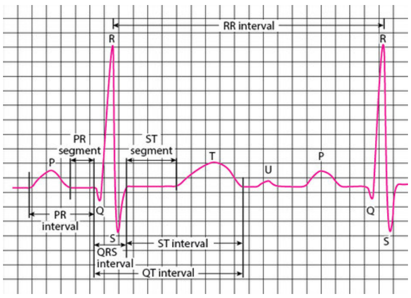

FIGURE 1. ECG signal with PQRST wave [2].

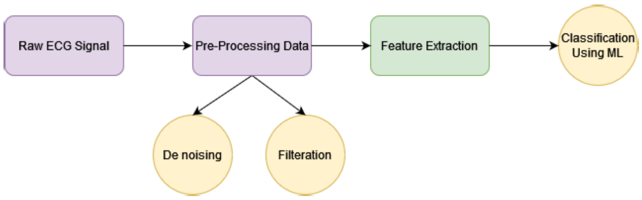

FIGURE 2. Overview of ECG acquisition and classification.

signal acquisition is done using real-time ECG monitoring instruments, which is a highly economical technology for medical purposes, some of the work is highlighted here which explains how frequent monitoring of cardiac signals helps in the analysis of any abnormalities [4]. With the increase in heart care cost the design of a cost-effective ECG monitoring system is most required, recently many researchers have designed and developed an ECG monitoring system that sends the medical report from any portable device [5]. The development of an ECG monitoring device will require an amplifier, filter, and sensor. The sensor that can be used for ECG acquisition is a non-invasive high-precision electrode that can be placed on the surface of the patient [6], [7]. The general approach to acquisition and analysis using Machine Learning (ML) is depicted in Figure 2. An input signal is taken from the MIT-BIH arrhythmia dataset, filter signal, denoise, and amplify ECG signal for analysis purposes, and Pre-processing of captured data and classification of ECG signal from normal to abnormal class is done using ML approaches. Improving the prediction of classifying will help early diagnosis of cardiovascular disease, however, considering insights from a review of pertinent state-of-the art research the metrics considered and prediction rate are less. The challenges faced in the literature are dealing with missing and highly imbalanced data, selecting a robust classification algorithm, navigating ECG signal complexity, and computational complexity, and understanding methodological limitations.

The primary contributions of this research are as follows:

- A systematic evaluation and comparison of various machine learning algorithms for ECG signal classification, providing insights into their relative strengths and weaknesses for this specific task.
- The investigation and incorporation of advanced feature extraction techniques, including wavelet transforms, HOS analysis, and FFT, to effectively capture the intricate patterns and dynamics of ECG signals, potentially improving classification performance.
- The optimization and fine-tuning of the best-performing machine learning algorithm(s), specifically Random Forests, through hyperparameter tuning, to achieve superior classification accuracy and robustness.
- The development of a reliable and accurate ECG signal classification approach, which can potentially aid in early diagnosis of cardiac disorders, remote patient monitoring, and clinical decision support systems.

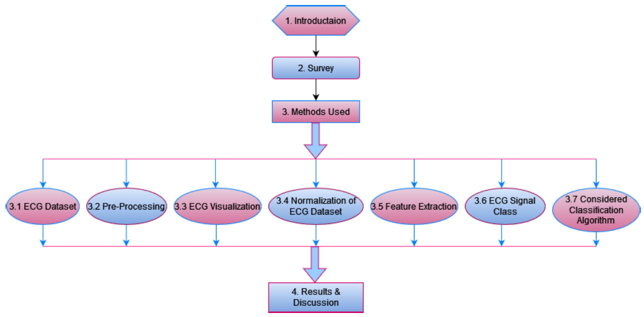

FIGURE 3. Organization of the paper.

Classification of these 5 classes is done using Machine Learning approach which is further explained in brief. The organization of this work is as follows shown in Figure 3: Section I is the introduction where ECG classification based on different models is defined, section II defines similar works, where other work based on Deep Learning (DL) is also highlighted along with some work related to low-cost ECG signal acquisition for real-time processing, section III defines methods for analyzing ECG signal form dataset collection to signal classification and section IV is results and discussion explained of the proposed approach.

II. RELATED WORKS

This review of the literature focuses on cutting-edge methods and creative strategies used in the research on efficient methods for classifying and processing ECG signals. Some previous research studies have focused on finding cardiovascular disease, particularly cardiac arrhythmia [10] where a learning-based model utilizes three separate machine learning techniques and three filter-based feature selection approaches on the cardiac arrhythmia dataset and the model chooses the best features. Random forest classifier with the gain ratio feature selection approach and a subset of 30 features had the greatest accuracy of 85.58%. Similarly, in [11] the work is based on the classification of automatic detection of arrhythmia with an accuracy of 99.84% using a discrete wavelet transform (DWT) heartbeat. Another work has been proposed using the evolutionary neural system for classifying 17 different myocardium dysfunctions with an accuracy of 98.8% [12]. Many research projects have been done using Deep Learning frameworks to develop a novel approach to automatically detecting myocardial infarction (MI) using ECG signals with and without noise removal. The Convolutional Neural Network (CNN) method for the automated detection of normal and MI ECG beats achieved an average accuracy of 93.53% and 95.22%, respectively [13]. The proposed algorithm is beneficial in clinical settings to assist doctors in making the diagnosis of MI since it can reliably detect unknown ECG signals even in the presence of noise. Similar to this, improving the performance of such models with an accuracy of 99.6% by developing a novel hybrid hierarchical attention-based bidirectional recurrent neural network with dilated CNN (HARDC) technique for arrhythmia classification is proposed [14]. This addresses issues that are brought about when conventional dilated CNN models fail to consider the relationship between contexts and gradient dispersion.

Multi-scale Convolutional Transformer Network (MCT-net), a unique method for ECG signal classification, is presented in [15]. It combines the best features of both architectures, using Transformer to extract global features through introspective processes and CNN to extract local features. Another study [16] evaluates several deep-learning techniques for the diagnosis of cardiovascular disease (CVD) using electrocardiograms (ECG). It evaluates multiple ECG signal coding methods and uses multimodal fusion methods to improve prediction accuracy. The study uses the PTB-XL ECG dataset, which contains 21,837 records and labels for four CVDs. The 1D-ECG representation performs better than multimodal models and image-based techniques. With a sensitivity of 79.67% and a specificity of 81.04%, the most effective model is GRU. In [17] the use of CNN to analyze ECG data without signal transformation or feature extraction is done. The project attempts to classify ECG images of individual heartbeats using CNNs and the Taguchi method. All fifteen types (five classes) in the MIT-BIH arrhythmia dataset are included in the study. The classification achieved an accuracy of 96.79%. Another study based on a deep learning-based system that detects irregularities in ECG data and uses that information to predict arrhythmias and heart failure combines Long Short Memory (LSTM) networks and CNN [18]. In [19] the proposed approach extracts features directly from input heartbeats using CNN. The Synthetic Minority Oversampling Technique (SMOTE) is used to solve class imbalance problems in the training dataset. The method achieves an average accuracy of 98.63%, an accuracy of 92.86%, a sensitivity of 92.41%, and a specificity of 99.06% when accurately classifying five different heartbeat types.

