



Decision Support System for Predicting Ventricular Arrhythmias Using Non-linear Features of ECG Signals

Monalisa Mohanty¹ · Pratyusa Dash² · Sukant Sabut³

Received: 2 May 2022 / Accepted: 14 February 2024
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2024

Abstract

Automated methods using computer-aided decision-making process are effectively used for timely detection of VAs which are the most life-threatening conditions. In this work, we have proposed a decision support system for detection of ventricular arrhythmias (VAs) with a low computational complexity using hybrid features derived from three different transform techniques i.e. DWT, EEMD, and VMD. The methods mainly consist of a windowing technique, signal decomposition, feature extraction, and classification. About 24 time–frequency based features were extracted with $22,721 \times 24$ and ranked for selection of higher ranked features having maximum information of disease. The reduced feature set consists of only 6 number of highly ranked features which are then classified with SVM and decision tree classifier for efficient recognition of VAs. Aim of the reduction in data size is to reduce the computational time. Our proposed method achieves high classification accuracy in hybrid-based features and by reducing the feature dimension, it reduces the computational complexity significantly. The accuracy of 99.83% and computational time of 4.82 s is achieved when considering all 24 features. In reduced feature set, an accuracy of 99.62% with a very less computational time of 2.71 s was obtained for decision tree classifier which indicates the importance of selecting the important features for classification of VAS. With superior classification accuracy and low computation complexity, this system can be utilized in clinical practice for the recognition of ventricular arrhythmias.

Keywords Sudden cardiac arrest · ECG signals · Decision support system · Hybrid features · Classifiers

Introduction

The studies have shown that the dynamics of the cardiac signal are very complex and non-linear in nature, therefore, the non-linear models are suitable for the analysis of ECG signals [1]. Each analysis methods to detect heart disease provides valuable diagnostic information with different conclusions requires further investigation in ECG signal analysis [2]. Hence some non-linear techniques have been developed for efficient detection of life-threatening ventricular arrhythmias such as VT/VF using ECG signal [3]. Sun et al. [4] proposed a non-linear method using the Hurst

index to characterize the VT and VF episode in ECG analysis. Both time and frequency-based features improves the classification of ECG beats [5]. Arafat et al. [6] proposed a threshold crossing sample count (TCSC) method which is in time domain analysis for accurate detection of VF episodes. Decomposition of the ECG signal using DWT, EMD, variational mode decomposition with machine learning techniques has been found to be effective in detecting cardiac arrhythmias [7–10].

The wavelet transform based features is an important approach in classifying the non-stationary ECG signals as it provides both time and frequency information of an ECG signal [11]. Rakshit et al. [12] have proposed an automatic method using wavelet transform and a Hilbert transform for R-peaks detection. The method achieved an average accuracy of 99.83%, with an error rate of 0.17%. Nazarahari et al. [13] presented a new wavelet functions WFs based on the hybrid approach that consists of GA-PSO framework for effective classification of the ECG signals. The discrete wavelet transforms and morphology-based features were successfully used to classify abnormal and normal class

✉ Sukant Sabut
sukanta207@gmail.com

¹ Department of Center of Internet of Things, ITER, SOA University, Bhubaneswar, India

² Department of Computer Science and Engineering, Heritage Institute of Technology, Kolkata, India

³ School of Electronics Engineering, KIIT Deemed to be University, Bhubaneswar, India

ECG signal with neural network classifier [14]. In a linear method, Martis et al. [15] applied the principal component analysis i.e., PCA based feature reduction technique in discrete wavelet transform (DWT) based features to discriminate between normal and arrhythmia classes. They achieved an overall classification accuracy of 98.78% with a neural network (NN) classifier. A novel method was presented by Garcia et al. [16] for automatic detection of AF episodes using the relative wavelet energy (RWE) of the TQ interval in the ECG signal. The method was able to detect the AF condition with an accuracy higher than 90% in less than 7 beats. A novel method based on stationary wavelet transform and SVM classifier was applied successfully for automatic detection of AF and achieves a sensitivity of 97.0% and specificity of 97.1%, respectively [17].

The empirical mode decomposition (EMD) has been also used for accurate detection of VF. In this method, the ECG signals are decomposed into different intrinsic mode functions (IMFs). Arafat et al. [18] have used EMD technique for decomposition of ECG signals and extracted orthogonality indices from the first three consecutive IMFs and used to detect VF condition. The method was applied for a window with a duration of 3 s and obtained the accuracy of detection as 99.70%. In an efficient process, the EMD method was used to discriminate between ventricular fibrillation and ventricular tachycardia with an accuracy of 91% using ECG signals [19]. Similarly, few studies have been reported the use of a variational mode of decomposition for effective application of detection of ventricular arrhythmias. Xu et al. [20] have proposed an adaptive-VMD process for decomposing the signals into five band-limited intrinsic modes (BLIMs). Then the feature extraction and classification were done using a Boosted-CART classifier to recognize the VT/VF condition and obtained an overall 98.29% of accuracy and 97.32% of sensitivity. Based on the non-linearity and nonstationary decomposition methods, Abdalla et al. [21] have developed a method called complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) for classification of arrhythmias using MIT-BIH database. It was used to find intrinsic mode functions (IMFs) and four different parameters were computed to construct the feature vector and classified with ANN classifier with an accuracy of 99.9%.

In this work we have proposed a hybrid feature set approach that combines the features based on DWT, EEMD, and VMD methods. The hybrid features having large dimension have been extracted to make a large dataset in order to achieve higher accuracy. The dimension of the feature set was reduced by selecting the useful informative features using a ranker search method. In order to balance between the reliable recognition system and computational burden, in this work, we have considered a hybrid feature set and ranked them to obtain the most significant information about

the VA conditions. The selected informative features are classified using the SVM and the decision tree classifier to recognize the life-threatening ventricular arrhythmias.

