BIOCYBERNETICS AND BIOMEDICALENGINEERING 41 (2021) 656–666

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/bbe

Original Research Article

An improved cardiac arrhythmia classification using an RR interval-based approach

Jagdeep Rahul [a,*], *Marpe Sora* [b], *Lakhan Dev Sharma* [c], *Vijay Kumar Bohat* [d]

[a] *Department of Electronics & Communication Engineering, Rajiv Gandhi University, India*
[b] *Department of Computer Science & Engineering, Rajiv Gandhi University, India*
[c] *School of Electronics Engineering VIT-AP University, India*
[d] *Department of Computer Science and Engineering, Bennett University, India*

A R T I C L E I N F O

Article history:
Received 17 November 2020
Received in revised form
15 April 2021
Accepted 16 April 2021
Available online 13 May 2021

Keywords:
ECG
Cardiac arrhythmia
Classifier
RR interval
PVC
PAC

A B S T R A C T

Accurate and early detection of cardiac arrhythmia present in an electrocardiogram (ECG) can prevent many premature deaths. Cardiac arrhythmia arises due to the improper conduction of electrical impulses throughout the heart. In this paper, we propose an improved RR interval-based cardiac arrhythmia classification approach. The Discrete Wavelet Transform (DWT) and median filters were used to remove high-frequency noise and baseline wander from the raw ECG. Next, the processed ECG was segmented after the determination of the QRS region. We extracted the primary feature RR interval and other statistical features from the beats to classify the Normal, Premature Ventricular Contraction (PVC), and Premature Atrial Contraction (PAC). The K-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), and Random Forest (RF) classifier were utilised for classification. Overall performance of SVM with Gaussian kernel achieved Se % = 99.28, Sp % = 99.63, +P % = 99.28, and Acc % = 99.51, which is better than the other classifiers used in this method. The obtained results of the proposed method are significantly better and more accurate.

1. Introduction

Cardiovascular diseases (CVDs) are the primary cause of death in the world. According to a World Health Organisation report, around 17.9 million people died of CVDs in 2016, accounting for 13% of total deaths. Almost 85% of total CVD deaths are due to stroke and heart attack [1]. Over three-fourths of total CVD deaths occur in low- and middle-income countries. People with CVDs present cardiovascular risks such as hypertension and diabetes that require early detection to prevent premature death. Heart failure is the most common CVD in developed and developing countries, especially among the elderly population [2]. Electrocardiogram (ECG) depicts electrical activity of the heart and is the

* Corresponding author.
E-mail address: jagdeeprahul24@gmail.com (J. Rahul).
https://doi.org/10.1016/j.bbe.2021.04.004

most widely used tool for detecting cardiovascular diseases in early stages [3]. ECG portrays of the sequence of events associated with electrical impulse conduction in the various chambers of the heart [4]. A standard ECG carries the P, QRS, T, and (occasionally) the U-wave in a series to complete one cardiac cycle [5]. The U-wave in the ECG is generally not found, due to its dependency on the T-wave amplitude and heart rate [6,7]. The P-wave is associated with atrial depolarisation; the QRS region represents the depolarisation of the ventricles; and the T-wave shows the repolarisation of the ventricles [8]. Any change in the characteristics of these waves may indicate the presence of arrhythmia in the heart. Accurate interpretation can help in the diagnosis and prognosis of cardiac arrhythmia [9].

The presence of cardiac arrhythmias in the heart can be easily detected through a machine-learning approach. Premature ventricular and atrial contraction (PVC and PAC) are arrhythmias that are usually not dangerous; they may occur in many healthy people. The presence of PVC may indicate low blood oxygen levels, which may result from chronic obstructive pulmonary disease (COPD) or pneumonia. Similarly, PAC may predict atrial fibrillation, which is a highly prevalent arrhythmia related to congestive heart failure [10]. With PAC, an electrical impulse is excited through an ectopic location rather than from the sinoatrial (SA) node in the atrium. Consequently, the atrium contracts prematurely, resulting in a non-sinus P-wave. Due to early contraction of the atrium, the P-wave generally merges with the T-wave. Moreover, it changes the activation timing of the atrioventricular (AV) node and creates an unusual delay, prolonging the PP interval in the ECG. Occasionally, PAC conduction becomes abnormal and widens the QRS complex, similar to bundle branch block, in which an electrical impulse is transmitted from the SA to the AV node but does not continue to the lower chambers through the bundle of His to activate them. Hence, an ectopic site activates the ventricles to depolarise the lower chambers in PVC. The PVC-affected cardiac cycle does not contain the P-wave. The PVC-affected QRS complex may be wider, taller, and sometimes negative in the ECG [11]. During the COVID-19 pandemic, patients with CVDs have been highly vulnerable to SARS-CoV-2, which can result in acute lung injury and arrhythmic complications [12].

Automatic beats classification is needed in ECG signal processing due to the presence of different types of arrhythmias. The manual classification of arrhythmias using the mathematical approach is tedious and time-consuming. The preprocessing of an ECG plays a crucial role in making the ECG noise-free before applying the machine-learning approach for arrhythmia classification. Noise in the ECG can alter the overall features of the signal, which may affect the accuracy of the classifier [13]. Researchers around the world have proposed many machine-learning approaches for automatic classification of ECG arrhythmias [14]. The machine-learning model based on the LSTM and CNN-SVM is used for arrythmias, congestive heart failure, and normal sinus rhythm classification in [15]. The approximate coefficients of the sixth-level decomposed ECG signal, using the tunable Q-wavelet transform, are selected for feature extraction of each beat in [16], which uses the SVM for the classification of the seven different classes. The real-time ventricular contraction detection method has been proposed, using the neural network weight fuzzy function to distinguish the PVC beats from the normal beat [17]. Automatic classification of normal, PVC, and left bundle branch block is proposed and implemented using the wavelet transform and SVM in [18]. Sixteen subclasses (within five classes) are utilised for classification using the general sparsed neural network in [19]. In this method, nine time frequencies and two high-level features were extracted from the QRS region. The method in [20] propose seventeen features, including normal, PAC, and PVC classes, classified based on the frequency components of the power spectral density feature set using SVM. The multi-layer perceptron and convolutional neural network are used in [21] for cardiac arrhythmia classification. Other cardiac diseases can also be detected through the machine learning approach; 3D display is presented in [22]. Other automatic classification methods for PVC beats are presented in [23] using k-means and Fuzzy C-Means, [24] using chaotic features, [25] using recurrence quantification analysis, and in [26] using template matching. [27] propose ECG arrhythmia classification for PAC using spectral correlation and SVM. An algorithm for ECG arrhythmia classification using principal component analysis (PCA) and an extreme learning machine is proposed in [28]. [29] propose an intelligent learning approach model for five types of arrhythmia, using the hidden Markov model (HMM) classifier.

The methods for ECG arrhythmia classification, including the PAC, are proposed in [30] using a parallel neural network,[31] using the fisher linear discriminant, and [32] using the relevance vector machine. Other classification methods applying deep learning and a bidirectional LSTM-based model have been used for automatic classification of cardiac arrhythmias [33]. Ventricular fibrillation and ventricular tachycardia arrhythmia episode are classified using the least square support vector machine (LS-SVM) classifier in [34]. The QRS complex detection can also be performed using the deep learning model [35].