Another work [20] suggested PSO-SVM optimized with Independent Component Analysis and a Genetic Algorithm, for categorizing ECG arrhythmia signals MIT-BIH Arrhythmia database. The SVM classifier is optimized using a Genetic Algorithm and Particle Swarm Optimisation. The findings demonstrate that, with a classification accuracy of 96%, the hybrid classifiers PSO-SVM-ICA and G-ICA outperform PCA, ICA, and PSO-SVM with ICA and G-ICA in terms of performance measures. Deep learning-based models have a problem maintaining accuracy on large-scale ECG data, for this, a proposed deep learning-based signal quality classification model is applied to dynamic monitoring [21]. To give an overview of another aspect of ECG signal monitoring, a wireless device can also be used that is flexible and helps in acquiring biosignal in real-time, [22], [23] suggests a wearable wireless sensor system that identifies arrhythmia using ML techniques to categorize them as healthy, non-healthy, or not specified. With early access to hospital support systems and less congestion in hospitals, the system can identify anomalies and arrhythmia disorders in their early phases. The use of ensemble approaches has grown in popularity recently across several fields, including biomedical signal processing. Although the application of ensemble techniques to the classification of ECG signals has been widely studied, learning from comparable approaches used in peptide analysis and disease prediction could offer important new insights. In this section, several relevant approaches are reviewed, including iAtbP-Hyb-EnC, AFP-CMBPred, pAtbP-EnC, iAFPs-EnC-GA, and pAVP-PSSMDWT-EnC, which employ ensemble techniques in the context of different diseases. In [24] to improve prediction accuracy, the iAtbP-Hyb-EnC model, which has been developed for the identification of antitubercular peptides, applies an ensemble learning strategy. As evidenced by its improved accuracy over current predictors, iAtbP-Hyb-EnC achieves robust performance by merging various classification algorithms and feature representation techniques. Similarly, in [25] AFP-CMBPred uses an ensemble-based framework to predict anticancer peptides. AFP-CMBPred provides enhanced predictive capacities by utilizing ensemble learning approaches, hence advancing the field of anticancer peptide research. In [26] the pAtbP-EnC model is suggested for precisely identifying antitubercular peptides in the field of peptide analysis. By using an ensemble method, pAtbP-EnC improves prediction accuracy by addressing the issues brought about by the fast expansion of peptide samples. Furthermore, the pAVP-PSSMDWT-EnC and iAFPs-EnC-GA models in [27] are made to find peptides linked to particular biological activities, like antiviral and antifungal characteristics. These models precisely predict peptide functionality by extracting features from sequential and evolutionary descriptors through the use of ensemble learning techniques. By drawing comparisons between these approaches and the classification of ECG signals, ensemble techniques could enhance the precision and dependability of ECG-based disease diagnosis. Ensemble models provide a strong foundation for handling the intricacies involved in ECG signal analysis by combining several feature extraction techniques and classification algorithms. By conducting this comparative investigation, this study tries to improve the field of ECG signal categorization and aid in the creation of useful diagnostic instruments for cardiovascular disorders by utilizing knowledge from adjacent fields.

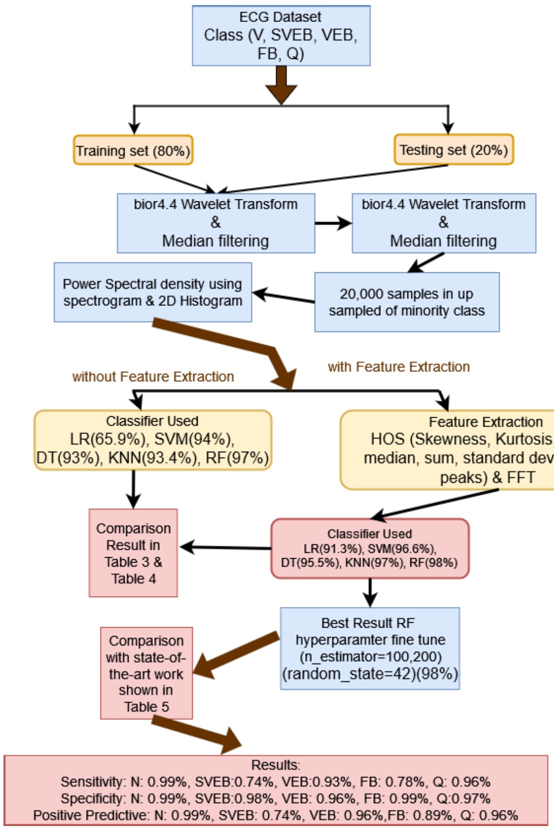

FIGURE 4. Methodology of a proposed approach.

TABLE 1. Dataset classification.

Class Name	AAMI Class	MIT-BIH Class
N	Normal Beat	Normal Beat Left Bundle Branch Block Beat Right Bundle Branch Block Beat Atrial Escape Beat Nodal Escape
SVEB	Superventricular Ectopic Beat	Artial Premature Beat Aberrated Atrial Premature Beat Superventricular Premature Beat Ventricular Escape Beat
VEB	Ventricular Ectopic Beat	Premature Ventricular Contraction Ventricular Escape Beat
FB	Fusion Beat	Fusion of Ventricular and Normal Beat
Q	Unknown Beat	Paced Beat Fusion of Paced and Normal Beat Unclassified Beat

TABLE 2. The training and Testing dataset's executive summary.

Class Name	Training Set	Testing set	Percentage%
N	72471	18115	82.8%
SVEB	2223	556	2.5%
VEB	5788	1448	6.6%
FB	641	162	0.7%
Q	6431	1608	7.3%
Total	87554	21889	

III. METHODS

The dataset considered is MIT-BIH Arrhythmia and various ML techniques that are used for the categorization of datasets. Figure 4 shows the entire methodology of a proposed approach.

A. DATASET DESCRIPTION

The training and testing of the model are first done using the MIT-BIH database, this dataset contains 48 annotated ECG recordings from 47 different subjects for 30 minutes and is sampled at the frequency of 360 Hz [8]. Continuous measurement of heart activity is done where annotations for various arrhythmias are included. This dataset has widely been used for developing classification algorithms for the detection of various heartbeat types and heart rate variability. There are a total of 15 classes or heartbeat types of arrhythmia classification: Normal rhythm beats (N), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Artial Premature Contraction (APC), Premature Ventricular Contraction (PVC), Ventricular Escape Beat (VEB), Nodal Escape Beat (NEB), Atrial Fibrillation (AF), Supraventricular Tachycardia (SVT), Ventricular Tachycardia (VT), Fusion of Ventricular and Normal Beat (FV), Fusion of Paced and Normal Beat (FP), Atrial Flutter (AFL), Second-Degree Atriventricular Block (AVB) and Unknown Beat (U). These are the different and wide range of heartbeat rhythm patterns that provide a dataset for analysis and classification algorithm development. Researchers [9] have described these 15 different ECG classes using the AAMI (Association for the Advancement of Medical Instrumentation) standard into 5 different classes that are N, SVEB, VEB, F, and Q. Table 1 shows the classification of the dataset by AAMI for the MIT-BIH dataset.

B. ECG DATASET

The ECG dataset is divided into training and testing sets with a ratio of 80:20. The description of different class data is shown in Table 1. Class-based assessment is employed. In previous research, [9], [28] [29], [30] [31] numerous studies have yielded promising outcomes in the classification of heartbeat segments based on the class of arrhythmia using the MIT-BIH Arrhythmia Database. This database contains a substantial amount of annotated heartbeat data, comprising a total of 109,446 labeled instances across five distinct arrhythmia classes. To facilitate model evaluation and performance assessment, the dataset was divided into training and testing subsets, following a ratio of 80:20. The heartbeat segments were acquired through meticulous annotation of beat locations, ensuring an accurate representation of the underlying arrhythmia patterns. Table 2 shows the traning and testing set of all class.

C. PRE-PROCESSING OF ECG SIGNAL

The dataset may contain noises like baseline drift, artifacts noise, and redundant data, so to denoise it is necessary to

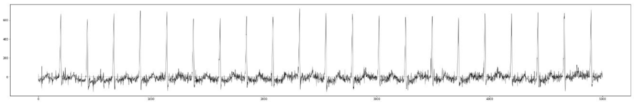

FIGURE 5. Original ECG signal.

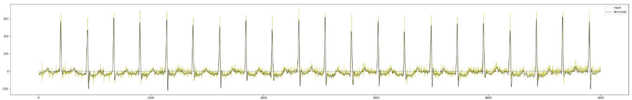

FIGURE 6. Raw ECG Vs denoised ECG signal.

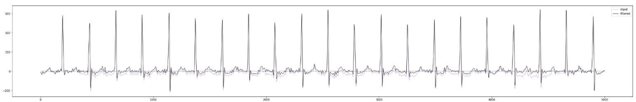

FIGURE 7. Filtered ECG signal.

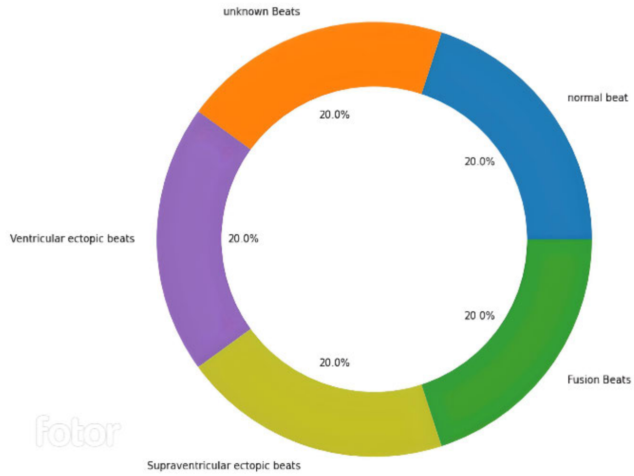

FIGURE 8. Class normalization of ECG dataset.

pre-process the data. Figure 5 shows the original ECG signal from the MIT-BIH dataset, which needs pre-processing.