Methodology

The flow chart for the projected ventricular arrhythmia classification system using hybrid feature has been shown in Fig. 1. The electrocardiogram signals were acquired from the first channel of both CUDB and VFDB databases. The acquired ECG signals are first got pre-processed using a high pass ($f_c = 1$ Hz), low pass ($f_c = 30$ Hz), and a notch filter to eliminate the baseline wander, high-frequency noise

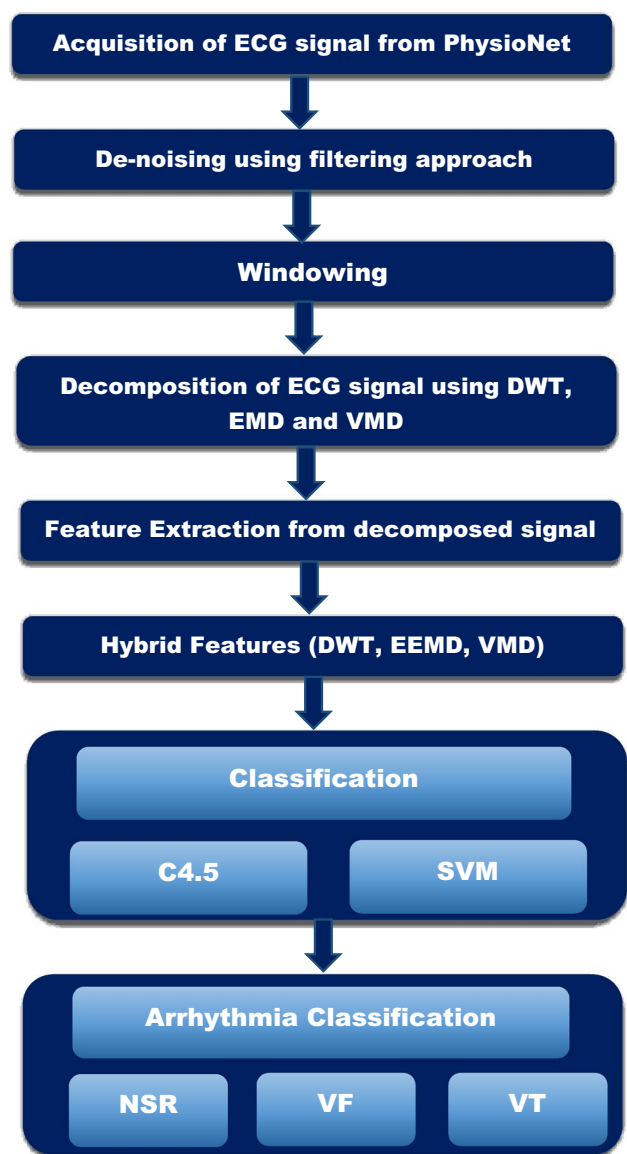


Fig. 1 The flow chart of proposed ventricular arrhythmia classification system

and power line interference respectively. Using a window length of 5 s, the acquired electrocardiogram signals have been decomposed using DWT, EEMD and VMD methods. From every mode, the 24 features have been extracted in the temporal, spectral, and statistical domain to make a large dataset having size $22,721 \times 24$. The dataset is then feed in machine learning approach for classification of the ventricular arrhythmias.

ECG Signal Decomposition

Discrete Wavelet Transform (DWT)

The wavelet transform has been broadly used for signal decomposition and feature extraction for the analysis of nonstationary signals like ECG [22]. In DWT approach, the signal is decomposed into the approximate coefficients as LPF and detail coefficients as HPF. The signal is then down sampled to half in each step. In this work, the ECG signal is decomposed into 8 levels by using Daubechies db6 mother wavelet transform. The decomposed signal of detail coefficients at D2-D8 and approximation coefficient A8 is consider for processing. In case of VT condition, the signal consists of detail coefficients (D1–D3) and approximation coefficient A3.

Let $X \rightarrow$ be the original signal

$n \rightarrow$ be the no. of decomposition levels

$A(n) \rightarrow$ be the approximation coefficients

$D(n) \rightarrow$ be the detail coefficients

$$S[n] = (X * A)[n] = \sum_{k=-\infty}^{\infty} X[k] A[n - k] \tag{1}$$

$$S_{low}[n] = \sum_{k=-\infty}^{\infty} X[k] A[2n - k] \tag{2}$$

$$S_{high}[n] = \sum_{k=-\infty}^{\infty} X[k] D[2n - k] \tag{3}$$

Ensemble Empirical Mode Decomposition (EEMD)

The effective time–frequency analysis of nonlinear and non-stationary signals, such as electrogram signals, can be done by using the EEMD technique [23]. Due to the signal intermittency, often the mode mixing problem arise which is the main bottleneck of the earlier EMD techniques. The EEMD technique provides a solution to the mode mixing problem by adding a finite noise to the original signal, while also retaining the decomposition information.

The decomposed data $Y(t)$ has been given as the intrinsic mode functions (IMFs) as:

$$Y(t) = \sum_{j=1}^n C_j \tag{4}$$

where the number of IMFs is given by n .

Then for successive observations, the modified data $Y_i(t)$ after adding random white noise has been given as:

$$Y_i(t) = Y(t) + N_i(t) \tag{5}$$

where N indicates the added white noise.

Based upon the decomposition, the required steps for the EEMD technique are given as follows:

- Initially, a random white noise has been added to the given data.
- Then the IMFs have been generated after decomposing the given data set.
- Repeat step 1 and step 2 by adding white noise series every time.
- Finally, obtain the mean of the signal i.e. ensemble of respective IMFs of the decomposition as the absolute result.

In the final mean of the corresponding IMFs, the added white noise series cancel each other. The average IMF resides within the filter windows and thus reduces the occurrence of mode mixing.