Methods based on the RR interval are proposed in [36–38]. The method proposed in [36] uses RR intervals and morphological features for supraventricular and ventricular ectopic beats detection, based on the linear discriminants classifier. In the method described by [37], different patterns of the PVC beat are detected in the presence of the normal and abnormal heartbeats; this method uses the temporal features of the QRS, T-wave, and RR intervals to detect the PVC beats in the ECG. The classification of the normal beats, PVC beats, and PAC is presented in the method developed by [38], which considers the RR-interval, template feature, width feature, and height feature of the QRS complex. However, while all of these methods [36–38] utilise the RR interval for classification of premature beats, none of them use statistical features along with the RR interval. To avoid data imbalance in the classifier, an equal number of beats should be selected from each class for better classification and justification. The method in [38] accounts for the different number of beats from each class for the classification, which may create a data imbalance for the classifier. Two-stage classification between the normal and abnormal class and thereafter, a three-stage classification among normal, arrhythmia, and atrial fibrillation was performed using the frequency, time, and geometric domain of HRV features [39]. In [40], the ECG signal was decomposed up to the 8th level and their coefficients were used to remove the noise. Furthermore, the amplitudes of ECG sub-waves and the PR and RR intervals are utilised for the classification of cardiac arrhythmias. Applying RR interval techniques, the EMD, VMD, and time-domain features have been used as an input to the cubic-SVM for classification of arrhythmias [41]. The sinusoidal regression features and RR intervals are used for classification based on the naïve bayes classifier in [42]. Supraventricular ectopic and ventricular ectopic beats are utilised for inter-patient classification, along with random projection and the RR interval, in [43]. In the method developed by [44], a deep learning approach based on the convolutional and recurrent-neural networks (CNN and RNN) is used to extract high-level features from the RR intervals to distinguish atrial fibrillation from the normal sinus rhythm. Our proposed method uses the RR interval with statistical features for classification of normal beat, PVC beat, and PAC beat.

In this paper, we select five different classifiers from the supervised machine learning approach for the classification of cardiac arrhythmias. The raw ECG signal is pre-processed to eliminate high-frequency noise and baseline wander, using the discrete wavelet transform (DWT) and median filter. Next, the processed ECG signal is segmented after the determination of the QRS region. The QRS region is extracted based on a predetermined window length for features extraction. The primary feature (RR interval) and other statistical features are estimated from the extracted beats for classification of the arrhythmias.

2. Materials

The Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmia Database (MIT-BIH AD) consists of 48 ECG records, with each record being at least thirty minutes in duration. This database consists of mixed ECG recordings from inpatients and outpatients who were affected by cardiac arrhythmias. The MIT-BIH AD database is sampled at 360 samples per seconds over the range of 10 mV, with 11-bit resolution [45,46]. The ECG records, which are used in our proposed method, are shown in Table 1.

In Table 1, all PAC beats were taken from the ECG records of MIT-BIH AD. A total of 2676 PAC beats are represented in the database, and all are considered for this classification. The same number of PVC beats are extracted from the common records with PAC beats. Finally, the same numbers of normal beats are also selected from the records, which are common in both PAC and PVC. A total of 8028 beats were extracted from the MIT-BIH AD database for training, testing, and validation of the proposed method.

3. Proposed methodology

We present a classification of PAC, PVC, and normal beats that applies machine-learning algorithms. This section describes pre-processing of the raw ECG signal with the help of the discrete wavelet transform (DWT) and two-stage median filter. Next, normal, PAC, and PVC beats are segmented, followed by the detection of the QRS complex from the processed ECG signal. Thereafter, nine features (RR interval and eight other statistical features) are extracted from the segmented beats. The extracted features are applied to the five different classifiers to evaluate the performance of the method. These classifiers are tuned to obtain the best results. The process flow of the proposed method is shown in Fig. 1.

3.1. *Pre-processing*

The raw ECG signal must be pre-processed at the initial stage due to the presence of baseline wander and high-frequency noises. These noises are highly dominating and can alter the features of the ECG, leading to an incorrect interpretation of a cardiac arrhythmia. High-frequency noises can be dealt with via various techniques, such as low pass filtering, least mean square filtering, moving average filtering, DWT, and many more. All filtering approaches except DWT change the ECG morphology after pre-processing [47]. Wavelet transform works on the basis function and mother wavelet, which is shown in Eq. 1.

$$Wx(a,\ b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\Psi^{*}\left(\frac{t-b}{a}\right)dt \tag{1}$$

The DWT allows decomposition of the signal into different scales due to its multi-scale features. The DWT is expressed in Eq. 2.

$$W(j,k) = \sum_{j}\sum_{k} x(k)e^{\frac{-j}{2}}\Psi(2^{-j}\ n-k) \tag{2}$$

Table 1 – ECG records from MIH-BIH AD considered in the proposed method for classification.

Types	Classes	Records
Arrhythmia	Premature Atrial Contraction (PAC) Beat	100 m, 101 m, 103 m, 108 m, 112 m, 113 m, 114 m,116 m, 117 m, 118 m, 121 m, 124 m, 200 m, 201 m, 202 m, 203 m, 205 m, 207 m, 209 m, 210 m, 213 m, 215 m, 219 m, 220 m, 222 m, 223 m, 228 m, 231, 232 m, and 233 m
	Premature Ventricular Contraction (PVC) Beat	100 m, 108 m, 114 m, 116 m, 118 m, 121 m, 124 m, 200 m, 201 m, 203 m, 202 m, 205 m, 207 m, 209 m, 210 m, 213 m, 215 m, 219 m, 223 m, 228 m, 231, and 233 m
Normal	Normal Beats	100 m, 108 m, 114 m, 116 m, 121 m, 124 m, 200 m, 202 m, 205 m, 207 m, 209 m, 210 m, 213 m, 219 m, 223 m, 228 m, 231, and 233 m

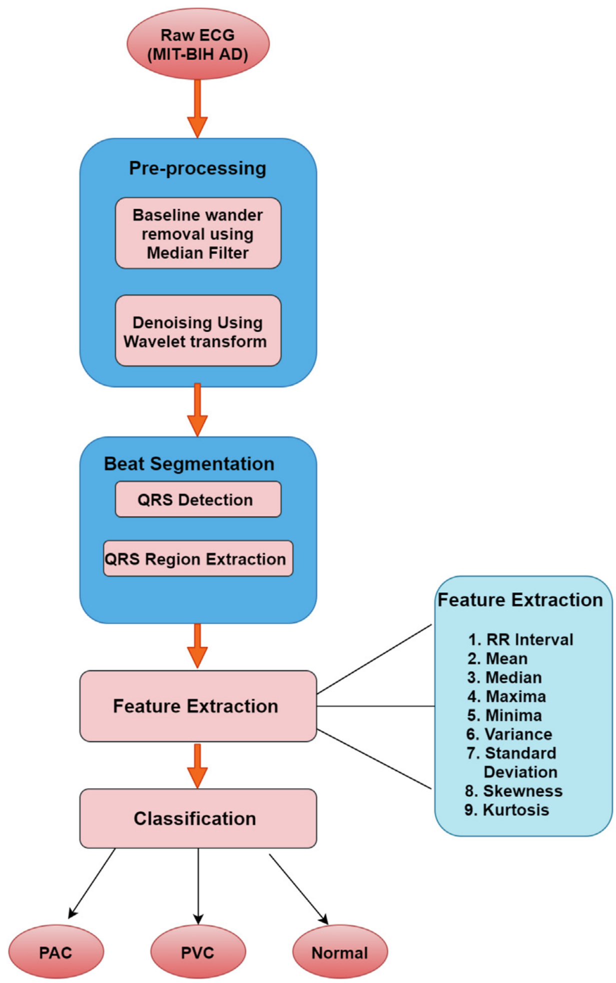

Fig. 1 – Process flow diagram of the proposed method.

