International Journal of Automation and Computing

17(4), August 2020, 551-561

DOI: 10.1007/s11633-019-1219-2

Automatic Classification of Cardiac Arrhythmias Based on Hybrid Features and Decision Tree Algorithm

Santanu Sahoo[1] Asit Subudhi[1] Manasa Dash[2] Sukanta Sabut[3]

[1]Department of Electronics and Communication Engineering, ITER, SOA University, Bhubaneswar 751030, India

[2]Department of Mathematics, Silicon Institute of Technology, Bhubaneswar 751024, India

[3]School of Electronics Engineering, KIIT deemed to be University, Bhubaneswar 751024, India

Abstract: Accurate classification of cardiac arrhythmias is a crucial task because of the non-stationary nature of electrocardiogram (ECG) signals. In a life-threatening situation, an automated system is necessary for early detection of beat abnormalities in order to reduce the mortality rate. In this paper, we propose an automatic classification system of ECG beats based on the multi-domain features derived from the ECG signals. The experimental study was evaluated on ECG signals obtained from the MIT-BIH Arrhythmia Database. The feature set comprises eight empirical mode decomposition (EMD) based features, three features from variational mode decomposition (VMD) and four features from RR intervals. In total, 15 features are ranked according to a ranker search approach and then used as input to the support vector machine (SVM) and C4.5 decision tree classifiers for classifying six types of arrhythmia beats. The proposed method achieved best result in C4.5 decision tree classifier with an accuracy of 98.89% compared to cubic-SVM classifier which achieved an accuracy of 95.35% only. Besides accuracy measures, all other parameters such as sensitivity (Se), specificity (Sp) and precision rates of 95.68%, 99.28% and 95.8% was achieved better in C4.5 classifier. Also the computational time of 0.65 s with an error rate of 0.11 was achieved which is very less compared to SVM. The multi-domain based features with decision tree classifier obtained the best results in classifying cardiac arrhythmias hence the system could be used efficiently in clinical practices.

Keywords: Electrocardiogram (ECG), cardiac arrhythmias, empirical mode decomposition (EMD), variational mode decomposition (VMD), hybrid features, decision tree classifier.

1 Introduction

The electrocardiogram (ECG) signal is an important tool for diagnosing the cardiac arrhythmias in clinical analysis. However, these signals are susceptible to various types of noises[1]. Different approaches using discrete wavelet transform (DWT), Hilbert transform and empirical mode decomposition (EMD) have been used successfully to detect the arrhythmias in ECG signals[2–5]. The wavelet and EMD methods were found to be effective for reducing noises from the ECG signals[6]. Feature extraction and classification are the most intensive points for the detection of cardiac arrhythmias[7]. In recent decades, several works have investigated the automatic detection of ECG beats. The performance of the classifiers mainly depends upon the time and frequency domain features, which can be extracted from the morphology of the ECG waveforms. Many efficient methods based on the autoregressive model coefficients[8], higher-order statistics[9], wavelet transform[10], support vector machines (SVMs)[11] are used to extract valuable features from the ECG signal. Integration of DWT and artificial neural network is used for correct classification of arrhythmia conditions[12].

The wavelet transform based time interval and morphological features were used for successful classification of the ECG beats[13]. Mitra et al.[14] proposed a three-stage technique based on stationary wavelet transform (SWT) for noise reduction, use of morphological and time interval features and a classification module to classify premature ventricular contraction (PVC) from other heart disorders. Huang et al.[15] introduced the EMD for analysis of the non-linear and non-stationary signals. EMD decomposes adaptively into intrinsic mode functions (IMFs) components that de-noises ECG signals. In a combined approach consisting of EMD and DWT, algorithms have been used for accurate reduction of noise from the ECG signals[16]. However, the EMD algorithm is sensitive to noise and computational time is high. Dragomiretskiy and Zosso[17] proposed the variational mode decomposition (VMD) model which is an alternative to the EMD algorithm for signal analysis. This technique is more robust than the EMD for sampling and de-noising the non-linear and non-stationary signals like ECG. Recently, the variational mode decomposition (VMD) have been used for de-noising the ECG signal and the detecting the cardiac arrhythmias[18, 19]. Maji et al.[20] used frequency domain features based on VMD for de-

Research Article

Manuscript received October 6, 2019; accepted December 18, 2019; published online March 5, 2020

Recommended by Associate Editor Hong Qiao

tecting the ventricular QRS complex of the ECG signal. The time-frequency domain features of EMD modes are helpful in classifying the ECG betas. The spectral energy of the ECG signal is observed in the range 0.5 Hz–30 Hz in the frequency domain analysis and the clinical information of the ECG waveforms lies in this range[11]. Thomas et al.[21] used dual-tree complex wavelet transform (DTCWT) based features for automatic classification of cardiac arrhythmias. Afkhami et al.[22] used morphological and statistical information to classify cardiac arrhythmias accurately. The optimum-path forest (OPF) is also used for fast and accurate arrhythmia detection[23]. Combined approach based on EMD and approximate envelope provides better R-peak detection of the EMG signal[24]. The dominant rescaled wavelet coefficients (DRWC) enhances the R-peaks and QRS complex by reducing the effects of other peaks of ECG[25]. Better denoising was done by Symlets wavelet function with modified S-median thresholding in order to preserve the time-frequency information the ECG signal[26, 27]. In a recent work, an optimization mechanism is used for an automatic feature learning scheme in ECG beat classification. The method is based on a global recurrent neural network (GRNN) for classification of ECG beats using morphological and temporal based features. The GRNN scheme achieves significant improvement in detection rate as compared to other algorithms[28]. In this paper, we propose an improved method that combines the EMD and VMD approach for accurate classification of arrhythmias. The ECG signals are obtained from the MIT-BIH Arrhythmia Database and de-noised with bandpass filters. The morphological and statistical features based on RR intervals, EMD, and VMD techniques were extracted from the ECG signals. All the features were combined to form a feature set and then classified with SVM and decision trees classifiers to detect the six types of arrhythmia beats. The obtained results were compared to other methods for confirming the superiority of the proposed scheme.