Variational Mode Decomposition (VMD)

The EMD approach is extensively used for recursive decomposition of a signal into various modes but has few limitations such as sampling and sensitivity to noise. Hence, to overcome such limitations, a non-recursive VMD model was developed which has the ability of concurrent extractions of modes. A signal, where the difference between the number of local extrema and zero-crossings is at most one, is defined as a mode [24, 25]. The IMFs of VMD are given as:

$$I_k(t) = \sum_{k=1}^N E_k(t) \text{Cos}(\theta_k(t)) \tag{6}$$

where N is the number of modes.

$\theta_k(t)$ is the phase ($\theta'_k(t) > 0$) and $E_k(t)$ is the envelope ($E_k(t) \geq 0$).

The variational substitute to the empirical mode decomposition or the empirical wavelet transform is given by VMD. This approach decomposes an input signal into different sub-signals i.e. IMFs which have the ability to reproduce the input with their sparsity properties.

Basically, 3 parameters are needed for VMD decomposition technique i.e. data fidelity constraint parameter (α), number of modes (N) and Lagrangian multiplier (λ). The parameter α is used to determine the frequency ranges in different modes. The parameter N has to be predefined for a signal in order to determine the number of decomposed modes. The parameter λ is used for quality check of reconstruction of an original signal from its IMFs.

The decomposition algorithm VMD is given as follows:

Step I. At count $n = 1$, the Lagrangian multiplier (λ_1) is initiated and the center frequency (w_k) and the spectrum (I_k) of the k th mode are generated.

Step II. For all modes i.e. for $k = 1$ to N,

- (a) the spectrum of the k th mode in the next iteration i.e. count = $n + 1$ is updated as:

$$I_k^{n+1}(t) = \frac{I(w) - \sum_{i < k} I_i^{n+1}(w) - \sum_{i > k} I_i^n(w) + ((\lambda^n(w)/2))}{1 + 2\alpha(w - w_k)^2} \tag{7}$$

- (b) For each mode, the center frequency is updated as follows:

$$w_k^{n+1} = \frac{\int_0^\infty w |I_k^{n+1}(w)|^2 dw}{\int_0^\infty |I_k^{n+1}(w)|^2 dw} \tag{8}$$

- (c) The Lagrangian multiplier is updated as follows:

$$\lambda_k^{n+1} = \lambda_k^n + \tau \left(I(w) - \sum_k I_k^{n+1}(w) \right) \tag{9}$$

Step III. Repeat step 2 until convergence occurs. The criterion for convergence is given as:

$$\sum_k \frac{\|I_k^{m+1} - I_k^m\|_2^2}{\|I_k^m\|_2^2} < \epsilon \tag{10}$$

Feature extraction

A total of 24 number of features having spectral, statistical and, temporal information conte have been extracted in order to classify the ventricular arrhythmia. The following features have been extracted to provide the precise classification results: Leakage Measure (LKG) [21, 26], Normalized spectral moment (FSMN, A1, A2, and A3), Auxiliary counts and Filter coefficient (C1, C2) [21, 26], Frequency Measure (FB) [21, 26], Covariance (CO) [21, 26], Area Measure (AB) [21, 26], Exponential Parameter (CROSS) [21, 26], Modified Exponential (MEA) [21, 26], Threshold crossing interval (TCI) [21, 26], Threshold crossing sample

count (TCSC) [21, 26], Hurst Parameter (Hurst) [27], Mean Absolute Value (VAL) [21, 26], Permutation Entropy (Perm) [28], Non-linear Features such as Shannon entropy (Shan_ent), Norm Entropy (Norm_ent), Log Entropy (Log_ent), Threshold Entropy (Th_ent) and, Sure Entropy (Sure_ent) [29], Kurtosis (Kur) [30], Skewness (Skw) [30]. Each and individual features are evaluated by the correlation attribute evaluation method and then ranked by a ranker search algorithm to improve the classification accuracy.

Feature Selection

The time–frequency domain features gather the relevant information in the signal for distinguishing the classes. Due to this reason, the feature selection and ranking play a key responsibility in the success of classification accuracy. To compute the order of variables according to their weightage value, a feature selection procedure called the Gain Ratio Attribute Evaluation method was used.

Gain Ratio Attribute Evaluation [31]: The information gain to the intrinsic information ratio is termed as information gain ratio i.e. IGR. Assuming T_r be the training examples set, A be the attributes set, and H be the entropy. The value (q,p) with $q \in T_r$ describes the worth of an accurate pattern q for attribute $p \in A$. The information gain for an attribute $p \in A$ is given as:

$$I_G(T_r, p) = H(T_r) - \sum_{v \in \text{value}(p)} \frac{|\{q \in T_r | \text{value}(q, p) = v\}|}{|T_r|} * H(\{q \in T_r | \text{value}(q, p) = v\}) \tag{11}$$

The intrinsic information is given as:

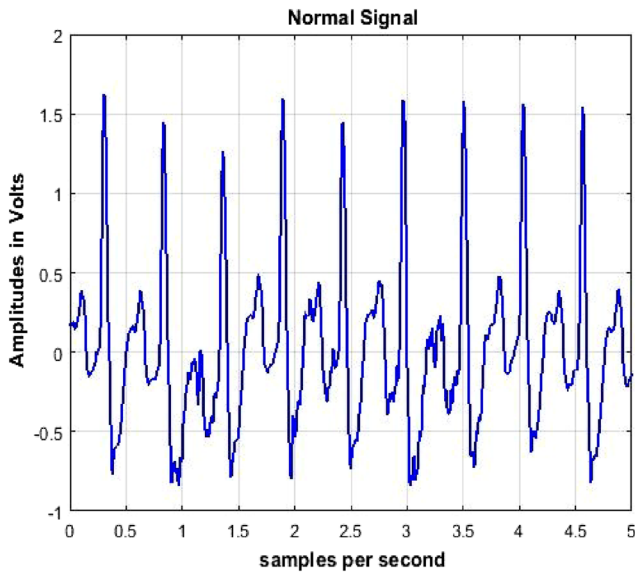
$$I_V(T_r, p) = - \sum_{v \in \text{value}(p)} \frac{|\{q \in T_r | \text{value}(q, p) = v\}|}{|T_r|} * \log_2 \left(\frac{|\{q \in T_r | \text{value}(q, p) = v\}|}{|T_r|} \right) \tag{12}$$