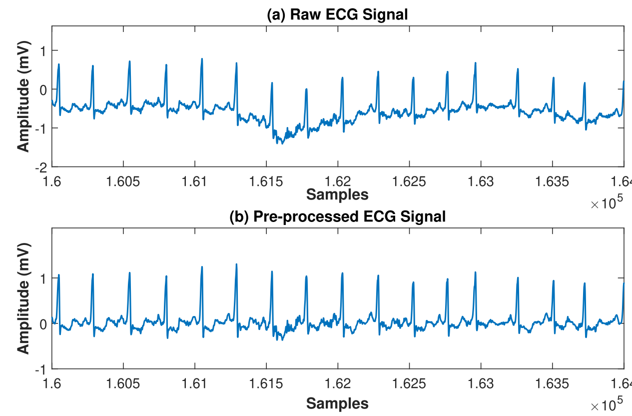

Fig. 2 – (a) Raw ECG Signal (b) Pre-processed ECG signal from MIT-BIH AD, record 223 m.

where, $x(t)$ is the ECG, α is the scaling, b is the translation parameter, and $x(t)\Psi^{*}$is a mother wavelet. ECG was denoised using the DWT, but baseline wander noise is still present in the signal. Two median filters connected in the cascade structure were utilised to remove the baseline wander noise without degrading the quality of the ECG [48]. The output of the median filter is calculated using Eq. 3.

$$y(n) = x(n) - med(x_{i-j}, \ldots, x, \ldots, x_{i+j}) \tag{3}$$

Where $y(n)$ is the pre-processed ECG and $x(n)$ is the ECG after performing the wavelet transform for denoising. The output of the second-stage median filter was further subtracted from the raw ECG to obtain the baseline wander free signal. The raw ECG and pre-processed ECG signals are shown in Fig. 2.

3.2. *Beat segmentation*

The PAC and PVC beats are associated with the delay of electric impulse conduction in the atrial and ventricle chambers of the heart. The QRS complex is a primary component of the ECG that represents ventricular depolarisation. The amplitude, duration, and shape of the QRS region can be used to determine the presence of a cardiac arrhythmia. Hence, QRS detection is a widely studied topic.

Many methods have been proposed in [49–52] for QRS complex detection in ECGs. We perform QRS complex detection using the method proposed in [52]. In PAC and PVC arrhythmia, QRS characteristics may differ from the normal QRS. The normal QRS is pointed and narrow, although QRS complexes may be wider in PAC and PVC due to the occurrence of beat in the refractory period [53]. Some aberrant PAC and PVC beats show an inverted QRS in normal ECG records. The normal QRS duration lies within the range of 80–100 ms, whereas premature beat duration is greater than 0.1s. A window of 0.125s is considered in this study for beat segmentation and QRS region extraction. Extraction of the normal QRS region from one cardiac cycle is shown in Fig. 3. The PVC and PAC beats are shown in Fig. 4 and Fig. 5, respectively.

3.3. *Feature extraction*

The dataset is prepared from the extracted QRS complex region of normal QRS complexes, PAC beats, and PVC beats. Feature extraction is an essential and widely used technique for data processing and for selecting the relevant feature from the input dataset. The greater number of features may indicate more information and better discriminative power, which would also increase the search space and computa-

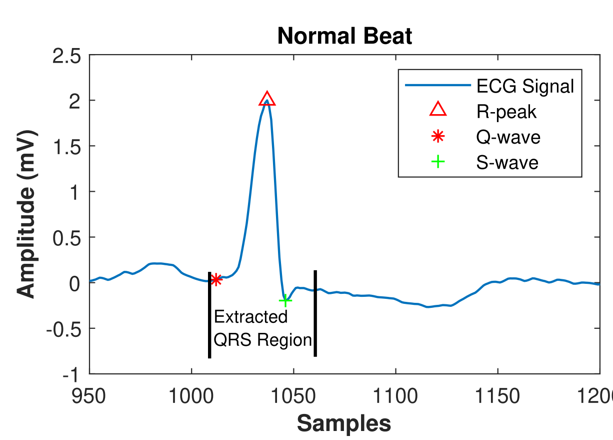

Fig. 3 – Normal beat from MIT-BIH AD 223 m, ECG record.

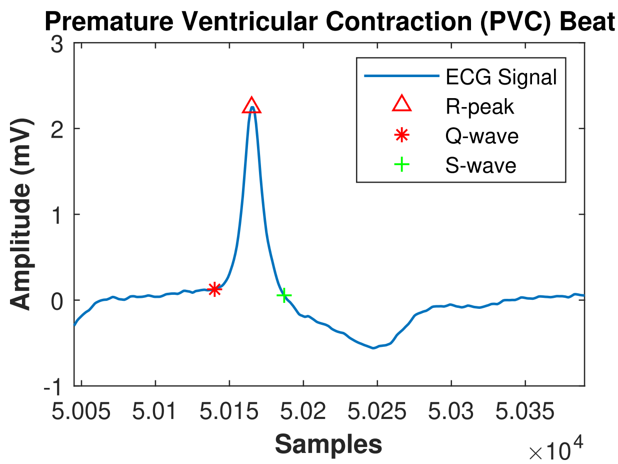

Fig. 4 – Premature Ventricular Contraction (PVC) beat from MIT-BIH AD 223 m, ECG record.

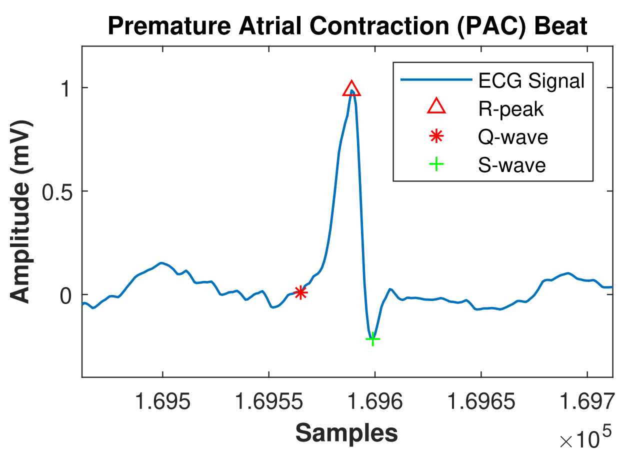

Fig. 5 – Premature Atrial Contraction (PAC) beat from MIT-BIH AD 223 m, ECG record.