2 Methodology

The flowchart of the proposed method for detection and classification of cardiac arrhythmia is shown in Fig. 1. The experimental study was evaluated on ECG signals obtained from the MIT-BIH Arrhythmia Database. The

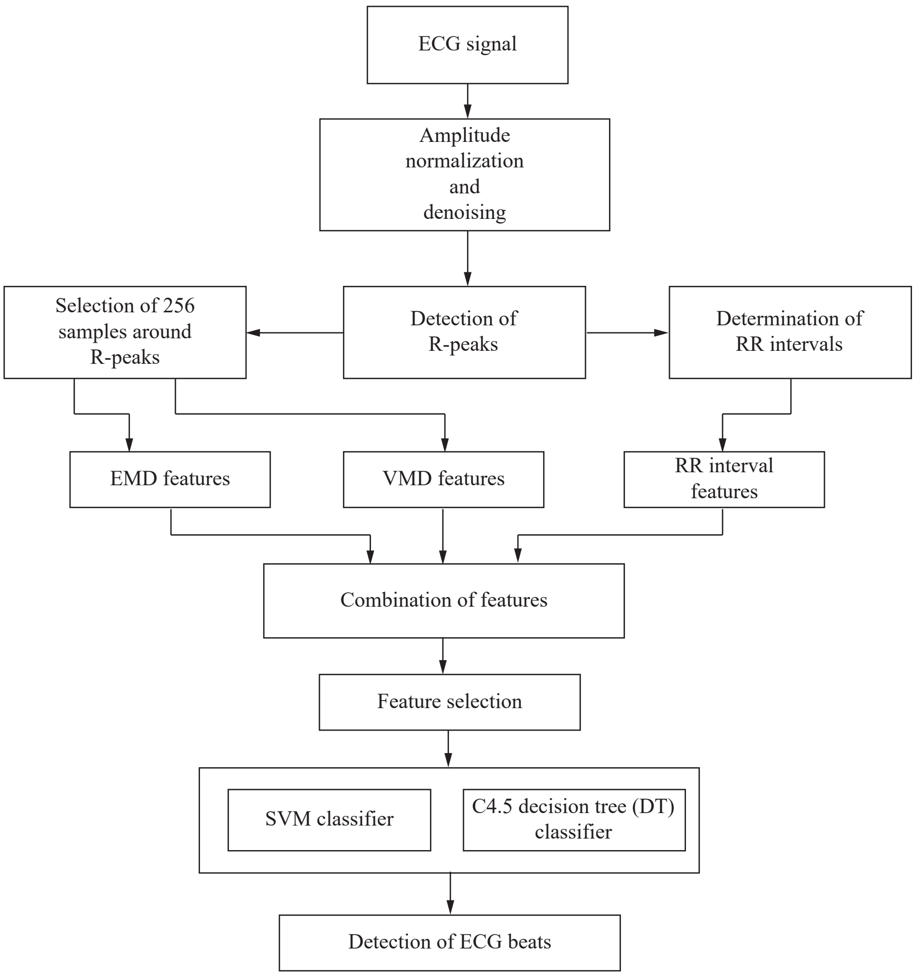

Fig. 1 Block diagram of the proposed method for detection of ECG beats

details of ECG beats used in this work are presented in Table 1. The method consists of de-noising the ECG signal, feature extraction and classification stages to classify six types of cardiac conditions. In the pre-processing stage, the signal was enhanced using a moving average filtering approach. Amplitude normalization is also included in this process. The R-peak of the pre-processed signal is performed using adaptive thresholding based Hilbert transform[29]. The features are extracted using RR intervals, EMD and VMD techniques and served as the input to the classifier that classifies the following types of ECG beats, i.e., normal (N), left bundle branch block (L), right bundle branch block (R), premature ventricular contraction (V), atrial premature beat (A) and paced beat (/). The feature set is comprised of eight morphological features from the 2nd and 3rd intrinsic mode function (IMF) of empirical mode decomposition (EMD), three features from the 3rd mode of variational mode decomposition (VMD) and four numbers of time-dependent features from RR intervals.

Table 1 Detail of ECG beats used in this study

Types of beat	Total records	No. of beats
Normal (N)	18	75052
Left bundle branch block (LBBB)	4	8075
Right bundle branch block (RBBB)	4	7259
Premature ventricular contraction (V)	12	7130
Atrial premature beat (A)	5	2546
Paced beats (/)	4	7028

2.1 Feature extraction

The information extracted from the heartbeat is considered to be the most important factor for successfully classifying the ECG beats. Various kind of time or frequency domain features are directly extracted from the heartbeats. Some of the published literature based on feature extraction is explained in this section. The most common feature calculated from the heartbeat is the RR interval, which is the interval between one to other heartbeats. The classification accuracy improves by considering the average RR interval and normalized RR interval[30]. The features based on variation of the QRS interval also provide good results[31]. The time domain features play an important role in improving the detection accuracy[32]. Various statistical features such as mean, standard deviation, energy and multi-domain features and higher order spectral features are used successfully in classification[9, 33, 34]. In this work, we have extracted the RR interval, EMD and VMD based features and combined them to make a feature set.

2.1.1 Empirical mode decomposition

The EMD method was first proposed by Haung et al.[15] for decomposition of non-linear and non-stationary signals into a finite and a small number of oscillations. EMD is spontaneous and adaptive with all IMFs which are derived from the dataset. Through shifting processes, the IMFs are estimated and its range varies from high frequency oscillations to low frequency oscillations. The low frequency oscillations are then decomposed into high frequency components and a residue, continuing the process till residue becomes monotonic.

1) The local maxima and local minima of a signal $w(t)$ are considered to construct the signal. Computation of mean of the signal is done by using the equation:

$$m(t) = \frac{w_u(t) + w_l(t)}{2}. \tag{1}$$

2) To get the difference, the mean is subtracted from the original signal which gives the first component $d(t)$ as

$$d(t) = w(t) - m(t). \tag{2}$$

3) If $d(t)$ satisfies the IMF conditions, then $f_1(t)$ contains the highest frequency components of the signal $w(t)$.

4) The second IMF component $d_1(t)$ is derived by the equation:

$$d_1(t) = d(t) - m(t). \tag{3}$$

5) The boundary conditions, at which iteration stops is

$$SD = \sum_{t=0}^{N} \frac{|d_{i-1}(t) - d_i(t)|^2}{d^2{}_{i-1}(t)}. \tag{4}$$

The residue of the signal can be calculated as

$$r(t) = w(t) - f_1(t). \tag{5}$$

6) For the n-th level decomposition, the signal is defined as

$$\tilde{w}(t) = \sum_{i=1}^{p-1} f_i(t) + r_n(t). \tag{6}$$

Different IMFs were derived from the signal by using the EMD approach. Only two IMFs are used for feature extraction, and a total of 8 features were extracted using this technique. By the following means, the features are extracted:

1) One cycle of ECG signal is obtained using a window of 256 samples from annotated R-peak.

2) The selected window is applied with EMD to extract required IMF up to the 4th level.

3) IMF 2 and IMF 3 were selected for required features of EMD because of its morphological similarities with QRS complex of ECG signal.