The information gain ratio is specified as:

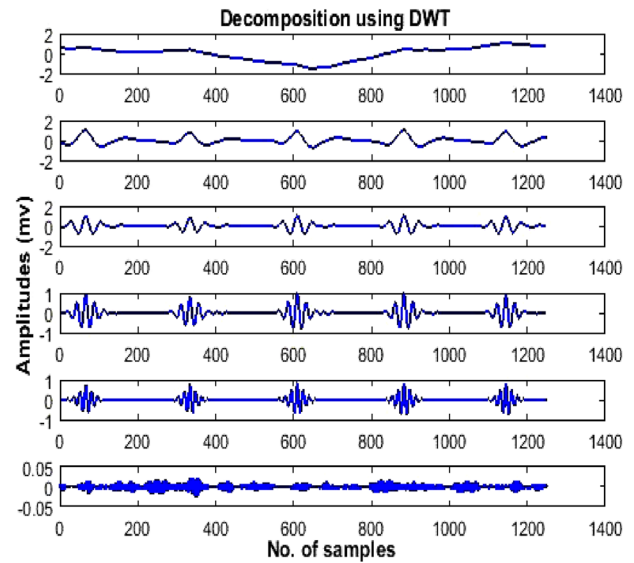
$$IGR(T_r, p) = \frac{I_G}{I_V} \tag{13}$$

Classification and Validation

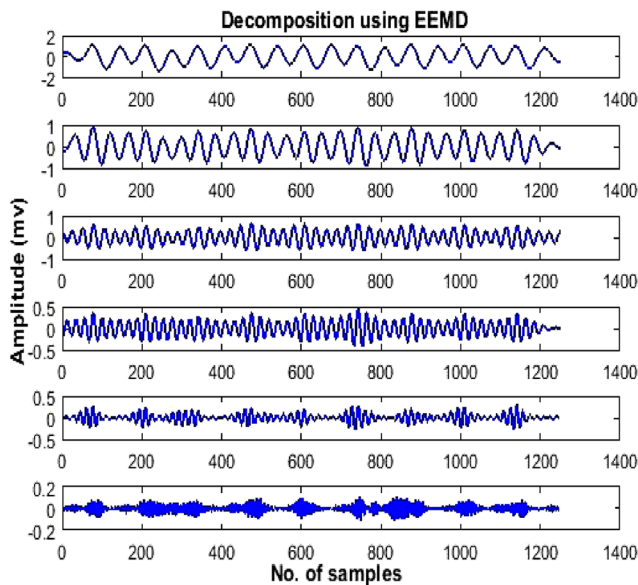
Since last decades, a number of approaches have been reported in the literature for the classification of cardiac arrhythmias using machine learning techniques. In proposed



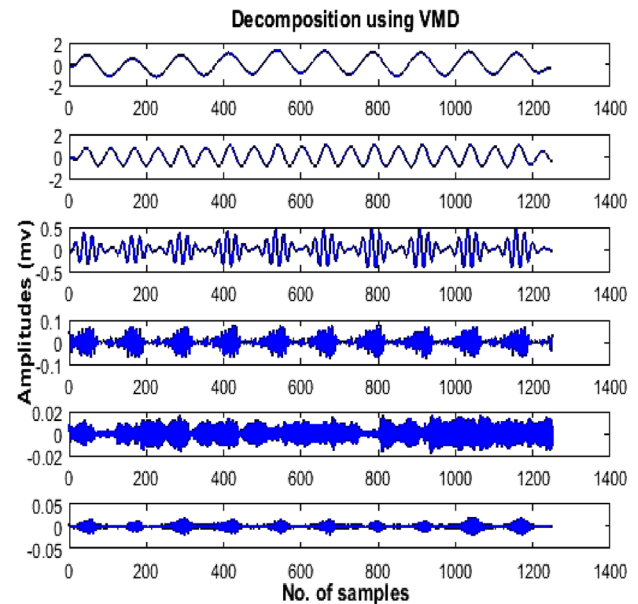
(a) Normal Signal



(b) Decomposition using DWT



(c) Decomposition using EEMD



(d) Decomposition using VMD

Fig. 2 For a window length of 5 s, this figure shows **a** normal signal, **b** DWT decomposition, **c** EEMD decomposition, and **d** VMD decomposition of record ‘cu03m’ of CUIDB database as an example of a normal signal

work, the extracted feature set was fed to the SVM and C4.5 classifiers for classification of VF, VT, and NSR arrhythmia conditions.

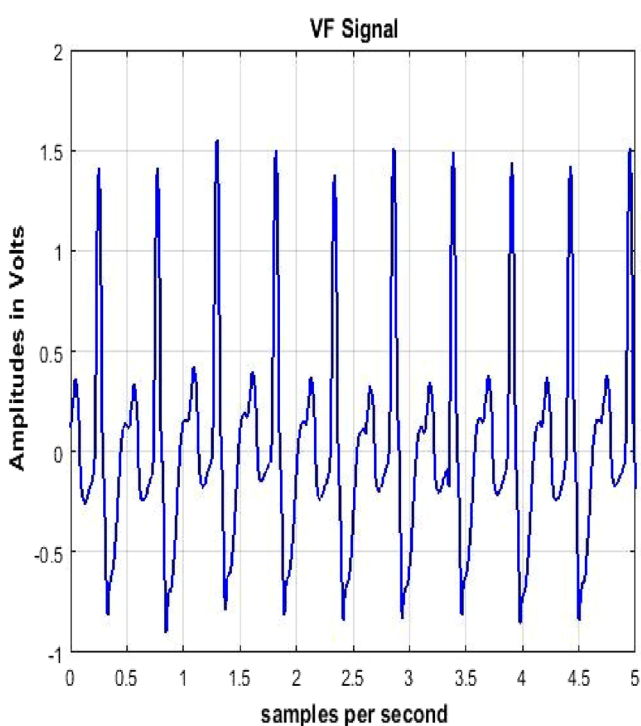
Support Vector Machine

Support vector machines are the supervised algorithms used in machine learning for signal classification [17]. The decision planes are used to outline the decision having