tional resources. The presence of irrelevant and redundant features in the large dataset can confuse the classifier, which degrades the performance of the classifier, leading to overfitting. Feature extraction is the mapping of the more extensive feature set into a smaller dimensional feature set derived from the larger dataset for better and faster classification. The types of the features extracted from the input dataset are the RR interval, mean, median, maximum, minimum, variance, standard deviation, skewness, and kurtosis, based on the Eqs. (4)–(12). A pictorial presentation of the RR interval extraction is given in Fig. 6. The statistical features are calculated from the dataset extracted from the selected QRS regions ($y[n]$) as follows:

$$RR_{interval} = R_{Peak}(current) - R_{Peak}(previous) \tag{4}$$

$$\mu_y = E[X] = \frac{1}{N}\sum_{1}^{n} y[n] \tag{5}$$

$$m_y = \begin{cases} \left(\frac{n}{2}+1\right), & when\ \ n\ \ is\ \ even \\ \left(\frac{n+1}{2}\right), & when\ \ n\ \ is\ \ odd \end{cases} \tag{6}$$

$$\theta_y = \max(y[n]) \tag{7}$$

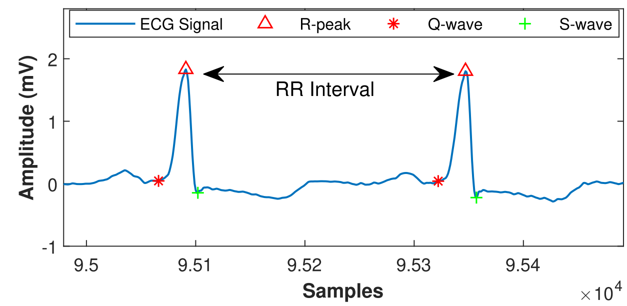

Fig. 6 – RR interval extraction between previous and current R-peak in the ECG signal.

$$\delta_y = \min(y[n]) \tag{8}$$

$$\sigma_y^2 = E\left[\left(X - \mu_y\right)^2\right] \tag{9}$$

$$\sigma_y = \sqrt{E\left[\left(X - \mu_y\right)^2\right]} \tag{10}$$

$$s = E\left[\left(\frac{X - \mu_y}{\sigma_y}\right)^3\right] \tag{11}$$

$$k = \frac{E\left[\left(X - \mu_y\right)^4\right]}{\left(E\left[\left(X - \mu_y\right)^2\right]\right)^2} \tag{12}$$

where, μ_y , $m_y, \theta_y, \delta_y, \sigma_y$, σ_y^2, s, and k represent the mean, median, maximum, minimum, standard deviation, variance, skewness, and kurtosis of the extracted QRS complex, respectively.

3.4. *Classification*

Classification is the process of placing the data into the given set of categories by mapping the attribute set to one of the predefined class labels. The classification algorithm is a type of supervised machine learning; it is categorised as either a linear or nonlinear model. In this study, we use the nonlinear classification algorithms k-Nearest Neighbours (kNN) [54], Support Vector Machine (SVM) [55], Naïve Bayes (NB) [56], Decision Tree (DT) [57], and Random Forest (RF) [58] to distinguish the PAC and PVC beats from the normal beats. The box plot of the primary feature (the RR interval) of normal, PVC, and PAC beats is shown in Fig. 7. The feature distribution of skewness and the RR interval of all extracted beats is shown in Fig. 8. Moreover, graphical presentations of all statistical features are shown in Figs. 9 and 10, respectively.

3.4.1. *Classifiers*

The k-NN is a non-parametric nonlinear classification algorithm that works with four different metric functions. This algorithm model learns during prediction time, not in training time: it is a lazy and instance-based learning classifier [59]. The classification performed in this algorithm is based on the Euclidean distance between two points [60]. SVM solves the classification problem efficiently by choosing the correct

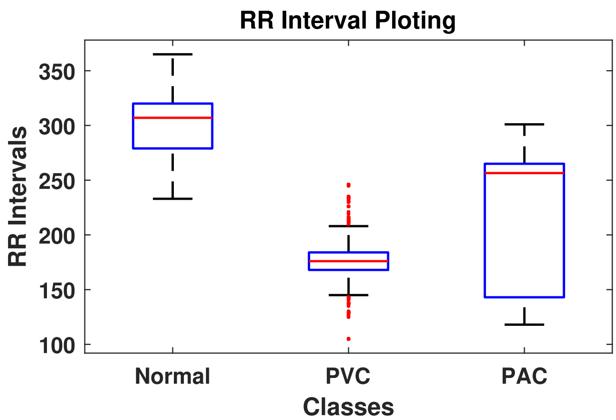

Fig. 7 – RR interval plotting of normal, PVC, and PAC beats.

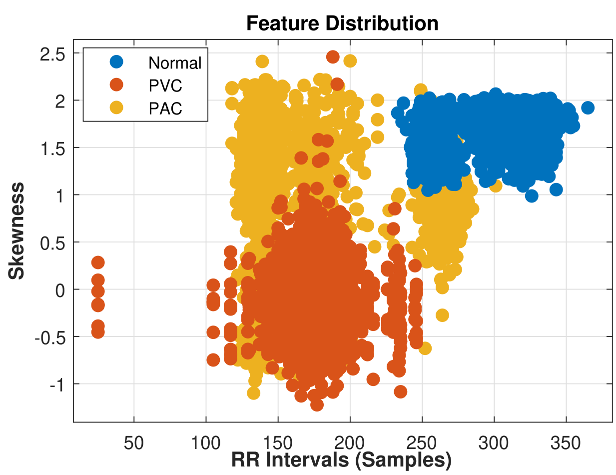

Fig. 8 – Feature distribution plot between the skewness and RR intervals of normal, PVC, and PAC beats.

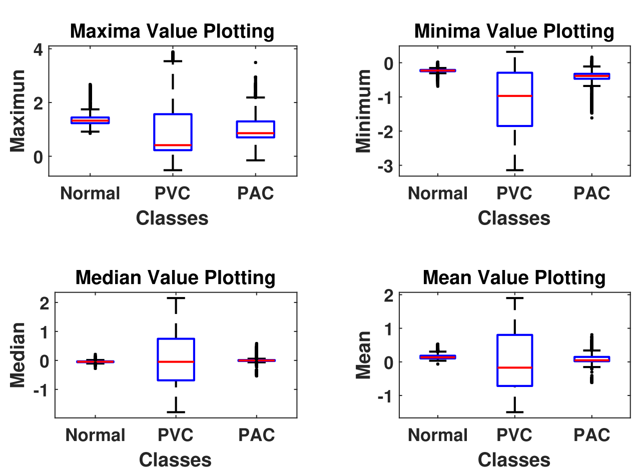

Fig. 9 – Statistical features: maxima, minima, median, and mean representation of normal, PVC, and PAC beats.

kernel trick. The SVM classifier uses the different kernel trick namely, the linear, polynomial, and Radial Basis Function (RBF) kernels to obtain better results [61,62].

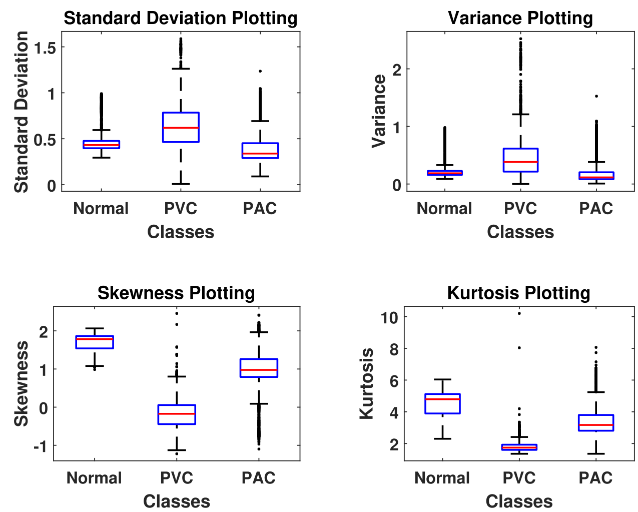

Fig. 10 – Statistical features: standard deviation, variance, skewness, and kurtosis representation of normal, PVC, and PAC beats.