4) The following statistical parameters are obtained from the selected IMFs:

a) Mean
b) Maximum
c) Minimum
d) Standard deviation

2.1.2 Variational mode decomposition

It is a non-recursive method used to decompose the signal into different modes that comprises low and high frequency components[35]. The mathematical description of VMD is given as

$$\min\left\{\sum_l\left\|\partial_t\left[\left(\sigma\left(t\right)+\frac{\mathrm{j}}{\pi t}\right)\times u_l\left(t\right)\right]\mathrm{e}^{-\mathrm{j}w_lt}\right\|_2^2\right\}\tag{7}$$

$$\sum_l u_l\left(t\right)=p\left(t\right)\tag{8}$$

where $p\left(t\right)$ is the input signal, $w_l\left(t\right)$ is the center frequency. $\{u_l\}=\{u_1,u_2,\cdots,u_l\}$ is the decomposed mode and the dirac distribution is denoted by $\sigma\left(t\right)$. Langragian multipliers and quadratic penalty terms exhibit better convergence at a finite weight and strict enforcement respectively and hence these terms are introduced for the reconstruction problem. Thus, the augmented Langragian arguments Γ can be expressed as

$$\Gamma\left(\{u_l\},\{w_l\},\lambda\right)=\alpha\sum_l\left\|\partial_t\left[\left(\sigma\left(t\right)+\frac{\mathrm{j}}{\pi t}\right)\times u_l\left(t\right)\right]\mathrm{e}^{-\mathrm{j}w_lt}\right\|_2^2+\\ \left\|p\left(t\right)-\sum_l u\left(t\right)\right\|_2^2+\lambda\left(t\right),p\left(t\right)-\sum_l u\left(t\right)\tag{9}$$

where λ is the dual ascent and α is the Lagrange multiplier. The minimization problem given in 7 is solved by an alternate direction method of multipliers (ADMM) by finding the saddle point of the augmented Langragian Γ.

As per the algorithm expressed, the next step is to update ${u_l}^n$ and w_l^n to ${u_l}^{n+1}$ and w_l^{n+1}, respectively. Equation (2) can be optimized with respect to ${u_l}^n$ as

$${u_l}^{n+1}=\underset{u_l\epsilon x}{\mathrm{argmin}}\left\{\alpha\sum_l\left\|\partial_t\left[\left(\sigma\left(t\right)+\frac{\mathrm{j}}{\pi t}\right)\times u_l\left(t\right)\right]\mathrm{e}^{-\mathrm{j}w_lt}\right\|_2^2+\\ \left\|p\left(t\right)-\sum_i u_i\left(t\right)+\frac{\lambda\left(t\right)}{2}\right\|_2^2\right\}.\tag{10}$$

Using Parseval Fourier isometric and Hermitian symmetry property, the equation is further optimized to:

$$\hat{u}_l^{n+1}\left(w\right)=\frac{\hat{p}\left(w\right)-\sum\limits_{i\neq l}\hat{u}_l\left(w\right)+\dfrac{\hat{\lambda}\left(w\right)}{2}}{1+2\propto\left(w-w_l\right)^2}\tag{11}$$

$$w_l^{n+1}\leftarrow\frac{\displaystyle\int_0^\infty w|\hat{u}_l\left(w\right)|^2\mathrm{d}w}{\displaystyle\int_0^\infty|\hat{u}_l\left(w\right)|^2\mathrm{d}w}.\tag{12}$$

Symbols with ^ indicate the frequency domain variables.

The signal is decomposed into different modes by the VMD approach based on frequency level. It is not affected by noise because it separates high frequency components at higher modes. On applying VMD, the input signals are decomposed into their modes consisting of the lowest frequency at mode 1 and high frequency noise at mode 4. The P-wave and T-wave of the ECG signal are finally eliminated and the high frequency QRS region of ECG signal is enhanced which is required for the analysis[20]. The complete decomposition process of the ECG signal is given in Fig. 2. In this method, the feature extraction techniques are as follows:

1) A single cycle of ECG signal is considered by selecting a window of 256 samples around R-peak.

2) Apply VMD approach up to the 4th level to obtain the required mode.

3) Select only mode 3 for analysis as it has morphological similarities with QRS complex.

4) From the selected mode, the features like skewness and kurtosis were obtained.

2.1.3 RR interval features

An adaptive threshold approach is preferred in detecting R-peaks in each stage as the limits of up and down threshold are not equal[15, 33]. The error is calculated and subtracted from the limits and the procedure continues until both the threshold limits become equal to obtain the final threshold. In this work, we have chosen RR interval as time domain features. Here we have taken the pre-RR interval feature which is defined as the difference between current and previous $\mathrm{R_{Loc}}$, the post-RR interval is defined as the difference between the current and following $\mathrm{R_{Loc}}$. The 1-min and 20-min averaged RR intervals are denoted by $\mathrm{RR_1}$ and $\mathrm{RR_2}$ which are presented by averaging the RR interval of 1 and 20 minutes, respectively.

2.2 Feature selection

Feature selection is a crucial step for choosing the relevant features subset that improves the classification accuracy[17]. Different features were extracted from the decomposed signal which is given in Table 2 and then ranked by the gain ratio attribute evaluation scheme for ranking them according to their weightage value to obtain the optimal features to improve the detection accuracy.