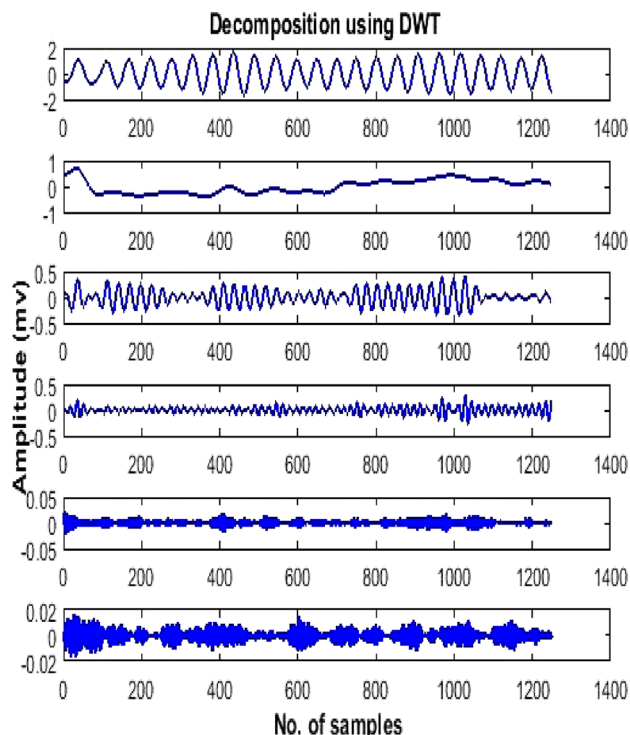
dissimilar class memberships. The hyperplane classifiers draw a line to discriminate between objects based on memberships functions. A hyperplane separates the variables into either class 0 or class 1.

Let the training data with N pairs i.e. $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ with $x_i \in \mathcal{R}^p$ and $y_i \in \{-1, 1\}$.

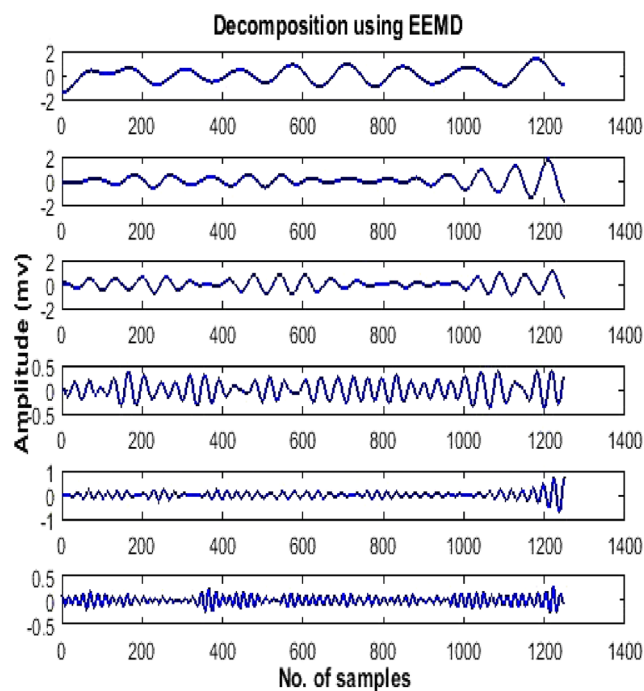
Let a hyper-plane be $\{x : f(x) = x^T \beta + \beta_0 = 0\}$, where β is a unit vector,



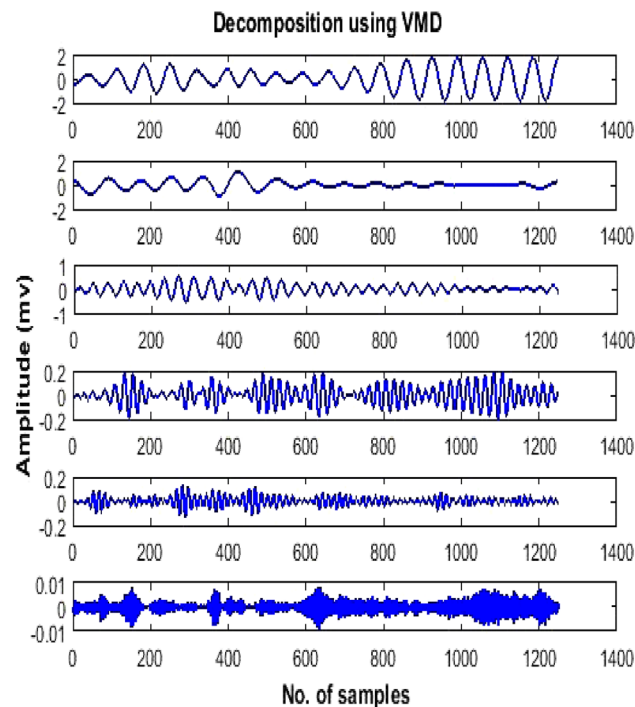
(a) VF Signal



(b) Decomposition using DWT



(c) Decomposition using EEMD



(d) Decomposition using VMD

Fig. 3 For a window length of 5 s, this figure shows **a** VF signal, **b** DWT decomposition, **c** EEMD decomposition, and **d** VMD decomposition of record 'cu07m' of CUDB database as an example of a VF signal