Naïve Bayes (NB) works on the event probability based on the prior knowledge of the condition [63]. In a real-life situation, features are dependent upon each other, but these features can be assumed independently for computing the results. A Decision Tree (DT) consists of the node, branch, and leaf that contain the predicted outcome [64]. To avoid overfitting in this classifier, the depth of the tree can be fixed at a certain level. This classifier is simple, fast, and easy to interpret for a small size. A decision tree uses entropy to partition the data and calculate the entropy of the attribute. The Random Forest (RF) algorithm is formally defined as a collection of decision tree classifiers: $f(x, \theta_k)$, $k = 1, 2, \ldots, k$; where θ_k is a random vector that meets the independent and identically distributed assumption [65]. Here, x is the input for which every decision tree in the random forest casts a unit vote for selecting its class label. The major advantage of RF is its ability to avoid the overfitting problem [66].

4. Results

The performance of the classifier was evaluated on the selected arrhythmia ECG records from the MIT-BIH AD database. The classifier's performance was assessed using the performance metrics Specificity rate (Sp %), Sensitivity rate (Se %), Positive predictivity rate (+P %), and Accuracy rate (Acc %). The calculation methods for these performance metrics are given in [60].

A total of 2676 PAC-affected beats were segmented and extracted from the MIT-BIH AD records. To avoid the imbalance data situation in this work, we have considered the same number of PVC, PAC, and normal beats extracted from the same database records. We selected 8028 instances for classification of cardiac arrhythmia, using the nine different extracted features. The Euclidean distance measure was used for the k-NN classifier. The Gaussian kernel was used in the SVM to assess performance. Similarly, DT, NB, and RF were utilised with the 100 number of split, Gaussian function, and 50 number of bags, respectively. These functions of the classifier achieve better results compared with other functions in the same classifier. Thus, a single function is considered in this study to assess the performance. The obtained results with confusion matrices from each classifier are reported in Table 2.

The obtained results from the five different classifiers, along with their confusion matrices, are presented in Table 2. The Sp %, Se %, +P %, and Acc % are calculated for each class in a particular classifier with respect to the other classes. Finally, the overall performance is calculated for each classifier. The k-NN classifier with the Euclidean metric has achieved overall performance of Se % = 99.15, Sp % = 99.57, +P % = 99.48, and Acc % = 99.43. The Gaussian kernel trick is used in the SVM to achieve an overall performance of Se % = 99.28, Sp % = 99.63, +P % = 99.28, and Acc % = 99.51. The Gini's diversity index with 100 splits in the DT classifier shows overall performance of Se % = 98.68, Sp % = 99.34, +P % = 98.68, and Acc % = 99.12. The Gaussian function with the NB classifier has achieved overall performance of Se % = 93.51, Sp % = 96.75, +P % = 93.86, and Acc % = 95.67. The results obtained with 50 number of bags for the RF classifier are Se % = 99.27, Sp % = 99.62, +P % = 99.29, and Acc % = 99.51.

5. Discussion

We propose cardiac arrhythmia classification using the RR interval-based approach and supervised machine-learning classifiers. Two different types of arrhythmias premature ventricular contraction (PVC) and premature atrial contraction (PAC) beats were selected from the ECG records available in the MIT-BIH AD. Three different classes were created using the normal, PVC, and PAC beats to evaluate the performance of the proposed method. Feature extraction was performed on each beat for further classifications. The RR interval as the primary feature, along with eight different statistical features, were extracted from the segmented QRS region. We have used five classifiers (k-NN, SVM, DT, NB, and RF) for classifications based on the features extracted from the arrhythmias and normal beats. Only the best result-oriented functions/kernels from each classifier are utilised to obtain the performance results. The summarised performance comparison of the best result from the five different classifiers is shown in Table 3.

The performance parameters were computed for each classifier in the proposed method. The best result, in terms of accuracy (Acc %) = 99.51, is achieved by the two classifiers SVM with Gaussian trick and Random Forest with 50 number of bags. Of the remaining three classifiers, Naïve Bayes performs less well than the k-NN and DT. The area under the curve of the SVM classifier with Gaussian kernel tricks is shown in Fig. 11.

The performance obtained from the SVM and RF in terms of the number of classes, the number of features, and accuracy is compared to the other existing state-of-the-art methods reported in Table 4. The results obtained with the proposed method are compared to the existing state-of-the-art methods in terms of accuracy. A comparison is also performed on the basis of classes and the classifier used. Three different classes (normal, PVC, and PAC) are used in the methods of [67] with SVM, k-NN, and Random Forest, [68] with neural network (NN), [69,70] with two-layered Hidden Markov Models (HMM), [71] with the artificial neural network (ANN), and [72] with the learning vector quantisation (LVQ) neural network. Three-class classification methods are proposed in [15,39]. They use SVM and ANN, respectively, for normal beats and arrythmia classification. The performance of the proposed method is found to be better than the performance

Table 2 – Performance assessment of the k-NN, SVM, DT, NB, and RF.

Classifier	Category	Normal	PVC	PAC	Se %	Sp %	+ P %	Acc %
k-NN	Normal	2675	0	1	99.96	99.94	99.89	99.95
	PVC	0	2640	36	98.65	99.48	98.95	99.20
	PAC	3	28	2645	98.84	99.31	98.62	99.15
	Overall				**99.15**	**99.57**	**99.48**	**99.43**
SVM	Normal	2674	1	1	99.93	99.98	99.96	99.96
	PVC	0	2660	16	99.40	99.25	98.52	99.30
	PAC	1	39	2636	98.51	99.68	99.36	99.29
	Overall				**99.28**	**99.63**	**99.28**	**99.51**
DT	Normal	2668	0	8	99.70	99.78	99.55	99.75
	PVC	0	2639	37	98.62	99.08	98.18	98.93
	PAC	12	49	2615	97.72	99.16	98.31	98.68
	Overall				**98.68**	**99.34**	**98.68**	**99.12**
NB	Normal	2441	0	235	91.22	99.98	99.96	97.06
	PVC	0	2544	132	95.07	97.14	94.33	96.45
	PAC	1	153	2522	94.25	93.14	87.30	93.51
	Overall				**93.51**	**96.75**	**93.86**	**95.67**
RF	Normal	2671	0	5	99.81	99.98	99.96	99.93
	PVC	0	2646	30	98.88	99.61	99.21	99.36
	PAC	1	21	2654	99.14	99.29	98.59	99.24
	Overall				**99.27**	**99.62**	**99.29**	**99.51**

Table 3 – Summarised performance comparison of five different classifiers.