To denoise the signal a wavelet transforms specifically 'bior4.4 wavelet' [32] is used where the input signal is decomposed into wavelet coefficients, and a nine-level wavelet decomposition is performed to maintain the signal no matter how much the signal fluctuates. The coefficients from 0 to cutoff low are set to zero similarly a coefficient from cutoff high to the end is set to zero hence removing the corresponding frequencies in the reconstructed signal. Figure 6 shows the input ECG signal with a black line and the denoised signal with a yellow line.

Multiple median Filtration of different widths is used for baseline fitting [33]. This approach aims to estimate and remove the baseline or low-frequency values from the signal, keeping only the signal of interest. Figure 7 shows the difference between the given ECG signal in the purple line and the filtered signal in black line.

D. NORMALIZING ECG DATASET

The considered dataset, MIT-BIH contains an imbalance of data that may suffer from misclassification because of the majority class. In the dataset, the maximum number of data approximately 82% belongs to the normal class, which may lead to severe causes in medical applications. Further, to normalize the data, following steps are used:

1) Divide the initial dataset into various subgroups according to the class:
   - Produce a subset that includes samples with values of N and so on.
2) Increase the sample size of the minority class subsets to match that of the majority class:
   - Create upsampled copies of each minority class subset by randomly selecting samples from each subset and replacing them.
   - State that there are 20,000 samples in the upsampled subsets for each class.
   - To ensure the reproducibility of the upsampled subsets, use a separate random state.
3) Combine the majority class subset and the upsampled subsets which creates a new dataset.

By boosting the number of samples in the minority classes through upsampling, this algorithm aims to correct the dataset's class imbalance, as shown in Figure 8. Ensuring that all classes are represented equally in the dataset, can help machine learning models that were trained on uneven data to perform better.

E. ECG VISUALIZATION

The overall distribution of the data and the frequency content of the signal over time can be seen in great detail on the spectrogram which is the strength of a signal with a frequency range in Figure 9. The length of each segment was set to 256 samples, and the overlap between segments was set to 128 samples for the computation of the spectrogram. To make power changes more visible, the spectrogram was plotted using (1)

$$logarithmicscaling(10 * np.log10(trainingdata)). \quad (1)$$

The strength of the frequency components at various time intervals is shown by the colour intensity in the figure with the help of the spectrogram visualization, it shows frequency content changes over time and acquires an understanding of the underlying dynamics and patterns of the signal.

A 2D histogram is one way to graphically represent the values of the distributed pixels within an ECG signal. It offers information on the frequency and occurrence of particular intensity levels across various signal regions. The y-axis displays the intensity values of the signal inside each interval, and the x-axis divides the ECG signal into discrete time intervals (bins). The bin's colour represents the number of occurrences or frequencies of various intensity values inside each time interval, which is used to produce the histogram. This makes it possible to see how the intensity values are

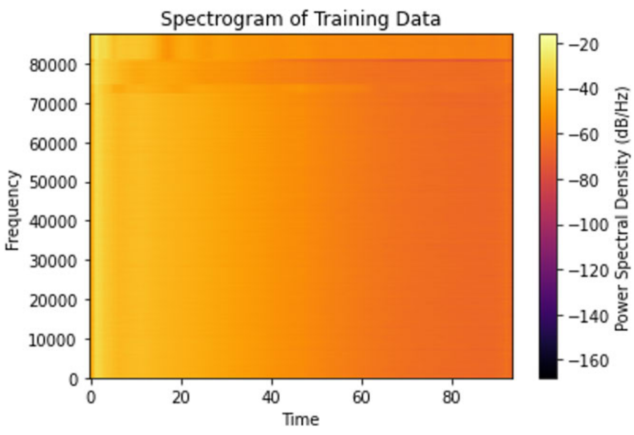

FIGURE 9. Power spectral density of training data.

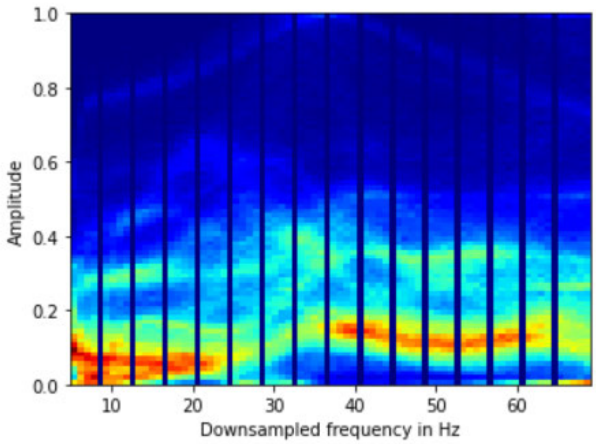

FIGURE 10. Class N.

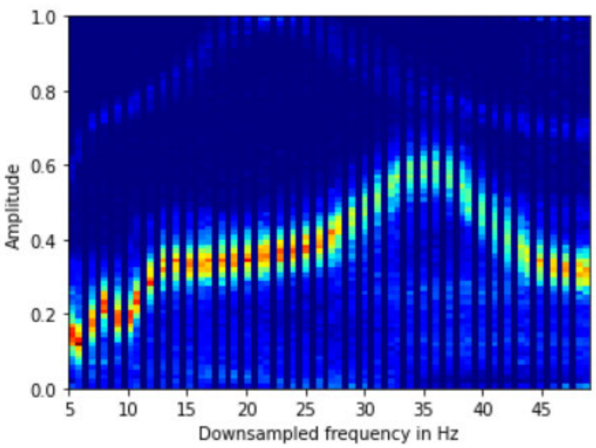

FIGURE 11. Class Q.

distributed throughout the ECG signal. It can aid in locating dominating intensity values or clusters shown in Figure 10 to Figure 14 for each class, evaluate the signal's symmetry or variability, and highlight any outliers or odd intensity patterns as done in the research study [34].

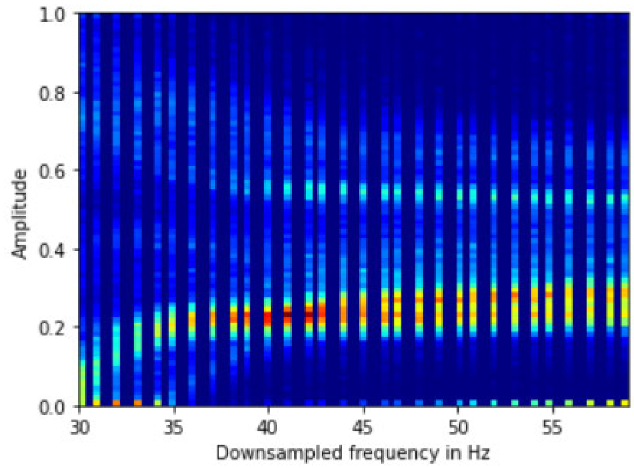

FIGURE 12. Class VEB.

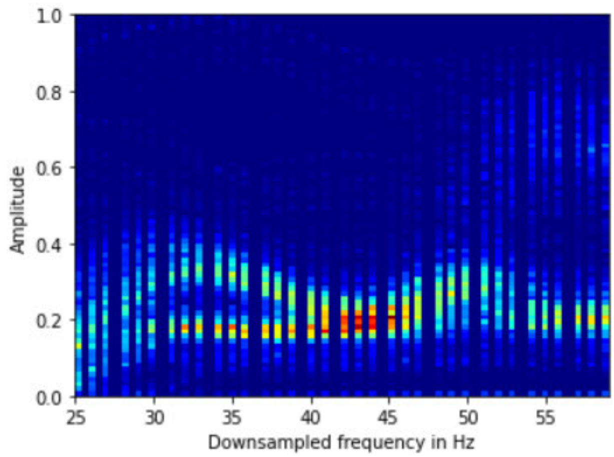

FIGURE 13. Class SVEB.

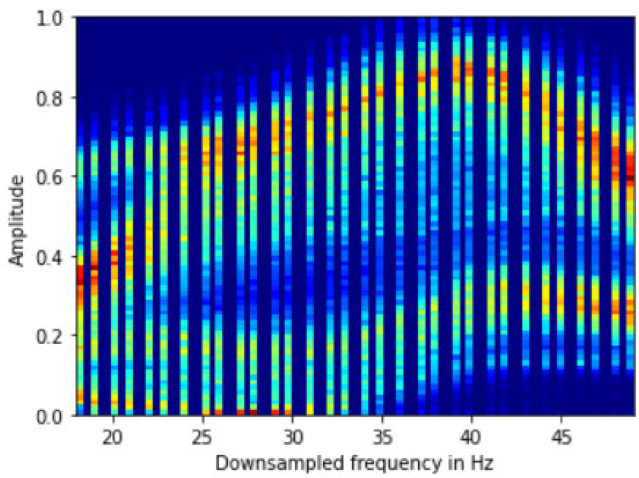

FIGURE 14. Class FB.