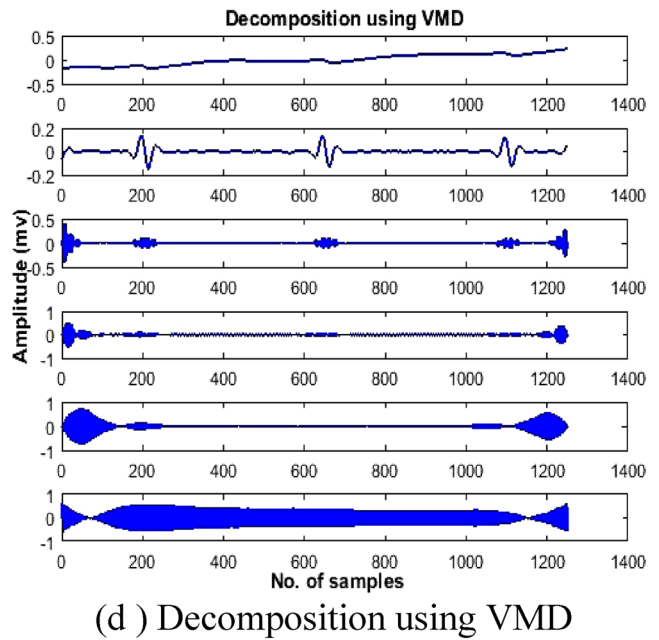
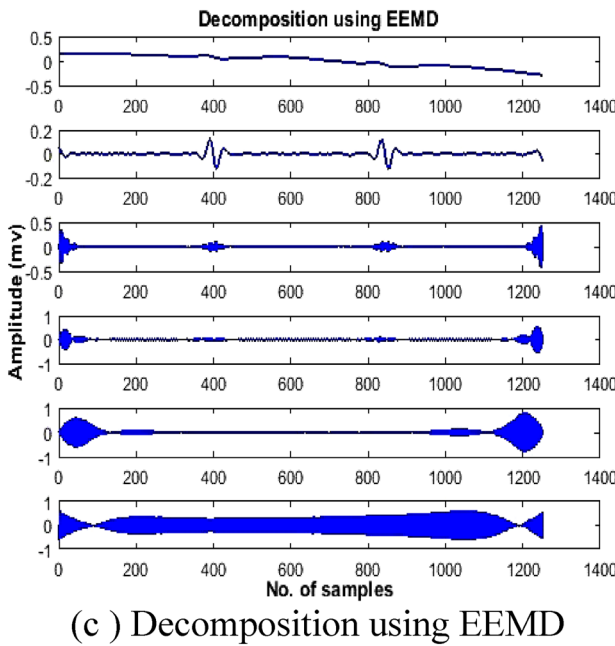
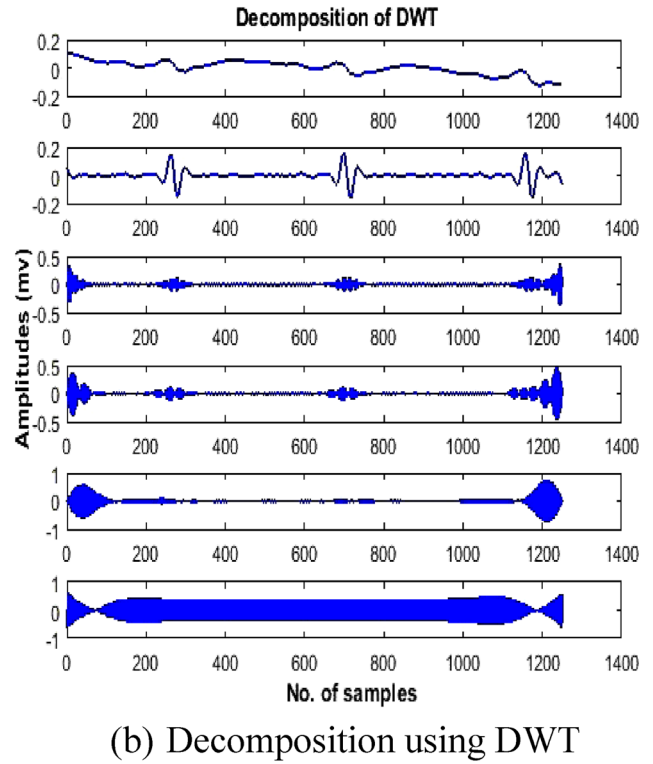
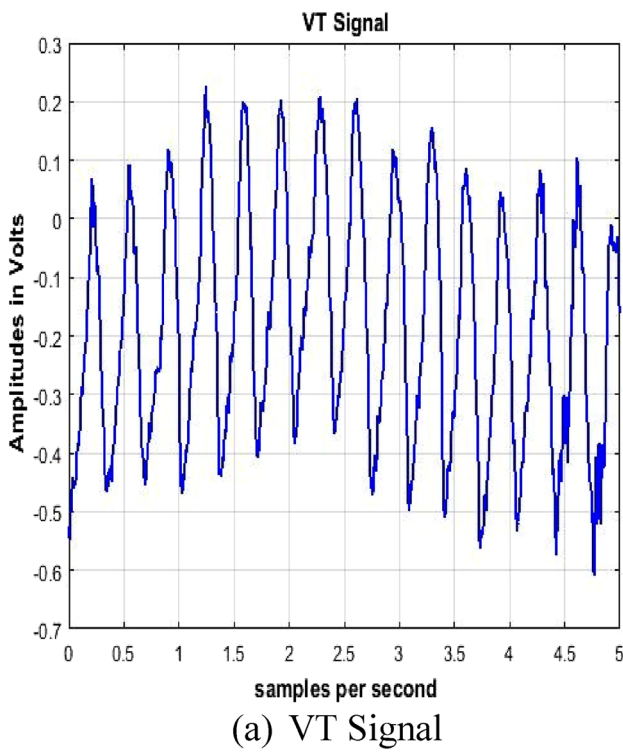


Fig.4 For a window length of 5 s, this figure shows **a** VT signal, **b** DWT decomposition, **c** EEMD decomposition, and **d** VMD decomposition of record ‘418 m’ of VFDB database as an example of a VT signal

$$G(x) = \text{sign}[x^T \beta + \beta_0] C \tag{14}$$

Using WEKA software tool, the tuning parameters of the SVM classifier chosen as:

- (a) Batch size is the preferred number of instances to process may be set to 100.
- (b) C: It is the complexity parameter. This value has been set to 1.0.