Classifier	Functions/kernels	Se %	Sp %	+P %	Acc %
k-NN	Euclidean	99.15	99.57	99.48	99.43
SVM	Gaussian	99.28	99.63	99.28	99.51
Decision Tree	Gini's diversity index	98.68	99.34	98.68	99.12
Naïve Bayes	Gaussian	93.51	96.75	93.86	95.67
Random Forest	Bags	99.27	99.62	99.29	99.51

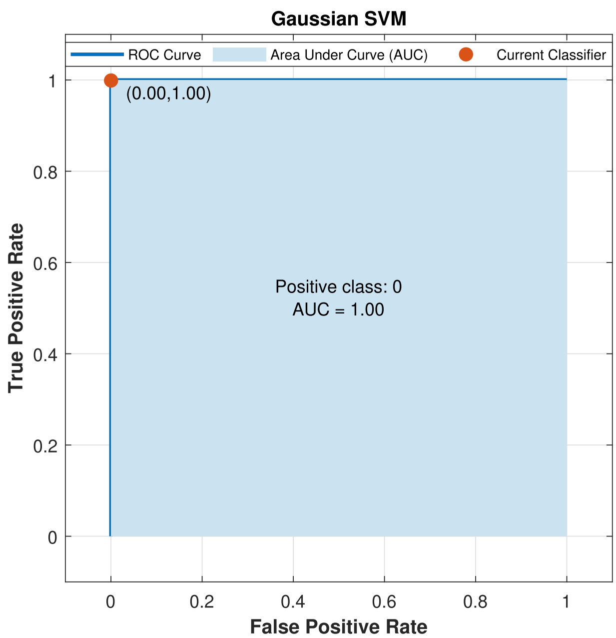

Fig. 11 – Area under Curve of SVM classifier with Gaussian kernel tricks.

obtained with the methods presented in Table 4 for three-class classification.

Two-class with PVC classification methods are proposed in [73] using wavelet-based statistical process control (SPC); [74] use SVM and [75] use the ANN classifier. Other two-class classification methods are proposed in [22] with kNN; [44] apply CNN and RRN; [40] use intelligent decision rules; [21] apply MLP and CNN; [42] use Naïve bayes; and [43] use SVM for normal and abnormal class. Our proposed method performance is marginally ahead of the method presented in Table 4 for two-class comparison in terms of detection accuracy. The methods proposed by [76,77] use four classes for the classification, using the SVM and DT. The method proposed in [1,16] considers eight different classes using 2D spectrograms and wavelet coefficients for classification. Other methods proposed in [19] use 16 classes with a general sparsed neural network (GSNN), while [20] utilises 17 classes with SVM and [41] uses 6 classes, with the DT classifier for the classification. All methods reported in Table 4 utilised the MIT-BIH AD for training, testing, and validation, except for [67,74,22,21]. Our proposed work's results are better than those of the other existing methods reported in Table 4. The method presented in [67] uses the MIMIC-III database, while [74] uses an artificially generated database for testing and validation purposes, [22] utilises the PTB database; [21] does not report the database used. The results achieved with our proposed method are better due to the use of the RR interval as a primary feature and the application of statistical features and a balanced dataset that clearly differentiates normal, PVC, and PAC beats in the classification. The proposed method is implemented on MATLAB 2020a software using the system configuration of i5-8th generation CPU @1.8 GHz with 16-GB memory. The computational time is calculated in terms of elapsed time range (ETR) for each algorithm. The ETRs for the KNN, SVM, RF, NB, and DT algorithms are calculated as 1.16s, 17.4s, 7.6s, 1.04s, and 1.32s, respectively.

Highlights of this work are as follows:

- We propose an improved RR interval-based PVC and PAC classification method using a machine-learning approach.
- The ECG signal was processed using the discrete wavelet transform.
- The proposed method uses the RR interval as a primary feature and eight different statistical features to evaluate performance.

The limitations and future scope of this work are as follows:

- Other primary features along with the statistical features can be used to classify other cardiac arrhythmias.
- A hybrid machine-learning approach can be utilised for further performance improvement.
- This experiment only considered the intra-patient validation, the results cannot reflect the generalization performance of the model.

6. Conclusion

In this paper, we propose an improved cardiac arrhythmia classification method using an RR interval-based approach. The ECG signal was pre-processed using the DWT and two-stage median filter to remove the high-frequency noises and baseline wander present in the raw ECG signal. The normal, PVC, and PAC beats were extracted from the QRS region, followed by detection of the QRS complex in the processed ECG signal. Next, we extracted nine different features, including the primary feature (the RR interval), from the beat segments. Five classifiers namely, k-NN, SVM, DT, NB, and RF were utilised to obtain the results, using the extracted fea-

Table 4 – Comparison of the proposed method with other existing state-of-the-art methods.

Authors	Features	Database	Classifier	C	F	Acc %
Proposed Work	RR interval + Statistical Features	MIT-BIH AD	SVM, Random Forest	3	09	99.51
Bashar et al.[67]	Density Poincare + statistical feature	MIMIC-III	SVM, k-NN, Random Forest	3	79	97.45
Inan et al.[68]	DWT + Timing Interval	MIT-BIH AD	NN	3	42	95.2
Chiu et al.[69]	Correlation coefficient	MIT-BIH AD	NR	3	20	99.30
Liang et al.[70]	HMM	MIT-BIH AD	Two-layer HMM	3	02	99.14
Akin et al.[71]	DWT	MIT-BIH AD	ANN	3	06	98
Liu et al.[72]	Energy + RR Interval	MIT-BIH AD	LVQ NN	3	02	98.90
Jung et al.[73]	Statistical + DWT	MIT-BIH AD	Wavelet SPC	2	02	97.9
de Oliveira et al.[74]	Geometric Feature	Artificial Data	SVM	2	12	99.5
Allami et al.[75]	Morphological + Statistical feature	MIT-BIH AD	ANN	2	10	98.6
Chen et al.[76]	PCA + DWT	MIT-BIH AD	SVM	4	19	97.80
Alarsan et al.[77]	DWT + Temporal Features	MIT-BIH AD	DT, Random Forests, GBT	4	16	98.03
Ullah et al.[1]	2-D Spectrograms	MIT-BIH AD	CNN	08	256X256	99.11
Çínar et al.[15]	Spectrogram Image	MIT-BIH AD	Hybrid Alexnet-SVM	03	9216	96.77
Jha et al.[16]	TQWT (Tunable Q-wavelet Trans.) Coefficients	MIT-BIH AD	SVM	08	12	99.27
Kirti et al.[39]	Heart Rate Variability	MIT-BIH AD	ANN	03	20	99
Sanamdikar et al.[19]	Time and Frequency Domain features	MIT-BIH AD	GSNN	16	20	98
Heo et al.[22]	3D ECG ST-T region	PTB Database	KNN	02	02	96.37
Plawiak et al.[20]	Power Spectral Density	MIT-BIH AD	Evolutionary-neural-based SVM	17	04 W	98.85
Andersen et al.[44]	RR intervals + l-beat Segmentation	CNN + RRN	02	30	97.80±0.61	
Arumugam et al.[40]	Time domain features	MIT-BIH AD	DWT and intelligent decision rule	02	08	NR
Sahoo et al.[41]	EMD + VMD + RR intervals	MIT-BIH AD	C4.5 Decision Tree Classifier	06	15	98.89
Savalia et al.[21]	ECG data	NR	MLP + CNN	02	60	88.7
Leutheuser et al.[42]	RR intervals	MIT-BIH AD	Naïve Bayes	02	03	90
Huang et al.[43]	Random projections, intervals	MIT-BIH AD	SVM + Threshold	02	15	94.6