2.3 Classifier

2.3.1 Support vector machine

A support vector machine is a supervised learning

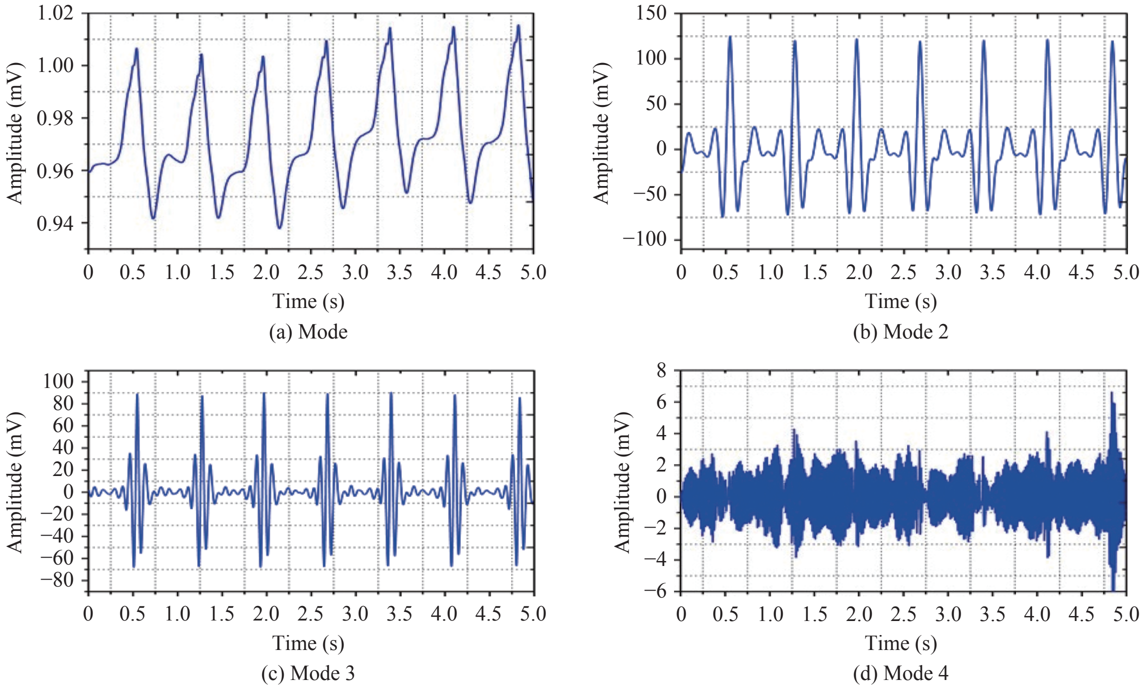

Fig. 2 Decomposition of ECG signal using VMD

Table 2 List of features extracted for the heartbeat classification

Features	Description	Features	Description
$RR_{(i)}$	Pre RR interval	$E_{min,2}$	Min of IMF_2 using EMD
$RR_{(i+1)}$	Post RR interval	$E_{max,3}$	Min of IMF_3 using EMD
RR_1	One minute average RR-interval	$E_{sd,2}$	Standard deviation of IMF_2 using EMD
RR_2	Twenty minute average RR-interval	$E_{sd,3}$	Standard deviation IMF_3 using EMD
$E_{m,2}$	Mean of IMF_2 (within QRS region) using EMD	$V_{kurt,3}$	Kurtosis of Mode 3 using VMD
$E_{m,3}$	Mean of IMF_3 (within QRS region) using EMD	$V_{sk,3}$	Skewness of Mode 3 using VMD
$E_{max,2}$	Max of IMF_2 using EMD	$V_{sd,3}$	Standard deviation of Mode 3 using VMD
$E_{max,3}$	Max of IMF_3 using EMD		

model used to classify the data set. It is a discriminative non-linear network structure that separates the data into different classes by a hyperplane[36]. The best classification accuracy can be obtained by minimizing the error in the training set and maximizing the margin between classes. The cubic SVM model has been used in this work that classifies six different types of ECG beats.

We consider that the training data have N pairs, i.e., $(x_1,y_1), (x_2,y_2), \cdots, (x_N,y_N)$ with $x_i \in \mathbf{R}^p$ and $y_i \in \{-1, 1\}$.

Let a hyper-plane be $\left\{x : f(x) = x^{\mathrm{T}}\alpha + \beta_0 = 0\right\}$, where α is a unit vector,

$$P(x) = \mathrm{sgn}\left[x^{\mathrm{T}}\alpha + \alpha_0\right] C. \tag{13}$$

2.3.2 C4.5 algorithm

C4.5 builds decision trees using the concept of information entropy from a set of training data as in done with the same way of an ID3 algorithm of decision tree classifier[37]. The learning in the decision tree classifies starts at the top point of the data and then moves down leaf according to the discrete values. The value at that leaf node provides the predicted output obtained from the value of the leaf node which is based on the learning process of the algorithms. The classification of the data starts at the root node and decision is made at each branch of the tree in order to predict the output class from the feature set. A decision tree is made based on the training instances of the ID3 algorithm. The decision making process improves by the tree with the use of if-then rules of the algorithm. The best classification characteristic is selected at each node best on highest information gain from the set of test attributes. The gain in information is based on the reduction in entropy that splits the data based on attribute values. The information gain at a node for an attribute A is calculated by the equation

$$Gain\,(I,A) = Entropy\,(E) - \sum_{v \in Values(A)} \left(\frac{|I_q|}{|I|} Entropy\,(S_q) \right) \quad (14)$$

where I is the set of instances, S_q is the subset of I for which attribute A has value q, and entropy of I is calculated as

$$Entropy\,(I) = -\sum_{i=1}^{C} P\log_2 P_i \quad (15)$$

where P_i is the proportion of orders in I that have the i-th class value as output attribute.

3 Evaluation results

The main objective of this paper was to detect the arrhythmia beats using feature selection and classification techniques in ECG signals. The performance of the classifier is evaluated in terms of its sensitivity (Se), specificity (Sp), positive predictivity (Pr) and accuracy (Acc) using a confusing matrix and stand equations. The hidden information of the ECG signals can be extracted in the frequency domain representation by EMD and VMD techniques. In this work, the R-peaks were detected using an adaptive thresholding based EMD technique. A window of 256 samples was selected around R-peaks to get one cardiac cycle. Applying EMD and VMD approaches across this window, the signal is decomposed into different frequency bands and a set of discriminating features were extracted from the decomposed signals. In the EMD approach, we have selected only IMF_2 and IMF_3 because of its morphological waveform is similar to the QRS complex. Each IMFs were evaluated by its mean, maximum value, minimum value and standard deviation. A total of eight features has been selected as a set of EMD features. The VMD technique is also proposed within the selected window to get required modes. The decomposition of the ECG signal is limited to four modes so as to avoid required high frequency noises. Mode three of the decomposed signal contains maximum morphology of the QRS complex, hence this mode is considered for feature extraction. The standard deviation, kurtosis and skewness from mode 3 is obtained and arranged to form a set of VMD features. The RR interval features give important information about heart rhythms. In this study, four RR intervals were extracted which include pre-RR interval, post-RR interval, 1-min and 20-min averaged RR intervals.

A combination of RR-interval, EMD and VMD based features were combined to get a set of 15 features which are then applied to the input of the classifier. Table 2 shows the list of extracted features from one cardiac cycle. We have used the gain ratio attribute evaluation scheme for getting the best discriminating features from the feature set. Table 3 shows the result of feature selection as per their weightage ranking. The selected features are classified with the C4.5 decision tree and a SVM classifiers in a ten-fold cross-validation process. A total of seven sets are used for training and remaining three sets are used for testing the classifier performance. The individual average performance of different cardiac beats (N, L, R, V, A, and /) is shown in Table 4. The performance evaluation results of the proposed method in 10-fold cross-validation are shown in Table 5 for both the classifiers. It was observed that the average results of the C4.5 algorithm yields better result than SVM classifier in terms of accuracy and computational time. Figs. 3–5 represent the result plot for sensitivity, specificity and accuracy of the classifiers. It is observed that with a confidence factor (CF) of 0.5, the C4.5 classifier gives better result than the cubic kernel-based SVM classifier. It is observed that the C4.5 classifier yields better results with accuracy of 98.89% at fold-6 compared and for the SVM classifier it is about 95.28% at fold-8 of for classifying six types of ECG beats. The computation time required to