Table 1 Rank of features by the correlation attribute evaluation with ranker search method

Features	Weight	Rank	Features	Weight	Rank
LKG	0.41	1	A1	0.12	13
TCSC	0.30	2	LOG_ENT	0.11	14
HURST	0.25	3	FB	0.10	15
CO	0.23	4	PERM	0.09	16
FSMN	0.22	5	A3	0.08	17
C1	0.22	6	KUR	0.08	18
SURE_ENT	0.17	7	C2	0.07	19
TCI	0.14	8	NORM_ENT	0.07	20
VAL	0.13	9	A2	0.05	21
MEA	0.13	10	SKW	0.03	22
CROSS	0.12	11	AB	0.03	23
TH_ENT	0.12	12	SHAN_ENT	0.02	24

Table 2 Reduced feature set with selected first 6 features having higher weightage when considering a threshold weightage value of 0.2

Features	Weight	Rank
LKG	0.408	1
TCSC	0.3027	2
HURST	0.2474	3
CO	0.2287	4
FSMN	0.223	5
C1	0.2221	6

- (c) Epsilon: It is used for round-off error. This value has been set to 1.0E−12.
- (d) Kernel: It indicates which kernel to use. We have chosen polykernel.

Decision Tree (C4.5) Algorithm

C4.5 is an algorithm is a class of decision tree was developed by Ross Quinlan which is an improved version of the ID3 algorithm [32]. It uses the perception of information entropy to construct decision trees the same as ID3. The classification begins at the root node. At each node, it selects the attribute that divides the data samples into two subsets. The attribute having peak normalized information gain is considered for making decisions. At each node, the information gain for an attribute Z, is estimated as:

$$Information\ Gain(I, Z) = Entropy(I) - \sum_{v \in Values(Z)} \left(\frac{|I_v|}{|I|} Entropy(I) \right) \tag{15}$$

where I be the set of instances and |I| be its cardinality, I_v is the subset of I for which attribute Z has value v. The entropy of I is estimated as:

$$Ent(I) = \sum_{i=1}^{numclass} -p \log_2 p_i \tag{16}$$

where p_i is the amount of instances in I that have the ith class value as output attribute.

The tuning parameters of the C4.5 classifier using WEKA software tool have been chosen by batch size of 100, confidence factor of 0.5 and the minimum instances per leaf is of 2.

Performance Measures

The classification parameters such as sensitivity (Se), specificity (Sp), precision (Pr) and accuracy (Acc) are is calculated using standard equations as follows:

$$Se (\%) = \frac{T_{-P}}{T_{-P} + F_{-N}} \times 100$$

$$Sp (\%) = \frac{T_{-N}}{T_{-N} + F_{-P}} \times 100$$

$$Pr (\%) = \frac{T_{-P} + T_{-N}}{T_{-P} + F_{-P}} \times 100$$

$$Acc (\%) = \frac{T_{-P} + T_{-N}}{T_{-P} + F_{-P} + T_{-N} + F_{-N}} \times 100$$

where T_{-P} denotes true positive, F_{-N} denotes false negative, T_{-N} denotes true negative and F_{-P} denotes false positive.

Results

The analysis on ventricular arrhythmias identification and classification was carried out by using all the available ECG signals in CUDB and VFDB databases of PhysioNet repository. The acquired signals are often contaminated with noises and interference; hence a pre-processing stage has been carried out for removal of unwanted artifacts and de-noising of the ECG signals. The de-noised ECG signals are decomposed into different levels using three different decomposition techniques (DWT, EEMD and VMD) and the results are as shown in Figs. 2, 3 and 4 respectively. A

set of 24 features has been extracted from the decomposed signals of all the three methods to make a hybrid feature set with large size data of 22,721 × 24. The hybrid feature set is

Table 3 Results of SVM algorithm with all 24 features at CP = 1

No. of feature	TP	FP	FN	TN	ACC	SE	SP	PPV
1	14,792	7928	7928	37,514	76.74	54.69	80.93	76.05
2	15,207	7513	7513	37,927	77.95	56.89	82.27	44.67
3	14,970	6709	6709	38,731	80.31	62.59	84.09	69.68
4	15,045	6654	6654	38,786	80.48	62.76	84.17	71.98
5	15,121	6575	6575	38,865	80.71	63.13	84.38	72.47
6	15,111	6567	6567	38,873	80.73	63.28	84.44	72.15
7	15,582	5874	5874	39,366	82.69	66.42	86.10	74.88
8	15,626	5416	5416	40,024	84.11	69.70	87.25	75.96
9	15,649	5453	5453	39,987	84.00	69.39	87.16	75.71
10	15,652	5448	5448	39,992	84.01	69.41	87.17	75.76
11	15,666	5431	5431	40,009	84.06	69.49	87.21	75.83
12	15,774	5236	5236	40,204	84.64	70.48	87.81	74.81
13	15,755	5207	5207	40,233	84.72	70.74	87.92	74.70
14	15,690	5100	5100	40,340	85.04	71.61	88.19	75.10
15	15,827	4990	4990	40,450	85.36	71.96	88.44	75.70
16	15,914	4800	4800	40,640	85.92	73.00	88.95	76.50
17	15,821	4780	4780	40,600	85.96	73.32	89.04	76.24
18	15,806	4823	4823	40,617	85.85	73.18	88.97	76.06
19	15,973	4681	4681	40,759	86.26	73.68	89.25	77.02
20	15,963	4665	4665	40,775	86.31	73.82	89.29	77.10
21	15,927	4577	4577	40,863	86.57	74.54	89.55	77.54
22	15,955	4545	4545	40,895	86.66	74.69	89.62	77.78
23	15,993	4540	4540	40,900	86.68	74.60	89.61	77.89
24	15,993	4536	4536	40,904	86.69	74.63	89.62	77.91