2D plots of 5 types of ECG segments class N, Class Q, Class VEB, Class SVEB, class FB which show the intensity distribution throughout the ECG signal.

F. FEATURE EXTRACTION

The amplitude, temporal, statistical, and frequency-based feature of ECG is extracted for analysis of the signal. R-peaks and their corresponding T-peaks, amplitude, QRS duration, and overall signal statistics are explained below:

1) Initialize an empty list called features to store the extracted features.
2) For each sample in the dataset:
    - To find an asymmetry of the distribution of ECG signal skew is calculated.

$$\text{Skewness} = \frac{\sum((X_i - \bar{X})^3)}{n \cdot \sigma^3} \tag{2}$$

    - To find the shape and distribution of the ECG signal

$$\text{Kurtosis} = \frac{\sum((X_i - \bar{X})^4)}{n \cdot \sigma^4} - 3 \tag{3}$$

    where X_i represents the individual observations, $\bar{X}$ is the mean of the observations, n is the sample size, and σ is the standard deviation.
    - For R-peak calculation peakutils.indexes() is used where peak detection is based on threshold=0.5 and minimum distance =100
    - For QRS duration it is calculated using np.diff()
3) Statistical metrics such as (mean, median, sum, and standard deviation) from the R-peak amplitudes.
4) Finally Fast Fourier Transform (FFT) converts the signal from the time domain to the frequency domain.

G. ECG SIGNAL CLASS

As discussed in the introduction, there are 5 classes for ECG signals normal (N) and abnormal (VEB,SVEB,FB,Q), after preprocessing and extracting features of individual signals it is classified as follows:

1) NORMAL (N)

The regular electrical activity of the heart during normal sinus rhythm is referred to as the ''normal beat class'' as shown in Figure 15 in the ECG readings. It depicts the typical pattern of cardiac depolarization and repolarization that take place in a healthy person devoid of any obvious abnormalities or arrhythmias [35], [36]. The following traits define the typical beat class:

- Regular Rhythm: Heartbeats that are considered normal have a regular rhythm and a constant pause between each beat.
- P Wave: A P wave, symbolizes the depolarization of the atria and is present during a regular beat. Before the QRS complex, there is a tiny upward deflection called the P wave.
- QRS Wave: The depolarization of the ventricles is represented by the QRS complex during a typical heartbeat. It is made up of the Q, R, and S waves. The QRS complex typically has a short duration and a limited width.

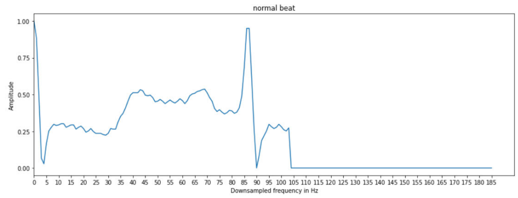

FIGURE 15. Normal heart beat.

- T Wave: The repolarization of the ventricles is represented by the T wave in a typical beat. The waveform that follows the QRS complex is often symmetrical and smooth.
- Heart Rate: The average adult's heart rate at rest is between 60 and 100 beats per minute when it is beating normally. Depending on elements like age, level of physical activity, and general health, it could differ slightly.

2) UNKNOWN BEAT (Q)

In ECG signals, a category known as the ''unknown beat class'' designates a beat whose precise nature or classification is unsure or ambiguous [36], [37]. It reflects situations where the ECG waveform either exhibits characteristics that make it difficult to assign to a single group or does not match any specified classes shown in Figure 16. It has the following characteristics:

- Morphological patterns with ambiguity: ECG waveforms that display morphological characteristics that differ from regular beats but do not match any specified aberrant patterns frequently fall under the category of unknown beats. These beats might have special qualities or special traits combined, making it challenging to categorize them precisely.
- The diversity of unidentified beats: The unidentified beat class may be diverse, including a variety of ECG anomalies that don't meet predetermined standards.
- Arbitrary classification: Since different experts or algorithms may interpret the same waveform in different ways, determining the class of an unknown beat might be arbitrary.
- Quality of Data and Noise: Unknown beats may be present in ECG recordings that are noisy or of poor quality. Accurate categorization can become increasingly difficult due to signal artifacts, electrode placement difficulties, or other sources of interference.
- Insufficient Reference Data: There may not be much of the unidentified beat class in the literature or datasets that are already available.

3) VENTRICULAR ECTOPIC BEAT (VEB)

In ECG signals, aberrant electrical impulses that come from the ventricles and cause premature ventricle contractions outside of the regular sinus rhythm are referred to as

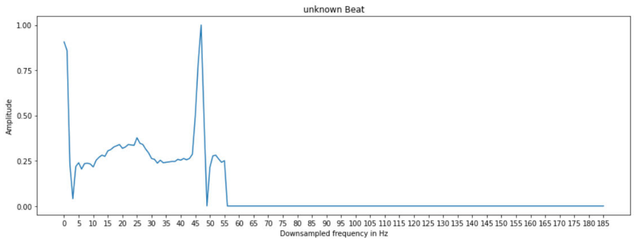

FIGURE 16. **Unknown beat.**

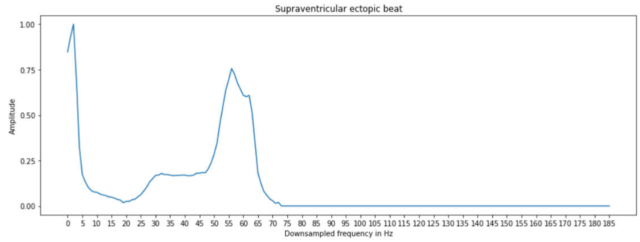

FIGURE 18. **Superventricular Ectopic beat.**

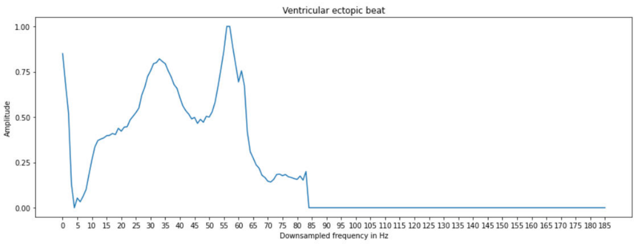

FIGURE 17. **Ventricular Ectopic beat.**

VEB [37], [38]. The VEB class is shown in Figure 17 and is described here, along with some of its distinguishing features:

- Ventricular Premature Depolarization: It occurs earlier in the cardiac cycle than anticipated, breaking the regular rhythm. They are distinguished by an early QRS complex that indicates ventricular depolarization and can be seen before the following anticipated normal beat.
- Broad QRS Complex: Compared to normal beats or other aberrant beats, the QRS complex in VEB is broader and distinct. Due to the aberrant conduction, the enlarged QRS complex represents a delayed or improper activation of the ventricles.
- Unusual T Wave: In contrast to regular beats, the T wave in VEB may have an opposing polarity or morphological abnormalities.
- Restitutionary Pause: A compensating pause follows a ventricular ectopic beat to give the heart time to regulate its electrical activity and resume the regular rhythm. This gap in the ECG data, which is longer than the typical pauses between normal beats, may be seen.
- Periodic or Recurrent Occurrence: Ectopic beats in the heart's ventricles can happen sometimes intermittently or regularly.

4) SUPRAVENTRICULAR ECTOPIC BEAT (SVEB)

The aberrant electrical impulses that come from the atria or the atrioventricular (AV) node and cause premature contractions that take place outside of the regular sinus rhythm are referred to as SVEB in ECG signals [37], [39]. The SVEB class shown in Figure 18 is described along with some of its features:

- Atrial pre-depolarization: It occurs earlier in the cardiac cycle than anticipated, breaking the regular rhythm. They can be identified by an early P wave or atrial depolarization.
- Narrow or regular QRS Complex: The ventricular depolarization is represented by the QRS complex, which is normally normal. Given that the premature beat originates from the atria or the AV node, the narrow QRS complex shows that electrical conduction across the ventricles is normal.
- Modifications to P Wave Morphology: Variations in the P wave's amplitude, length, and shape, which reflect the aberrant atrial depolarization brought on by the ectopic beat.
- Restitutionary Pause: SVEB followed by a compensatory pause before the regular beat, just like VEB. This delay, which appears as a gap in the ECG recording, enables the heart to reset may be seen.
- Periodic or Recurrent Occurrence: SVEB can occur sporadically or infrequently, known as isolated PACs, or they can happen frequently, referred to as supraventricular ectopy.