Table 3 Ranking of features using gain ratio evaluation

Weightage of feature	Rank	Feature name	Weightage of feature	Rank	Feature name
0.35	1	$E_{m,3}$	0.10	9	$E_{min,2}$
0.26	2	$V_{kurt,3}$	0.10	10	RR_1
0.24	3	$V_{sd,3}$	0.09	11	$RR_{(i+1)}$
0.22	4	$V_{sk,3}$	0.08	12	RR_2
0.20	5	$E_{sd,3}$	0.07	13	$E_{sd,2}$
0.18	6	$E_{max,3}$	0.06	14	$E_{max,2}$
0.15	7	$E_{min,3}$	0.05	15	$RR_{(i)}$
0.12	8	$E_{m,2}$			

Table 4 Average result of each class in SVM and C4.5 classifiers

Methods	Class	Acc (%)	Se (%)	Sp (%)	Pr (%)
SVM	N	90.44	88.13	92.59	91.71
	L	97.55	86.09	99.07	92.44
	R	98.26	88.22	99.64	97.07
	V	92.26	69.53	94.28	51.99
	A	97.43	82.15	99.20	92.26
	/	95.77	88.79	96.50	72.65
C4.5 DT	N	98.48	98.37	98.58	98.36
	L	99.36	96.76	99.68	97.36
	R	99.45	97.39	99.70	97.58
	V	98.07	91.40	98.88	90.87
	A	99.20	95.49	99.58	95.89
	/	98.77	94.64	99.32	94.77

Abbreviation: Normal (N), left bundle branch block (L), right bundle branch block (R), premature ventricular contraction (V), atrial premature beat (A) and paced beats(/).

Table 5 Comparison results between C4.5 and cubic-SVM classifier at each fold in 10-fold cross-validation process

Folds	C4.5						Cubic-SVM					
	Se (%)	Sp (%)	Pr (%)	Acc (%)	Time (s)	MAE	Se (%)	Sp (%)	Pr (%)	Acc (%)	Time (s)	MAE
2	94.52	99.11	94.76	98.62	0.75	0.02	83.21	96.77	82.38	95.11	3.66	0.23
3	95.18	99.21	95.53	98.78	0.69	0.01	83.68	96.85	82.85	95.24	2.42	0.23
4	95.94	99.34	95.92	98.95	0.63	0.01	83.89	96.89	82.98	95.29	2.38	0.23
5	95.76	99.29	95.98	98.91	0.66	0.01	83.85	96.88	83.06	95.29	2.42	0.23
6	95.98	99.36	96.19	98.99	0.64	0.01	83.89	96.9	83.16	95.32	2.56	0.23
7	95.96	99.32	96.05	98.94	0.61	0.01	84.01	96.91	83.21	95.33	2.58	0.23
8	95.82	99.32	95.81	98.92	0.61	0.01	84.05	96.92	83.27	95.35	2.25	0.23
9	95.98	99.32	95.85	98.92	0.63	0.01	83.89	96.9	83.09	95.31	2.48	0.23
10	95.94	99.33	96.13	98.96	0.63	0.01	83.89	96.9	83.09	95.31	2.29	0.23
Average	95.67	99.29	95.80	98.89	0.65	0.011	83.82	96.88	83.01	95.28	2.56	0.23

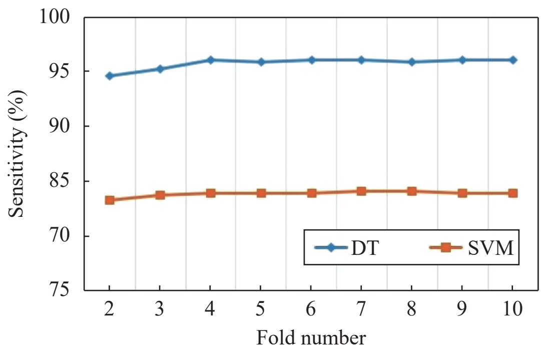

Fig. 3 Sensitivity for ECG beats at 10-fold cross validation scheme

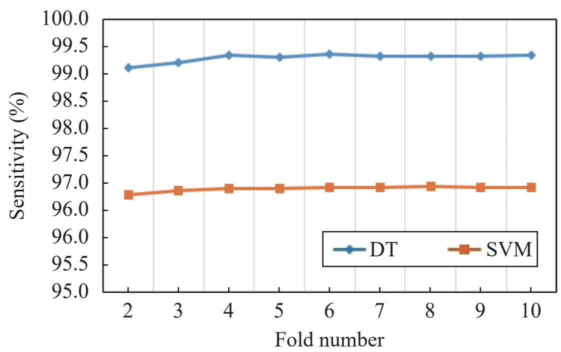

Fig. 4 Specificity performance of the classifiers

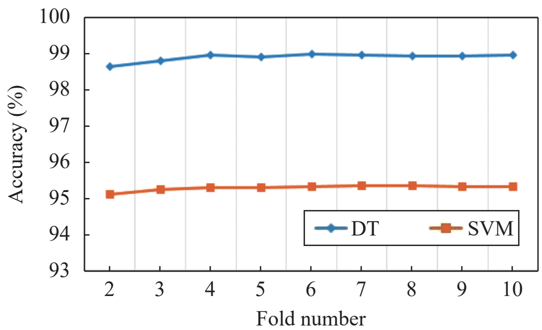

Fig. 5 Accuracy for DT and SVM classifier

build the classifiers at different folds is shown in Fig. 6. The computational time for the C4.5 classifier was of 0.65 s with error rate of 0.11 which is very less compared to SVM classifier that shows the high computational time of 2.56 s with detection error rate of 0.23. The performance of mean absolute error (MAE) and root mean square error values of both the classifiers is plotted in Fig. 7. It indicates that the decision tree algorithm provides better results with respect to all measured parameters with least errors in classifying the arrhythmia conditions.