C = Number of Classes, F = Number of features, W = Window, NR = Not Reported.

tures dataset. We employed the MIT-BIH AD database to train and test the proposed method. The 10-cross fold validation method was used to validate the performance of the classifiers. The overall performances of the SVM and the Random Forest classifier are better than those of the other classifiers used in this method. The obtained results demonstrate that the proposed method can be used for automatic cardiac arrhythmia classification of PAC and PVC beats.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Ullah A, Anwar S, Bilal M, Mehmood R. Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation. Remote Sens 2020;12(10):1685.
[2] Organization WH. World health statistics 2016: Part ii: Global health indicators [Tech. rep.]. World Health Organization; 2016.
[3] Sharma LD, Sunkaria RK. Stationary wavelet transform based technique for automated external defibrillator using optimally selected classifiers. Measurement 2018;125:29–36.
[4] Rahul J, Sora M, Sharma LD. An overview on biomedical signal analysis. Int J Recent Technol Eng 2019;7:206–9.
[5] Sharma LD, Sunkaria RK. Detection and delineation of the enigmatic u-wave in an electrocardiogram. Int J Inf Technol 2019:1–8.
[6] Rahul J, Sora M. A novel adaptive window based technique for t wave detection and delineation in the ecg. Bio-Algorithms Med-Syst 16(1). .
[7] Raj S, Ray KC, Shankar O. Development of robust, fast and efficient qrs complex detector: a methodological review. Australas Phys Eng Sci Med 2018;41(3):581–600.
[8] Sharma LD, Sunkaria RK. Novel t-wave detection technique with minimal processing and rr-interval based enhanced efficiency. Cardiovasc Eng Technol 2019;10(2):367–79.
[9] Coviello JS. ECG interpretation made incredibly easy! Lippincott Williams & Wilkins; 2020.
[10] Kusumoto F. ECG interpretation: from pathophysiology to clinical application. Springer Nature; 2020.
[11] Gomes JA. Signal averaged electrocardiography: concepts, methods and applications. Springer Science & Business Media; 2012.
[12] Kochi AN, Tagliari AP, Forleo GB, Fassini GM, Tondo C. Cardiac and arrhythmic complications in patients with covid-19. J Cardiovasc Electrophysiol 2020;31(5):1003–8.
[13] Houssein EH, Kilany M, Hassanien AE. Ecg signals classification: a review. Int J Intell Eng Inf 2017;5(4):376–96.
[14] Parvaneh S, Rubin J, Babaeizadeh S, Xu-Wilson M. Cardiac arrhythmia detection using deep learning: a review. J Electrocardiol 2019;57:S70–4.

[15] Çinar A, Tuncer SA. Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ecg signals using lstm and hybrid cnn-svm deep neural networks. Comput Methods Biomech Biomed Eng 2020:1–12.
[16] Jha CK, Kolekar MH. Cardiac arrhythmia classification using tunable q-wavelet transform based features and support vector machine classifier. Biomed Signal Process Control 2020;59 101875.
[17] Lim JS. Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system. IEEE Trans Neural Networks 2009;20(3):522–7.
[18] Faziludeen S, Sabiq P. Ecg beat classification using wavelets and svm. In 2013 IEEE Conference on Information & Communication Technologies, IEEE, 2013. pp. 815–818. .
[19] Sanamdikar ST, Hamde ST, Asutkar VG. Analysis and classification of cardiac arrhythmia based on general sparsed neural network of ecg signals. SN Appl Sci 2020;2(7):1–9.
[20] Pławiak P, Abdar M. Novel methodology for cardiac arrhythmias classification based on long-duration ecg signal fragments analysis. In Biomedical Signal Processing, Springer, 2020. pp. 225–272. .
[21] Savalia S, Emamian V. Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks. Bioengineering 2018;5(2):35.
[22] Heo J, Lee JJ, Kwon S, Kim B, Hwang SO, Yoon YR. A novel method for detecting st segment elevation myocardial infarction on a 12-lead electrocardiogram with a three-dimensional display. Biomed Signal Process Control 2020;56 101700.
[23] Jenny NZN, Faust O, Yu W. Automated classification of normal and premature ventricular contractions in electrocardiogram signals. J Med Imag Health Inf 2014;4 (6):886–92.
[24] Yousefi MR, Khezri M, Bagheri R, Jafari R. Automatic detection of premature ventricular contraction based on photoplethysmography using chaotic features and high order statistics. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2018. pp. 1–5..
[25] Hock TS, Faust O, Lim T-C, Yu W. Automated detection of premature ventricular contraction using recurrence quantification analysis on heart rate signals. J Med Imag Health Inf 2013;3(3):462–9.
[26] Malek AS, Elnahrawy A, Anwar H, Naeem M. Automated detection of premature ventricular contraction in ecg signals using enhanced template matching algorithm. Biomed Phys Eng Express 2020;6(1) 015024.
[27] Khalaf AF, Owis MI, Yassine IA. A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines. Expert Syst Appl 2015;42 (21):8361–8.
[28] Kim J, Shin HS, Shin K, Lee M. Robust algorithm for arrhythmia classification in ecg using extreme learning machine. Biomed Eng Online 2009;8(1):31.
[29] Sangaiah AK, Arumugam M, Bian G-B. An intelligent learning approach for improving ecg signal classification and arrhythmia analysis. Artif Intell Med 2020;103 101788.
[30] Van Steenkiste G, van Loon G, Crevecoeur G. transfer learning in ecg classification from human to horse using a novel parallel neural network architecture. Scientific Rep 2020;10 (1):1–12.
[31] Elgendi M, Jonkman M, De Boer F. Premature atrial complexes detection using the fisher linear discriminant. In: 2008 7th IEEE International Conference on Cognitive Informatics, IEEE. p. 83–8.
[32] Gayathri S, Suchetha M, Latha V. Ecg arrhythmia detection and classification using relevance vector machine. Proc Eng 2012;38:1333–9.
[33] He R, Liu Y, Wang K, Zhao N, Yuan Y, Li Q, Zhang H. Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional lstm. IEEE Access 2019;7:102119–35.
[34] Tripathy RK, Zamora-Mendez A, De la O Serna JA, Paternina MRA, Arrieta JG, Naik GR. Detection of life threatening ventricular arrhythmia using digital taylor fourier transform. Front Physiol 2018;9:722.
[35] Xiang Y, Lin Z, Meng J. Automatic qrs complex detection using two-level convolutional neural network. Biomed Eng Online 2018;17(1):1–17.
[36] De Chazal P. Detection of supraventricular and ventricular ectopic beats using a single lead ecg. In Conference proceedings:...Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, vol. 2013, 2013. p. 45. .
[37] Manikandan MS, Ramkumar B, Deshpande PS, Choudhary T. Robust detection of premature ventricular contractions using sparse signal decomposition and temporal features. Healthcare Technol Lett 2015;2(6):141–8.
[38] Cai Z, Li J, Johnson AE, Zhang X, Shen Q, Zhang J, Liu C. Rule-based rough-refined two-step-procedure for real-time premature beat detection in single-lead ecg. Physiol Meas. .
[39] Kirti S, Harsh J. Shruti, Multistage classification of arrhythmia and atrial fibrillation on long-term heart rate variability. J Eng Sci Technol 2020;15:1277–95.
[40] Arumugam M, Sangaiah AK. Arrhythmia identification and classification using wavelet centered methodology in ecg signals. Concurrency Comput: Practice Exp 2020;32(17) e5553.
[41] Sahoo S, Subudhi A, Dash M, Sabut S. Automatic classification of cardiac arrhythmias based on hybrid features and decision tree algorithm. Int J Autom Comput 2020:1–11.
[42] Leutheuser H, Gradl S, Eskofier BM, Tobola A, Lang N, Anneken L, Arnold M, Achenbach S. Arrhythmia classification using rr intervals: improvement with sinusoidal regression feature, in. In: 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), IEEE. p. 1–5.
[43] Huang H, Liu J, Zhu Q, Wang R, Hu G. A new hierarchical method for inter-patient heartbeat classification using random projections and rr intervals. Biomed Eng Online 2014;13(1):90.
[44] Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. Expert Syst Appl 2019;115:465–73.
[45] Moody GB, Mark RG. The impact of the mit-bih arrhythmia database. IEEE Eng Med Biol Mag 2001;20(3):45–50.
[46] PhysioBank P. Physionet: components of a new research resource for complex physiologic signals. Circulation v101 i23. e215–e220. .
[47] Lin H-Y, Liang S-Y, Ho Y-L, Lin Y-H, Ma H-P. Discrete-wavelet-transform-based noise removal and feature extraction for ecg signals. Irbm 2014;35(6):351–61.
[48] Moore AW, Jorgenson JW. Median filtering for removal of low-frequency background drift. Anal Chem 1993;65(2):188–91.
[49] Sharma LD, Sunkaria RK. A robust qrs detection using novel pre-processing techniques and kurtosis based enhanced efficiency. Measurement 2016;87:194–204.
[50] Sharma T, Sharma KK. Qrs complex detection in ecg signals using locally adaptive weighted total variation denoising. Comput Biol Med 2017;87:187–99.
[51] Sharma A, Patidar S, Upadhyay A, Acharya UR. Accurate tunable-q wavelet transform based method for qrs complex detection. Comput Electr Eng 2019;75:101–11.