5) FUSION BEAT (FB)

In ECG signals, the term "fusion beat class" refers to a special kind of beat that happens when a regular sinus beat and a VEB overlap in the cardiac cycle, creating a composite waveform that demonstrates traits of both beats. FB happens when the normal conduction channel and the abnormal pathway connected to the VEB are both activated at the same time [28], [37]. Figure 19 shows FB and some of its features are:

- Combination of Normative and Pathological Features: FB combines elements of both regular sinus and VEB in their appearance. A normal P wave, which represents atrial depolarization, a normal or narrow QRS complex, which represents ventricular depolarization, and an altered T wave, among other anomalies, may all be seen in the waveform.
- Modular Morphology: Depending on the timing and level of overlap between the regular sinus beat and the VEB, the morphology of FB can vary greatly. The resulting waveform is distinctive and can be distinguished from other beat patterns because it may have blended.
- Transitional Timing: It happens when the time of the VEB and the regular sinus beat within the cardiac

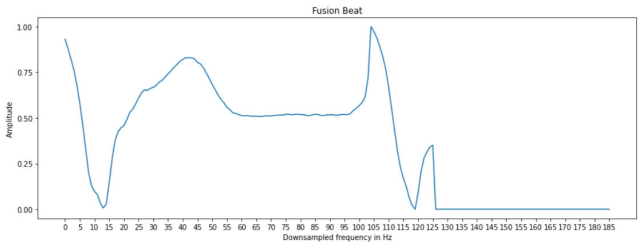

FIGURE 19. Fusion beat.

cycle coincide. During a particular phase, the electrical activation coming from both pathways unites, fusing the two waveforms.

- Fusion Complex Exists: It is represented by the fusion complex, which combines both typical and atypical characteristics. Morphological alterations are possible, a typical P wave followed by a QRS complex with characteristics modified by the VEB.

H. CLASSIFICATION USED

The classification of the data varies depending on the number of factors involved and is contingent upon the desired outcome of the research. To determine the most suitable approach for ECG classification, machine learning classification techniques such as Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression are employed.

1) DECISION TREE

Several straightforward and interpretable decision rules make up the tree structure. In the medical area, where researchers and practitioners frequently need an understanding of the decision-making process, this might be helpful [40]. DT can automatically identify pertinent features from the input data that relates to the morphology, rhythm, or timing of heartbeats. This can aid in accurately classifying ECG signals by highlighting key features. This algorithm may efficiently mimic the nonlinearities that ECG signals can exhibit, which include complicated patterns and changes. DT algorithm is capable of efficiently handling enormous datasets or when real-time classification is necessary. DT can be merged with other algorithms to create ensemble approaches like Random Forest or Boosting, which enhance classification performance. The following parameter is set during simulation.

- Criterion to measure the quality of split is "gini"
- Splitter = "best" which chooses the best split point
- max_depth = 100
- min_sample_split (min number of samples required to perform a split) =2
- min_sample_leaf (minimum size of a leaf node to avoid further splitting) = 1
- max_fetaure = "none" all features are considered equall
- class_weight = "none" all classes are treated equally

Researchers have shown that decision tree is one of the most widely used classification techniques for ECG signals.

2) LOGISTIC REGRESSION

Binary classification problems are ideally suited for logistic regression. It is frequently necessary to distinguish between normal and pathological ECG patterns when classifying ECG signals. Logistic regression uses a linear decision boundary to represent the association between the input features (ECG signal properties) and the target class's log odds. Despite the complexity of ECG signals, logistic regression can still be useful when the distinction between classes can be roughly represented by a linear function. It captures the linear correlations between class probabilities and characteristics. Logistic regression is computationally effective and is capable of handling huge datasets with a variety of features. ECG signal databases are frequently large and contain a lot of properties or features that were taken from the signals. Such multidimensional data can be handled by logistic regression, which can produce accurate classification results [41]. The parameters considered during the simulation are as follows:

- random_state = 0 (same sequence of random number is generated)
- penalty "L2" for regularization
- Solver "lbfgs" for multi classification problem
- max_iter =100 (maximum epochs solver will run to optimize the LR model)

While logistic regression is widely employed as a classification technique, researchers often favor other machine learning approaches for ECG classification due to its inherent linearity assumption, dependence on manually engineered features, and susceptibility to overfitting.

3) SUPPORT VECTOR MACHINE

SVM handles nonlinear data, resistance to noise, and capacity for handling high-dimensional data. Through the use of kernel functions, SVMs can capture nonlinear correlations between input characteristics and class labels. SVMs may successfully differentiate several classes of ECG signals by translating the input data into a higher-dimensional feature space. ECG signals are frequently represented by a large number of characteristics or data points, or high-dimensional data. SVMs are effective at handling high-dimensional data, even when there are more features than samples. They construct the decision boundary using a subset of the training data known as support vectors, which aids in preventing overfitting and lowering computational costs.

Electrical interference, muscular movement, and improper electrode placement can all cause noise to interfere with the ECG data. Compared to some other classifiers, such as neural networks, SVMs are less sensitive to noise. SVMs are more resistant to noisy data points because of the maximum margin concept, which seeks to identify the decision border that maximizes the separation between various classes [42]. The parameters that are considered during the simulation are:

- svm_model: Creating a pipeline for SVM classification using SVC.
- crossvalidation = 5 for feature selection
- params-grid: hyperparameter combinations for the SVM Classifier The dictionaries contain the following key-value pairs:
  - parameters for different SVM kernels, such as 'rbf', 'linear', 'poly', and 'sigmoid'.
  - Define values for other hyperparameters like 'C' and 'gamma'
- For each SVM model type (linear, polynomial, RBF, sigmoid):
  - Create an SVM model with the best hyperparameters obtained from grid search which is rbf SVM.

This explains the effects of various percentile values on feature selection, views the performance of the classifier, and identifies the most effective SVM hyperparameters. This iterative method aids in improving the classification model's accuracy and ability to generalize to new data.

4) K-NEAREST NEIGHBOR (KNN)

ECG signals frequently display intricate patterns and non-linear connections between the input features and the appropriate classes. KNN is a non-parametric technique that successfully captures non-linear decision boundaries without relying on presumptions regarding the distribution of the underlying data. The classification of ECG signals may include the use of numerous features that have been derived from the signals. High-dimensional feature spaces can be handled by KNN without incurring a large processing burden [43]. The parameters that are considered during the simulation are:

- knn = KNeighborsClassifier(n_neighbors=5)
  n_neighbors: The number of neighbors to consider for classification is set to 5

5) RANDOM FOREST

Random Forest is a prominent algorithm used for ECG classification. RF is frequently used for ECG categorization for the following reasons: Robustness to Noise, ECG signals are complicated and nonlinear by nature. The non-linear correlations between the input features and the target classes can be captured by RF. It can simulate intricate decision boundaries, enabling precise ECG signal classification. RF is a technique for ensemble learning that blends various decision trees to produce predictions. A random selection of features and samples is used to train each tree, which lowers the possibility of overfitting and boosts generalization efficiency. Due to the ensemble method, RF is more resistant to noise and variability in ECG readings. This algorithm is capable of processing high-dimensional data with numerous features, which is frequently the case in the categorization of ECG signals. With minimal computational overhead, it can effectively handle feature sets with hundreds or thousands of features [44]. The parameters used for the simulation are:

TABLE 3. The summary of the performance of a ML model before feature extraction for overall class N, SVEB, VEB, FB, Q.

Model	Accuracy%	Precision %	Recall %	F1-score %
Decision Tree	0.93	0.712	0.835	0.763
Logistic Regression	0.659	0.442	0.762	0.472
Random Forest	0.974	0.912	0.873	0.892
Polynomial SVM	0.919	0.958	0.919	0.933
rbf SVM	0.94	0.964	0.94	0.948
KNN	0.934	0.716	0.8968	0.78

- n-estimators (number of decision tree to be used) = 100
- max-depth = None (no limit on the depth of tree)
- min-sample-split (min number of samples required to perform a split) =2
- min-sample-leaf (min size of a leaf node to avoid further splitting) = 1
- max-feature = "auto" or "sqrt" square root of the total number of features is considered at each split
- = "true"

By enabling the evaluation of feature importance, tree architectures, and prediction routes, RF offers insights into the decision-making process. RF is a flexible and strong algorithm that can manage the difficulties associated with ECG categorization. It is a well-liked option for analyzing ECG signals and precisely classifying them into several classes.