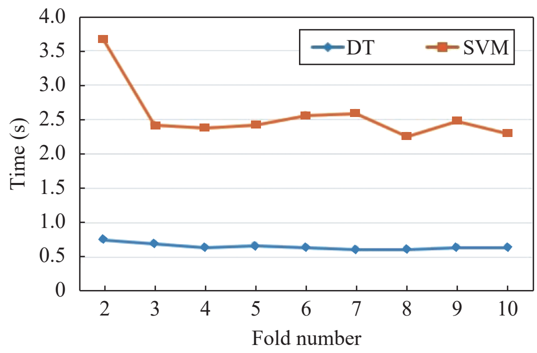

Fig. 6 Computational time of the classifiers

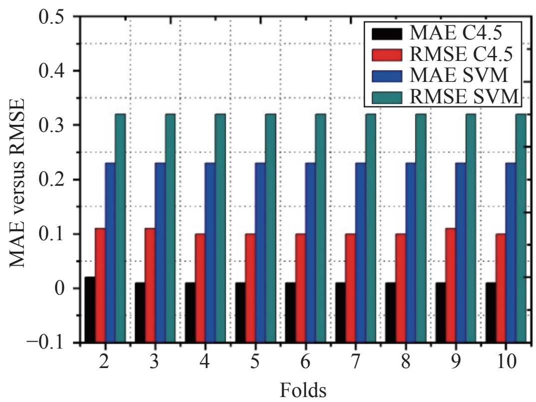

Fig. 7 Mean absolute error and root mean square error of the classifier

4 Discussions

De-noising and extracting important features from time-varying ECG signals are the important tasks in developing an automated system for detection of cardiac arrhythmias. Past studies have emphasized on different de-noising and classification methods to classify the ECG signals. It has been seen from the obtained result that the wavelet coefficients, QRS morphology and statistical based features are the best suitable approaches in order to discriminate the arrhythmia conditions. The time and frequency domain features based on Teager energy functions achieved the classification accuracy of 95% in classifying five types of beats using the neural network (NN) classifier[38]. The DWT based features along with the combined neural network classifier achieved an accuracy of 96.94% in classifying four types of ECG betas[39]. Martis et al.[40] achieved an accuracy of 93.48% with the use of morphological and time-based features and classified five ECG beats by the LSSVM classifier. In a study, Osowski et al.[41] presented a reliable heartbeat recognition system based on the higher order statistics (HOS) and Hermite characterization of QRS complex features to classify the ECG signals with the SVM classifier. They obtained the classification accuracy of 77% and 94.26% for Hermite preprocessing and HOS respectively with a SVM classifier. The usage of multiclass SVM together with discrete cosine transform improves the classification accuracy of 95.2% in classifying ECG beats[42].

In 2008, Asl et al.[43] classified six types of arrhythmias with an accuracy of 97.65% using heart rate variability (HRV) features and the NN classifier. In 2010, Fei[44] identified four types of ECG betas with an accuracy of 95.65% in the SVM classifier using time interval bi-orthogonal spline wavelet features. Kaya et al.[45] proposed an efficient method for detection of cardiac arrhythmias using feature reduction approaches and classified with based SVM and K-nearest neighbors (K-NN) classifiers. The genetic algorithm, principal component analysis (PCA) and independent component analysis (ICA) methods were used for feature reduction. They achieved an accuracy of 98.86% and 99.11% using PCA, ICA features using K-NN classifier and an accuracy of 98.92% was obtained in SVM classifier with the use of statistical and temporal features. In recent applications, the decision tree classifier is successfully applied in biomedical signal analysis. The best classification accuracy of 97.98% was obtained by random forest classifier with implementing ML-libs and Scala language on the Apache Spark framework[45]. A hybrid model has been reported to classify cardiac arrhythmias where the genetic algorithm was used for selecting optimal features, and the decision tree with the C4.5 algorithm was applied to classify and train the model. The method achieved the accuracy values of 86.96%, and sensitivity of 88.88% for the two-class mode[46]. In this work, we proposed an automated method based on signal de-noising, extraction of important informative features with EMD and VMD techniques and classified the feature set with the C4.5 decision tree and cubic-SVM classifier. The method achieved high accuracy of 98.89% in the decision tree and 95.28% in the cubic-SVM classifier, which is higher than the previously published results as reported in Table 6.

5 Conclusions

Despite enormous research works in the area of ECG analysis for automatic detection of cardiac arrhythmias, still a method is required for precise analysis to classify ECG beats by using time-frequency based features. The purpose of this work was to investigate the automatic approach based on a decision tree classifier in classifying ECG beats in terms of accuracy. We proposed a robust method for detection of cardiac arrhythmia conditions using time-frequency based features and a decision tree classifier. The detection scheme is based on denoising of ECG

Table 6 Comparison of results with the published article

Literature	Methods	Classifier	Classes	Accuracy (%)
Li et al.[7]	DWT and multi domain features	SVM	5	98.8
Kamath[38]	Teager energy function features	NN	5	95.0
Martis et al.[40]	Principal components of bispectrum features	SVM	5	93.48
Gulera and Ubeyli[39]	Statistical features	NN	4	96.94
Osowski et al.[41]	HOS features	SVM	13	94.26
Acir[42]	Best features set	LSSVM	6	95.2
Fei[44]	Time intervals features	SVM	5	95.65
Asl et al.[43]	Heart rate variability features	SVM	4	97.65
Alarsan and Younes[46]	Machine learning approach	Random forest	6	97.98
Ayar and Sabamoniri[47]	Genetic algorithm	Decision tree	2	86.96
Proposed method	EMD, VMD and RR interval features	Decision tree	6	98.89

signals and extracting important morphological features using EMD and VMD and time-domain features using RR intervals techniques. The combined features are given to the input of cubic-SVM and decision tree classifiers to perform classification of the different arrhythmias. Results illustrate that extracted important features yield the best detection accuracy of 98.89% in the decision tree classifier, which indicates its superiority in detecting cardiac arrhythmia. Thus, the presented automated approach can be effectively used for detection of cardiac arrhythmias. Further analysis can be done by the use of deep learning approaches for better accuracy at a faster rate in decision making systems in cardiac disease diagnosis.

References

[1] H. Sivaraks, C. A. Ratanamahatana. Robust and accurate anomaly detection in ECG artifacts using time series motif discovery. *Computational and Mathematical Methods in Medicine*, vol. 2015, Article number 453214, 2015. DOI: 10.1155/2015/453214.

[2] O. El B′charri, R. Latif, K. Elmansouri, A. Abenaou, W. Jenkal. ECG signal performance de-noising assessment based on threshold tuning of dual-tree wavelet transform. *BioMedical Engineering OnLine*, vol. 16, no. 2, Article number 26, 2017. DOI: 10.1186/s12938-017-0315-1.

[3] E. C. Erkus, V. Purutcuoglu, E. Detection of abnormalities in heart rate using multiple Fourier transforms. *International Journal of Environmental Science and Technology*, vol. 16, no. 9, pp. 5237–5242, 2019. DOI: 10.1007/s13762-019-02252-3.

[4] H. Sharma, K. K., Sharma. ECG-derived respiration based on iterated Hilbert transform and Hilbert vibration decomposition. *Australasian Physical & Engineering Sciences in Medicine*, vol. 41, no. 2, pp. 429–443, 2018. DOI: 10.1007/s13246-018-0640-0.