[52] Rahul J, Sora M, Sharma LD. Exploratory data analysis based efficient qrs-complex detection technique with minimal computational load. Phys Eng Sci Med 2020;43(3):1049–67.

[53] Spector ZZ, Meliones C, Idriss SF. arrhythmias and pacing. In: Ungerleider RM, Meliones JN, Nelson McMillan K, Cooper DS, Jacobs JP, editors. Critical Heart Disease in Infants and Children (Third Edition), 27. Philadelphia: Elsevier; 2019. p. 326–350.e3. https://doi.org/10.1016/B978-1-4557-0760-7.00027-9. URL: https://www.sciencedirect.com/science/article/pii/B9781455707607000279.

[54] Saini I, Singh D, Khosla A. Qrs detection using k-nearest neighbor algorithm (knn) and evaluation on standard ecg databases. J Adv Res 2013;4(4):331–44.

[55] Melgani F, Bazi Y. Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans Inf Technol Biomed 2008;12 (5):667–77.

[56] Sahayadhas A, Sundaraj K, Murugappan M, Palaniappan R. A physiological measures-based method for detecting inattention in drivers using machine learning approach. Biocybern Biomed Eng 2015;35(3):198–205.

[57] Wieben O, Afonso VX, Tompkins WJ. Classification of premature ventricular complexes using filter bank features, induction of decision trees and a fuzzy rule-based system. Med Biol Eng Comput 1999;37(5):560–5.

[58] Kropf M, Hayn D, Schreier G. Ecg classification based on time and frequency domain features using random forests. In: 2017 Computing in Cardiology (CinC), IEEE, 2017. pp. 1–4..

[59] Porwik P, Orczyk T, Lewandowski M, Cholewa M. Feature projection k-nn classifier model for imbalanced and incomplete medical data. Biocybern Biomed Eng 2016;36 (4):644–56.

[60] Sharma LD, Sunkaria RK. Stationary wavelet transform based technique for automated external defibrillator using optimally selected classifiers. Measurement 2018;125:29–36.

[61] Aggarwal Y, Das J, Mazumder PM, Kumar R, Sinha RK. Heart rate variability features from nonlinear cardiac dynamics in identification of diabetes using artificial neural network and support vector machine. Biocybern Biomed Eng. .

[62] Asghar MA, Khan MJ, Rizwan M, Mehmood RM, Kim S-H. An innovative multi-model neural network approach for feature selection in emotion recognition using deep feature clustering. Sensors 2020;20(13):3765.

[63] Wiggins M, Saad A, Litt B, Vachtsevanos G. Evolving a bayesian classifier for ecg-based age classification in medical applications. Appl Soft Comput 2008;8(1):599–608.

[64] Pathan S, Kumar P, Pai R, Bhandary SV. Automated detection of optic disc contours in fundus images using decision tree classifier. Biocybern Biomed Eng 2020;40(1):52–64.

[65] Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. Random forests for classification in ecology. Ecology 2007;88(11):2783–92.

[66] Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. Pattern Recogn Lett 2006;27 (4):294–300.

[67] Bashar S, Han D, Fearass Z, Ding E, Fitzgibbons T, Walkey A, McManus D, Javidi B, Chon K. Novel density poincare plot based machine learning method to detect atrial fibrillation from premature atrial/ventricular contractions. IEEE Trans Biomed Eng. .

[68] Inan OT, Giovangrandi L, Kovacs GT. Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Trans Biomed Eng 2006;53(12):2507–15.

[69] Chiu C-C, Lin T-H, Liau B-Y. Using correlation coefficient in ecg waveform for arrhythmia detection. Biomed Eng: Appl Basis Commun 2005;17(03):147–52.

[70] Liang W, Zhang Y, Tan J, Li Y. A novel approach to ecg classification based upon two-layered hmms in body sensor networks. Sensors 2014;14(4):5994–6011.

[71] Akin ZE, Bilgin S. Classification of normal beat, atrial premature contraction and ventricular premature contraction based on discrete wavelet transform and artificial neural networks. In: 2017 Medical Technologies National Congress (TIPTEKNO), IEEE, 2017. pp. 1–4..

[72] Liu X, Du H, Wang G, Zhou S, Zhang H. Automatic diagnosis of premature ventricular contraction based on lyapunov exponents and lvq neural network. Comput Methods Programs Biomed 2015;122(1):47–55.

[73] Jung Y, Kim H. Detection of pvc by using a wavelet-based statistical ecg monitoring procedure. Biomed Signal Process Control 2017;36:176–82.

[74] de Oliveira BR, de Abreu CCE, Duarte MAQ, Vieira Filho J. Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection. Comput Methods Programs Biomed 2019;169:59–69.

[75] Allami R. Premature ventricular contraction analysis for real-time patient monitoring. Biomed Signal Process Control 2019;47:358–65.

[76] Chen X, Wang Y, Wang L, et al. Arrhythmia recognition and classification using ecg morphology and segment feature analysis. IEEE/ACM Trans Comput Biol Bioinf 2018;16 (1):131–8.

[77] Alarsan FI, Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data 2019;6(1):1–15.