IV. RESULTS AND DISCUSSION

In this section, we will discuss the result of our experiment. Furthermore, for further analysis and classification, various ML models have been employed for the accurate classification of ECG signals. For ML the parameters used for the performance evaluation are accuracy, precision, recall, specificity, positive predictive value (ppv) for all classes that are calculated by the values of True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP) the value of F1-score is calculated by (4)

$$F1 = 2 * (precision * recall)/(precision + recall) \tag{4}$$

This research presents a comparative analysis of various supervised learning algorithms, including DT, SVM, KNN, logistic regression, and RF, for the classification of ECG signals. The classification performance of these algorithms was evaluated before the feature extraction process, as depicted in Table 3. The findings shed light on the effectiveness of different algorithms in handling ECG signal classification tasks prior to feature extraction.

Random forest demonstrated promising results, suggesting its suitability for ECG signal classification tasks. This study focuses on the effective classification of ECG signals into normal and pathological classes using the Fast Fourier Transform (FFT) as a feature extraction technique. FFT is used to extract pertinent frequency-domain characteristics, improving classification performance. The findings show that FFT-based feature extraction makes a major contribution to the trustworthy diagnosis of heart diseases by accurately

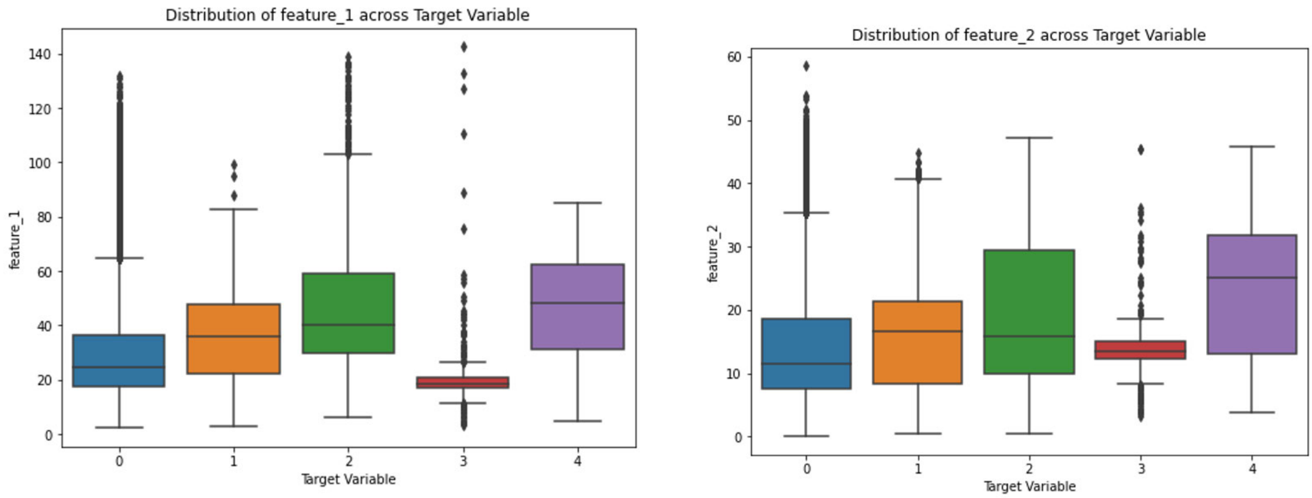

FIGURE 20. Box plot of a magnitude spectrum of different target variable class (N,SVEB,VEB,FB,Q).

TABLE 4. The summary of the performance of a ML model after feature extraction for overall class (N,SVEB,VEB,FB,Q).

Model	Accuracy%	Precision %	Recall %	F1-score %
Decision Tree	0.955	0.806	0.806	0.808
Logistic Regression	0.9130	0.792	0.55	0.618
Random Forest	0.98	0.972	0.812	0.878
rbf SVM	0.966	0.944	0.756	0.822
KNN	0.97	0.934	0.836	0.878

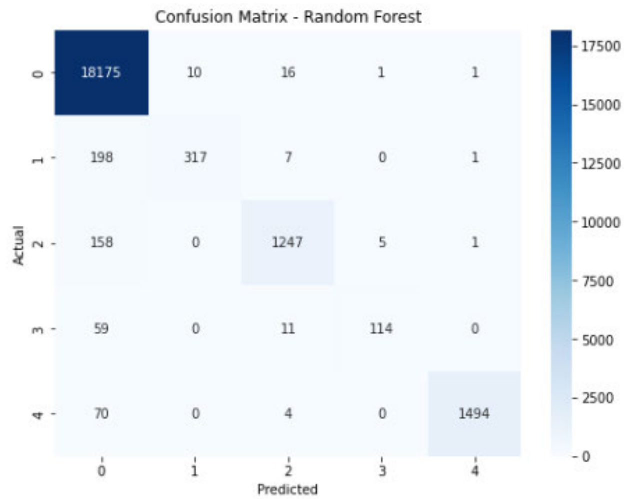

FIGURE 21. Confusion Matrix of RF classifier.

classifying ECG signals. FFT is used for the ECG signal using equation5

$$X[k] = \sum_{n=0}^{N-1} (x[n] \cdot \exp(-j \cdot 2\pi \cdot k \cdot n/N)) \tag{5}$$

where $X[k]$ represents the output signal at frequency index k, $x[n]$ is the input signal at time index n, N is the total number of samples, and exp denotes the exponential function.

To show the effect of the extracted feature over each class of ECG signal a box plot is shown in figure 20. It is observed from the plot that N beat and F beat can be classified, whereas SVEB, VEB, and Q beat are difficult to distinguish as this beat has some common characteristics associated with it.

The classification results, in terms of performance measures, obtained after performing FFT on the ECG signals before classification are presented in Table 4.

The classifier with the best performance RF, was applied to a testing set to create a confusion matrix, which was then used to evaluate the effectiveness of the classification model. The 98% overall accuracy of the results demonstrated the model's capacity to appropriately classify ECG signals into the appropriate groups. A confusion matrix is a useful tool for assessing how well a classification model is working. To analyze true positives, true negatives, false positives, and false negatives, it gives a thorough breakdown of the expected and actual class labels shown in Figure 21.

The sensitivity values attained by our approach are as follows: normal beats (N) −0.99%, SVEB −0.74%, VEB −0.93%, fusion beats (FB) −0.78%, and unknown beats (Q) −0.96%. Sensitivity measures the proportion of correctly identified positive instances within each class. Furthermore, the specificity values achieved are N −0.99%, SVEB −0.98%, VEB −0.96%, FB −0.99%, and Q −0.97%. Specificity represents the proportion of accurately identified negative instances within each class. Additionally, the positive predictive values obtained are N - 0.99%, SVEB - 0.74%, VEB - 0.96%, FB - 0.89%, and Q - 0.96%. Positive predictive value denotes the proportion of correctly predicted positive instances within each class.

These performance metrics demonstrate the effectiveness of our approach in accurately classifying ECG signals across different heartbeat classes. The high sensitivity values

TABLE 5. Performance comparison with some of the state of art-work result.