[5] D. D. Taralunga, M. Ungureanu, B. Hurezeanu, I. Gussi, R. Strungaru. Empirical mode decomposition applied for non-invasive electrohysterograhic signals denoising. In *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Milan, Italy, pp. 4134–4137, 2015. DOI: 10.1109/EMBC.2015.7319304.

[6] N. Q. Li, P. Li. An improved algorithm based on EMD-wavelet for ECG signal de-noising. In *Proceedings of International Joint Conference on Computational Sciences and Optimization*, IEEE, Sanya, China, pp. 825–827, 2009. DOI: 10.1109/CSO.2009.178.

[7] H. Q. Li, D. Y. Yuan, Y. X. Wang, D. Y. Cui, L. Cao. Arrhythmia classification based on multi-domain feature extraction for an ECG recognition system. *Sensors*, vol. 16, no. 10, Article number 1744, 2016. DOI: 10.3390/s16101744.

[8] D. F. Ge, B. P. Hou, X. J. Xiang. Study of feature extraction based on autoregressive modeling in EGG automatic diagnosis. *Acta Automatica Sinica*, vol. 33, no. 5, pp. 462–466, 2007. DOI: 10.1360/aas-007-0462.

[9] Y. Kutlu, D. Kuntalp. Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients. *Computer Methods and Programs in Biomedicine*, vol. 105, no. 3, pp. 257–267, 2012. DOI: 10.1016/j.cmpb.2011.10.002.

[10] S. C. Saxena, V. Kumar, S. T. Hamde. Feature extraction from ECG signals using wavelet transforms for disease diagnostics. *International Journal of Systems Science*, vol. 33, no. 13, pp. 1073–1085, 2002. DOI: 10.1080/002077202 10167159.

[11] F. Alonso-Atienza, E. Morgado, L. Fernandez-Martinez, A. Garcia-Alberola, J. L. Rojo-Alvarez. Detection of life-threatening arrhythmias using feature selection and support vector machines. *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 832–840, 2014. DOI: 10.1109/TBME.2013.2290800.

[12] Y. Ozbay. A new approach to detection of ECG arrhythmias: Complex discrete wavelet transform based complex valued artificial neural network. *Journal of Medical Systems*, vol. 33, no. 6, pp. 435–445, 2009. DOI: 10.1007/s10916-008-9205-1.

[13] S. Sahoo, B. Kanungo, S. Behera, S. Sabut. Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities. *Measurement*, vol. 108, pp. 55–66, 2017. DOI: 10.1016/j.measurement.2017.05.022.

[14] S. Mitra, M. Mitra, B. B. Chaudhuri. A rough-set-based inference engine for ECG classification. *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 6, pp. 2198–2206, 2006. DOI: 10.1109/TIM.2006.884279.

[15] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. A. Zheng, N. C. Yen, C. C. Tung, H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998. DOI: 10.1098/rspa.1998.0193.

[16] M. A. Kabir, C. Shahnaz. Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains. *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 481–489, 2012. DOI: 10.1016/j.bspc.2011.11.003.

[17] K. Dragomiretskiy, D. Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014. DOI: 10.1109/TSP.2013.2288675.

[18] U. Maji, S. Pal, S. Majumder. Estimation of arrhythmia episode using variational mode decomposition technique. In *Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, IEEE, Pisa, Italy, pp. 767–772, 2015. DOI: 10.1109/I2MTC.2015.7151365.

[19] A. Mert. ECG feature extraction based on the bandwidth properties of variational mode decomposition. *Physiological Measurement*, vol. 37, no. 4, pp. 530–543, 2016. DOI: 10.1088/0967-3334/37/4/530.

[20] U. Maji, M. Mitra, S. Pal. Detection and characterisation of QRS complex in VMD domain. In *Proceedings of Michael Faraday IET International Summit*, IET, Kolkata, India, pp. 586–590, 2015. DOI: 10.1049/cp.2015.1697.

[21] M. Thomas, M. K. Das, S. Ari. Automatic ECG arrhythmia classification using dual tree complex wavelet based features. *AEU – International Journal of Electronics and Communications*, vol. 69, no. 4, pp. 715–721, 2015. DOI: 10.1016/j.patrec.2015.11.018.

[22] R. G. Afkhami, G. Azarnia, M. A. Tinati. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recognition Letters*, vol. 70, pp. 45–51, 2016. DOI: 10.1016/j.dsp.2006.11.009.

[23] E. S. da S Luz, T. M. Nunes, V. H. C. de Albuquerque, J. P. Papa, D. Menotti. ECG arrhythmia classification based on optimum-path forest. *Expert Systems with Applications*, vol. 40, no. 9, pp. 3561–3573, 2013. DOI: 10.1016/j.eswa.2012.12.063.

[24] H. Q. Li, X. F. Wang, L. Chen, E. B. Li. Denoising and R-peak detection of electrocardiogram signal based on EMD and improved approximate envelope. *Circuits, Systems, and Signal Processing*, vol. 33, no. 4, pp. 1261–1276, 2014. DOI: 10.1007/s00034-013-9691-3.

[25] A. Ghaffari, H. Golbayani, M. Ghasemi. A new mathematical based QRS detector using continuous wavelet transform. *Computers & Electrical Engineering*, vol. 34, no. 2, pp. 81–91, 2008.

[26] L. Su, G. L. Zhao. De-noising of ECG signal using translation-invariant wavelet de-noising method with improved thresholding. *IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, Shanghai, China, pp. 5946–5949, 2005. DOI: 10.1109/IEMBS.2005.1615845

[27] M. A. Awal, S. S. Mostafa, M. Ahmad, M. A. Rashid. An adaptive level dependent wavelet thresholding for ECG denoising. *Biocybernetics and Biomedical Engineering*, vol. 34, no. 4, pp. 238–249, 2014. DOI: 10.1016/j.bbe.2014.03.002.

[28] G. J. Wang, C. S. Zhang, Y. P. Liu, H. Z. Yang, D. P. Fu, H. Q. Wang, P. Zhang. A global and updatable ECG beat classification system based on recurrent neural networks and active learning. *Information Sciences*, vol. 501, pp. 523–542, 2019. DOI: 10.1016/j.ins.2018.06.062.

[29] H. Rabbani, M. P. Mahjoob, E. Farahabadi, A. Farahabadi. R peak detection in electrocardiogram signal based on an optimal combination of wavelet transform, Hilbert transform, and adaptive thresholding. *Journal of Medical Signals & Sensors*, vol. 1, no. 2, pp. 91–98, 2011. DOI: 10.4103/2228-7477.95292.

[30] C. Ye, M. T. Coimbra, B. V. K. V. Kumar. Arrhythmia detection and classification using morphological and dynamic features of ECG signals. In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, Buenos Aires, Argentina, pp. 1918–1921, 2010. DOI: 10.1109/IEMBS.2010.5627645.