State of the art work	Feature Extraction	Classifier	Parameter	N	SVEB	VEB	FB
Proposed approach	RR feature, Morphological feature,HOS, FFT	Optimized RF	Accuracy	0.977	0.98	0.977	0.978
			Sensitivity	0.99	0.74	0.93	0.78
			Specificity	0.99	0.98	0.96	0.99
			PPV	0.99	0.74	0.96	0.89
			F1-score	0.99	0.75	0.93	0.75
			Precision	0.97	0.975	0.97	0.95
Liu, S.,et al., [9]	Spectral Feature, Bispectrum and 2D graph fourier transform (GFT)	SVM	Accuracy	0.960	0.982	0.977	0.988
			Sensitivity	0.993	0.715	0.767	0.462
			F1-Score	0.833	0.831	0.865	0.63
Marinho, L,et al., [29]	Fourier, Higher Order Statistics (HOS), Goertzel, Kurtosis	Naïve bias	Accuracy	0.94			
			Sensitivity	0.996	0.2	0.736	-
			Specificity	0.489	0.998	0.998	-
Bhattacharyya, S. et al., [28]	Statistical domain, Spectral domain, Temporal domain	Ensemble of RF and SVM by WMA	Accuracy	0.9821			
			Sensitivity	-	0.742	0.942	-
			PPV	-	0.9009	0.959	-
Zou, C.,et al., [30]	RR interval, Segment Label	RF with CNN	Accuracy	0.96			
			F1-score	0.98	074	0.93	-
			Precision	-	0.76	-	-
			Sensitivity	-	0.78	-	-
Chen, S.et al., [31]	Projected and dynamic feature, RR intervals	SVM	Accuracy	0.9846			
Mar, T.et al., [45]	Time domain, frequency domain, non-linear features	SFFS, linear, quadratic discriminant,MLP	Accuracy	0.89			
Afkhami,R et al., [46]	Statistical and mixture modelling feature,HOS, GMM	Decision Tree	Accuracy	0.9615			
			Sensitivity	0.9737	0.865	0.959	0.118
-			Positive Predictive	0.984	0.909	0.7763	0.2421
Lin,C et al,. [47]	RR interval, Morphological feature	linear discriminant	Sensitivity	0.916	0.814	0.862	-
			Positive Predictive	0.993	0.316	0.777	-
Garcia,G et al., [48]	TVCG and complex networks	PSO,SVM	Accuracy	0.924			
			Sensitivity	0.94	0.62	0.873	-
			Positive Predictive	0.98	0.53	0.594	-

indicate the model's ability to correctly identify positive instances, while the high specificity values reflect its proficiency in accurately classifying negative instances. The positive predictive values provide insights into the precision of the classification model within each class.

The overall accuracy of 98% showcases the robustness of the proposed approach in accurately predicting the heartbeat classes of ECG signals. These results suggest the potential of the approach in assisting with the early detection and diagnosis of cardiac abnormalities, particularly thoserelated to SVEB and VEB, which can have significant implications for patient health and well-being. Table 5 presents a comparative analysis demonstrating the superior performance of the proposed approach in efficient and automated arrhythmia diagnosis through heartbeat classification from ECG signals, surpassing state-of-the-art methods proposed in previous years. The study aimed to evaluate the effectiveness of various approaches for arrhythmia diagnosis and classification. Our proposed approach exhibited significantly higher accuracy and robustness compared to the existing state-of-the-art methods.

The RF algorithm's robustness is evident in its ability to classify ECG signals accurately, even in the presence of noise and data variability. Its decision tree-based ensemble approach provides insights into the decision-making process, enhancing interpretability. Furthermore, RF can handle high-dimensional ECG data with numerous features, making it suitable for complex signal analysis tasks. However, RF's computational complexity can be a drawback, especially for large datasets or when considering a wide range of hyperparameters. There is also a risk of bias if the individual decision trees are not diverse enough or if the training data is imbalanced. Additionally, while the overall decision-making process of RF is interpretable, the individual decision trees within the ensemble can be complex, making it challenging to understand the contribution of each feature or tree. RF's performance can also be sensitive to hyperparameter settings, necessitating careful tuning and optimization for optimal results.

V. CONCLUSION

In this work, a machine learning-based approach for classifying ECG data into various types of heartbeats, and distinguishing between healthy and unhealthy patterns is identified. RF classifiers give a more accurate performance before and after feature extraction as compared with other classifiers like SVM, LR, KNN, and DT. The proposed approach leverages methodology that explores and incorporates advanced feature extraction techniques, such as wavelet transform for signal denoising, morphological features, statistical features, and HOS analysis for effective feature extraction along with FFT on the performance of ECG signal classification. RF is optimized to enhance the accuracy of heartbeat classification. The result shows the proposed approach gives an overall accuracy of 98%, precision of 97.2%, sensitivity of 81.2%, and F1-score of 87.8% surprising state-of-the-art work to the best of our knowledge.

Furthermore, future work includes integrating Deep Learning frameworks into the existing classification model has the advantage of leveraging their powerful learning capabilities, enabling the model to learn intricate representations of the ECG signals. This can lead to improved accuracy by effectively capturing the intricate patterns and variations in the data [49], [50]. Low-cost ECG circuit development for real-time ECG signal processing [51] is also one of the areas that can be explored for live ECG signal analysis.

REFERENCES

[1] S. Deb, S. M. R. Islam, J. RobaiatMou, and M. T. Islam, "Design and implementation of low cost ECG monitoring system for the patient using smart device," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2017, pp. 774–778.

[2] M. J. Goldman, *Principles of Clinical Electrocardiography*, 1970, p. 400.

[3] S. Nita, S. Bitam, and A. Mellouk, "An enhanced random forest for cardiac diseases identification based on ECG signal," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 1339–1344.

[4] Z. Gao, J. Wu, J. Zhou, W. Jiang, and L. Feng, "Design of ECG signal acquisition and processing system," in *Proc. Int. Conf. Biomed. Eng. Biotechnol.*, May 2012, pp. 762–764.

[5] K. B. Jadhav and U. M. Chaskar, "Design and development of smart phone based ECG monitoring system," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 1568–1572.

[6] Y. Ye-Lin, J. M. Bueno-Barrachina, G. Prats-Boluda, R. R. de Sanabria, and J. Garcia-Casado, "Wireless sensor node for non-invasive high precision electrocardiographic signal acquisition based on a multi-ring electrode," *Measurement*, vol. 97, pp. 195–202, Feb. 2017.

[7] B. Babusiak, S. Borik, and M. Smondrk, "Two-electrode ECG for ambulatory monitoring with minimal hardware complexity," *Sensors*, vol. 20, no. 8, p. 2386, Apr. 2020.

[8] 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, May 2001.

[9] S. Liu, J. Shao, T. Kong, and R. Malekian, "ECG arrhythmia classification using high order spectrum and 2D graph Fourier transform," *Appl. Sci.*, vol. 10, no. 14, p. 4741, Jul. 2020.

[10] N. Singh and P. Singh, "Cardiac arrhythmia classification using machine learning techniques," in *Engineering Vibration, Communication and Information Processing: ICoEVCI 2018, India*. Singapore: Springer, 2019, pp. 469–480.

[11] S. M. Anwar, M. Gul, M. Majid, and M. Alnowami, "Arrhythmia classification of ECG signals using hybrid features," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–8, Nov. 2018.

[12] P. Pławiak, "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system," *Expert Syst. Appl.*, vol. 92, pp. 334–349, Feb. 2018.

[13] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci.*, vols. 415–416, pp. 190–198, Nov. 2017.

[14] M. S. Islam, K. F. Hasan, S. Sultana, S. Uddin, P. Lio', J. M. W. Quinn, and M. A. Moni, "HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN," *Neural Netw.*, vol. 162, pp. 271–287, May 2023.

[15] S. Zhang, C. Lian, B. Xu, Y. Su, and A. Alhudhaif, "12-lead ECG signal classification for detecting ECG arrhythmia via an information bottleneck-based multi-scale network," *Inf. Sci.*, vol. 662, Mar. 2024, Art. no. 120239.

[16] H. Narotamo, M. Dias, R. Santos, A. V. Carreiro, H. Gamboa, and M. Silveira, "Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches," *Biomed. Signal Process. Control*, vol. 93, Jul. 2024, Art. no. 106141.

[17] S.-F. Li, M.-L. Huang, and Y.-S. Wu, "Combining the Taguchi method and convolutional neural networks for arrhythmia classification by using ECG images with single heartbeats," *Mathematics*, vol. 11, no. 13, p. 2841, Jun. 2023.

[18] A. Eleyan and E. Alboghbaish, "Electrocardiogram signals classification using deep-learning-based incorporated convolutional neural network and long short-term memory framework," *Computers*, vol. 13, no. 2, p. 55, Feb. 2024.

[19] F. Khan, X. Yu, Z. Yuan, and A. U. Rehman, "ECG classification using 1-D convolutional deep residual neural network," *PLoS One*, vol. 18, no. 4, Apr. 2023, Art. no. e0284791.

[20] M. Ramkumar, C. G. Babu, G. S. Priyanka, and R. S. Kumar, "ECG arrhythmia signals classification using particle swarm optimization-support vector machines optimized with independent component analysis," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1084, no. 1, 2021, Art. no. 012009.

[21] X. Zhang, J. Li, Z. Cai, L. Zhao, and C. Liu, "Deep learning-based signal quality assessment for wearable ECGs," *IEEE Instrum. Meas. Mag.*, vol. 25, no. 5, pp. 41–52, Aug. 2022.

[22] A. Farooq, M. Seyedmahmoudian, and A. Stojcevski, "A wearable wireless sensor system using machine learning classification to detect arrhythmia," *IEEE Sensors J.*, vol. 21, no. 9, pp. 11109–11116, May 2021.

[23] A. B. Jani, R. Bagree, and A. K. Roy, "Design of a low-power, low-cost ECG & EMG sensor for wearable biometric and medical application," in *Proc. IEEE SENSORS*, Oct. 2017, pp. 1–3.

[24] S. Akbar, A. Ahmad, M. Hayat, A. U. Rehman, S. Khan, and F. Ali, "IAtbP-Hyb-EnC: Prediction of antitubercular peptides via heterogeneous feature representation and genetic algorithm based ensemble learning model," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104778.

[25] F. Ali, S. Akbar, A. Ghulam, Z. A. Maher, A. Unar, and D. B. Talpur, "AFP-CMBPred: Computational identification of antifreeze proteins by extending consensus sequences into multi-blocks evolutionary information," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 105006.

[26] S. Akbar, A. Raza, T. Al Shloul, A. Ahmad, A. Saeed, Y. Y. Ghadi, O. Mamyrbayev, and E. Tag-Eldin, ''PAtbP-EnC: Identifying anti-tubercular peptides using multi-feature representation and genetic algorithm-based deep ensemble model,'' *IEEE Access*, vol. 11, pp. 137099–137114, 2023.

[27] S. Akbar, F. Ali, M. Hayat, A. Ahmad, S. Khan, and S. Gul, ''Prediction of antiviral peptides using transform evolutionary & SHAP analysis based descriptors by incorporation with ensemble learning strategy,'' *Chemometric Intell. Lab. Syst.*, vol. 230, Nov. 2022, Art. no. 104682.

[28] S. Bhattacharyya, S. Majumder, P. Debnath, and M. Chanda, ''Arrhythmic heartbeat classification using ensemble of random forest and support vector machine algorithm,'' *IEEE Trans. Artif. Intell.*, vol. 2, no. 3, pp. 260–268, Jun. 2021.

[29] L. B. Marinho, N. D. M. M. Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. R. Filho, and V. H. C. de Albuquerque, ''A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification,'' *Future Gener. Comput. Syst.*, vol. 97, pp. 564–577, Aug. 2019.

[30] C. Zou, A. Müller, U. Wolfgang, D. Rückert, P. Müller, M. Becker, A. Steger, and E. Martens, ''Heartbeat classification by random forest with a novel context feature: A segment label,'' *IEEE J. Translational Eng. Health Med.*, vol. 10, pp. 1–8, 2022.

[31] S. Chen, W. Hua, Z. Li, J. Li, and X. Gao, ''Heartbeat classification using projected and dynamic features of ECG signal,'' *Biomed. Signal Process. Control*, vol. 31, pp. 165–173, Jan. 2017.

[32] S. Shridhar, Y. Karuna, S. Saladi, and R. Reddy, ''Denoising of ECG signals using wavelet transform and principal component analysis,'' in *Proc. Int. Conf. Sustain. Comput. Sci., Technol. Manag. (SUSCOM)*. Jaipur, India: Amity Univ. Rajasthan, Feb. 2019.

[33] W. Hao, Y. Chen, and Y. Xin, ''ECG baseline wander correction by mean-median filter and discrete wavelet transform,'' in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 2712–2715.

[34] G. H. Choi, K. Lim, and S. B. Pan, ''Identification system based on resolution adjusted 2D spectrogram of driver's ECG for intelligent vehicle,'' *Mobile Inf. Syst.*, vol. 2022, pp. 1–13, Jan. 2022.

[35] H. Hindarto, I. Anshory, and A. Efiyanti, ''Classification of heart signal using wavelet Haar and backpropagation neural network,'' *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 403, no. 1, 2018, Art. no. 012069.

[36] M. K. Das and S. Ari, ''ECG beats classification using mixture of features,'' *Int. Scholarly Res. Notices*, vol. 2014, pp. 1–12, Sep. 2014.

[37] W.-J. Rappel, ''The physics of heart rhythm disorders,'' *Phys. Rep.*, vol. 978, pp. 1–45, Sep. 2022.

[38] S. U. Hassan, M. S. Mohd Zahid, T. A. Abdullah, and K. Husain, ''Classification of cardiac arrhythmia using a convolutional neural network and bi-directional long short-term memory,'' *Digit. Health*, vol. 8, Jan. 2022, Art. no. 205520762211027.

[39] D. Zhang, Y. Chen, Y. Chen, S. Ye, W. Cai, and M. Chen, ''An ECG heartbeat classification method based on deep convolutional neural network,'' *J. Healthcare Eng.*, vol. 2021, pp. 1–9, Sep. 2021.

[40] L. V. Kumari and Y. P. Sai, ''Classification of ECG beats using optimized decision tree and adaptive boosted optimized decision tree,'' *Signal, Image Video Process.*, vol. 16, no. 3, pp. 695–703, 2022.

[41] V. Mahesh, A. Kandaswamy, C. Vimal, and B. Sathish, ''ECG arrhythmia classification based on logistic model tree,'' *J. Biomed. Sci. Eng.*, vol. 2, no. 6, p. 405, 2009.

[42] Z. Ge, Z. Zhu, P. Feng, S. Zhang, J. Wang, and B. Zhou, ''ECG-signal classification using SVM with multi-feature,'' in *Proc. 8th Int. Symp. Next Generation Electron. (ISNE)*, Oct. 2019, pp. 1–3.

[43] S. Faziludeen and P. Sankaran, ''ECG beat classification using evidential K-nearest neighbours,'' *Proc. Comput. Sci.*, vol. 89, pp. 499–505, Jan. 2016.

[44] P. Yang, D. Wang, W.-B. Zhao, L.-H. Fu, J.-L. Du, and H. Su, ''Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification,'' *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102138.

[45] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, ''Optimization of ECG classification by means of feature selection,'' *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2168–2177, Feb. 2011.

[46] R. G. Afkhami, G. Azarnia, and M. A. Tinati, ''Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals,'' *Pattern Recognit. Lett.*, vol. 70, pp. 45–51, Jan. 2016.

[47] C.-C. Lin and C.-M. Yang, ''Heartbeat classification using normalized RR intervals and morphological features,'' *Math. Problems Eng.*, vol. 2014, pp. 1–11, Jul. 2014.

[48] G. Garcia, G. Moreira, D. Menotti, and E. Luz, ''Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO,'' *Sci. Rep.*, vol. 7, no. 1, p. 10543, Sep. 2017.

[49] W. Ullah, I. Siddique, R. M. Zulqarnain, M. M. Alam, I. Ahmad, and U. A. Raza, ''Classification of arrhythmia in heartbeat detection using deep learning,'' *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–13, Oct. 2021.

[50] Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. B. M. Pauzi, Q. Zhang, and P. Y. Lim, ''Deep learning-based ECG arrhythmia classification: A systematic review,'' *Appl. Sci.*, vol. 13, no. 8, p. 4964, Apr. 2023.

[51] N. Faruk, A. Abdulkarim, I. Emmanuel, Y. Y. Folawiyo, K. S. Adewole, H. A. Mojeed, A. A. Oloyede, L. A. Olawoyin, I. A. Sikiru, M. Nehemiah, A. Ya'u Gital, H. Chiroma, J. A. Ogunmodede, M. Almutairi, and I. A. Katibi, ''A comprehensive survey on low-cost ECG acquisition systems: Advances on design specifications, challenges and future direction,'' *Biocybern. Biomed. Eng.*, vol. 41, no. 2, pp. 474–502, Apr. 2021.

TANUJA SUBBA received the B.Tech. and M.Tech. degrees in computer science and engineering from Sikkim Manipal Institute of Technology, Sikkim Manipal University (SMU), in 2015 and 2017, respectively.

Since 2021, she has been an Assistant Professor with the Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, SMU. She is the author of five articles. Her research interests include human–computer interaction, biosignals, machine learning, robotics, and artificial intelligence.

TEJBANTA CHINGTHAM received the Bachelor of Engineering (B.E.) degree in computer science and engineering from Bharathiar University, Coimbatore, Tamil Nadu, in 1999, the M.Tech. degree from IIT Guwahati, in 2003, and the Ph.D. degree from BIT Mesra, in 2012, in the area of specialization of robotics and computational intelligence.

In 2000, he joined as an Assistant Professor with the Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majitar. Over the years, he has progressed in his academic journey and currently holds the position of a Professor with Sikkim Manipal Institute of Technology, where he continues to contribute to the field of computer science and engineering. He has more than 45 articles and a patent. His research interests include human–computer interaction, computational intelligence, immune systems, robot path planning, and robotics.