[31] P. de Chazal, M. O'Dwyer, R. B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004. DOI: 10.1109/TBME.2004.827359.

[32] M. Kropf, D. Hayn, G. Schreier. ECG classification based on time and frequency domain features using random forests. In *Proceedings of Computing in Cardiology*, IEEE, Rennes, France, 2017. DOI: 10.22489/CinC.2017.168-168.

[33] S. Sahoo, P. Das, P. Biswal, S. Sabut. Classification of heart rhythm disorders using instructive features and artificial neural networks. *International Journal of Medical Engineering and Informatics*, vol. 10, no. 4, pp. 359–381, 2018. DOI: 10.1504/IJMEI.2018.095085.

[34] Z. J. Yao, J. Bi, Y. X. Chen. Applying deep learning to individual and community health monitoring data: A survey. *International Journal of Automation and Computing*, vol. 15, no. 6, pp. 643–655, 2018. DOI: 10.1007/s11633-018-1136-9.

[35] E. Ebrahimzadeh, M. S. Manuchehri, S. Amoozegar, B. N. Araabi, H. S. Soltanian-Zadeh. A time local subset feature selection for prediction of sudden cardiac death from ECG signal. *Medical & Biological Engineering & Computing*, vol. 56, no. 7, pp. 1253–1270, 2018. DOI: 10.1007/s11517-017-1764-1.

[36] S. Siuly, V. Bajaj, A. Sengur, Y. C. Zhang. An advanced analysis system for identifying alcoholic brain state through EEG signals. *International Journal of Automation and Computing*, vol. 16, no. 6, pp. 737–747, 2019. DOI: 10.1007/s11633-019-1178-7.

[37] W. M. Chong, C. Le Goh, Y. T. Bau, K. C. Lee. Fast numerical threshold search algorithm for C4.5. In *Proceedings of the 3rd International Conference on Advanced Applied Informatics*, IEEE, Kitakyushu, Japan, pp. 930–935, 2014. DOI: 10.1109/IIAI-AAI.2014.183.

[38] C. Kamath. ECG beat classification using features extracted from Teager energy functions in time and frequency domains. *IET Signal Processing*, vol. 5, no. 6, pp. 575–581, 2011. DOI: 10.1049/iet-spr.2010.0138.

[39] I. Guler, E. D. Ubeylı. ECG beat classifier designed by combined neural network model. *Pattern Recognition*, vol. 38, no. 2, pp. 199–208, 2005. DOI: 10.1016/j.patcog.2004.06.009.

[40] R. J. Martis, U. R. Acharya, K. M. Mandana, A. K. Ray, C. Chakraborty. Cardiac decision making using higher order spectra. *Biomedical Signal Processing and Control*, vol. 8, no. 2, pp. 193–203, 2013. DOI: 10.1016/j.bspc.2012.08.004.

[41] S. Osowski, L. T. Hoai, T. Markiewicz. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 582–589, 2004. DOI: 10.1109/TBME.2004.824138.

[42] N. Acır. Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm. *Neural Computing & Applications*, vol. 14, no. 4, pp. 299–309, 2005. DOI: 10.1007/s00521-005-0466-z.

[43] B. M. Asl, S. K. Setarehdan, M. Mohebbi. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, vol. 1, pp. 51–64, 2008. DOI: 10.1016/j.artmed.2008.04.007.

[44] S. W. Fei. Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine. *Expert Systems with Applications*, vol. 37, no. 10, pp. 6748–6752, 2010. DOI: 10.1016/j.eswa.2010.02.126.

[45] Y. Kaya, H. Pehlivan, M. E. Tenekeci. Effective ECG beat classification using higher order statistic features and genetic feature selection. *Biomedical Research*, vol. 28, no. 17, pp. 7594–7603, 2017.

[46] F. I. Alarsan, M. Younes. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *Journal of Big Data*, vol. 6, no. 1, Article number 81, 2019. DOI: 10.1186/s40537-019-0244-x.

[47] M. Ayar, S. Sabamoniri. An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm. *Informatics in Medicine Unlocked*, vol. 13, pp. 167–175, 2018. DOI: 10.1016/j.imu.2018.06.002.

Santanu Sahoo received the B. Eng. degree in electronics & communication engineering from Utkal University, India in 2004, the M. Eng. degree in communication system engineering and the Ph. D. degree in electronics engineering from Siksha "O" Anusandhan University, India in 2009 and 2018, respectively. He is working as an associate professor in Department of ECE, Institute of Technical Education & Research, Siksha "O" Anusandhan University, India. He has over 15 years of teaching and research experience.

His research interests include biomedical signal and machine learning approaches.

E-mail: santanusahoo@soa.ac.in
ORCID iD: 0000-0001-8324-3101

Asit Subudhi received the B. Eng. degree in electronics & communication from BPUT, India in 2005, the M. Eng. degree in communication system engineering from BPUT, India in 2010, and received the Ph. D. degree in electronic engineering from SOA University, India in 2018. Presently, he is working as associate professor in Department of ECE, ITER, SOA University, India. He has over 12 years of experience in teaching and his research expertise focuses on signal and image processing, VLSI design. He is a member of IEEE.

His research interests include signal and image processing.

E-mail: asitsubudhi@soa.ac.in

Manasa Dash received the M. Sc., M. Phil and Ph. D. degrees in mathematics in 1995, 1997 and 2014 respectively from Utkal University, India. Presently, she is working as an associate professor in Silicon Institute of Technology, India. She has published many papers in referred international journals and conferences. She has over 18 years of experience in both teaching and research.

Her research interests include optimization engineering, linear algebra, numerical methods, random fourier series, and probability & statistics.

E-mail: manasa@silicon.ac.in

Sukanta Sabut received the B. Eng. in electronics & communication engineering from Visvesvaraya Technological (VT) University of Karnataka, India in 2001, the M. Tech. degree in biomedical instrumentation engineering from the VT University of Karnataka, India in 2005, and the Ph. D. degree in medical science & technology from IIT Kharagpur, India in 2011. He was working as an associate professor, SOA University, India from 2013 to 2017. Presently, he is working as an associate professor in School of Electronics Engineering, KIIT Deemed to be University, India. He is a member of IEEE, IET, IFESS, Rehabilitation council of India and the Institution of Engineers (India). He is the author of more than 63 articles in reputed international journals and conferences. He has over 19 years of experience in both teaching and research.

His research interests include biomedical signal and image analysis, machine learning for medical, neural and rehabilitation engineering.

E-mail: sukanta207@gmail.com (Corresponding author)
ORCID iD: 0000-0003-2896-8331