Table 4 Performance evaluation of the SVM algorithm with a reduced feature set at CP = 1

Selected features	TP	FP	FN	TN	ACC	SE	SP	PPV
1	14,792	7928	7928	37,514	76.74	54.69	80.93	76.05
2	15,207	7513	7513	37,927	77.95	56.89	82.27	44.67
3	14,970	6709	6709	38,731	80.31	62.59	84.09	69.68
4	15,045	6654	6654	38,786	80.48	62.76	84.17	71.98
5	15,121	6575	6575	38,865	80.71	63.13	84.38	72.47
6	15,111	6567	6567	38,873	80.73	63.28	84.44	72.15

then evaluated with correlation attribute evaluation process along with ranker search method as shown in Table 1. All 24 hybrid features are assigned with a rank according to the weightage of respective features. By taking a threshold value of 0.2, a set of only first 6 features having highest rank has been selected which is as shown in Table 2.

The features set is then classified with SVM and C4.5 classifier for comparative analysis. The best result was obtained with CP = 1 for SVM algorithm for all 24 hybrid

features. The results for the evaluation of different feature combinations are as shown in Table 3, which confirms that the combination of all 24 hybrid features taken together gives the highest accuracy rate of 86.69% with a sensitivity of 74.63%. Taking the reduced feature set, the accuracy rate was found to be 80.73% along with the sensitivity of 63.28% as shown in Table 4. The performance analysis for feature set has been done by C4.5 algorithm with different confidence factors (CF) and the best result was obtained with CF = 0.5. Then all 24 features were evaluated for different combinations using a C4.5 algorithm at CF = 0.5 and the evaluated results are as shown in Table 5. It confirms that the larger the number of features higher the classification accuracy but the computational time also increases. By taking

Table 5 Results of C4.5 algorithm with 24 hybrid features at CF=0.5

No. of features	TP	FP	FN	TN	ACC	SE	SP	PPV
1	16,144	3517	3517	41,923	89.68	81.03	92.09	83.69
2	17,346	535	535	44,905	98.43	97.64	98.84	97.36
3	17,499	364	364	45,076	98.93	98.35	99.20	98.28
4	17,530	301	301	45,139	99.12	98.66	99.35	98.51
5	17,656	156	156	45,284	99.54	99.25	99.65	99.31
6	17,662	131	131	45,309	99.62	99.41	99.71	99.40
7	17,694	81	81	45,359	99.76	99.65	99.82	99.63
8	17,690	76	76	45,364	99.78	99.70	99.84	99.63
9	17,678	95	95	45,345	99.72	99.61	99.79	99.55
10	17,681	91	91	45,349	99.73	99.62	99.80	99.57
11	17,680	94	94	45,346	99.72	99.61	99.79	99.56
12	17,696	73	73	45,367	99.74	99.70	99.84	99.65
13	17,701	72	72	45,368	99.74	99.69	99.84	99.65
14	17,696	74	74	45,366	99.75	99.69	99.84	99.65
15	17,695	77	77	45,363	99.78	99.67	99.83	99.65
16	17,711	57	57	45,383	99.79	99.75	99.88	99.73
17	17,712	57	57	45,383	99.80	99.75	99.88	99.73
18	17,704	66	66	45,374	99.81	99.71	99.86	99.67
19	17,710	57	57	45,383	99.81	99.75	99.87	99.74
20	17,708	63	63	45,377	99.82	99.72	99.86	99.70
21	17,708	61	61	45,379	99.82	99.73	99.87	99.70
22	17,708	59	59	45,381	99.82	99.75	99.87	99.71
23	17,708	61	61	45,379	99.82	99.73	99.87	99.70
24	17,709	58	58	45,382	99.83	99.75	99.87	99.72

Table 6 Results of C4.5 algorithm with a reduced feature set at CF=0.5

Selected features	TP	FP	FN	TN	ACC	SE	SP	PPV
1	16,144	3517	3517	41,923	89.68	81.03	92.09	83.69
2	17,346	535	535	44,905	98.43	97.64	98.84	97.36
3	17,499	364	364	45,076	98.93	98.35	99.20	98.28
4	17,530	301	301	45,139	99.12	98.66	99.35	98.51
5	17,656	156	156	45,284	99.54	99.25	99.65	99.31
6	17,662	131	131	45,309	99.62	99.41	99.71	99.40

Table 7 Comparative analysis between the classifiers using 24 features and 6 features

Classifiers	All 24 features (best results)				Only 06 features (best results)			
	Acc	Se	Comp. time (Sec)	MAE (%)	Acc	Se	Comp. time (Sec)	MAE (%)
SVM	86.69	74.63	85.9	27.48	80.73	63.28	9.35	3.04
C4.5	99.83	99.75	4.82	0.31	99.62	99.41	2.71	0.7

Bold value indicates the best result

all twenty-four features give a high accuracy rate of 99.83% with a sensitivity of 99.75% with a computational time of 4.82 s. In a reduced feature set, the best accuracy rate was

found to be 99.62% along with the sensitivity of 99.41% as shown in Table 6. A comparative analysis has been done for both the classifiers by taking all 24 hybrid features and

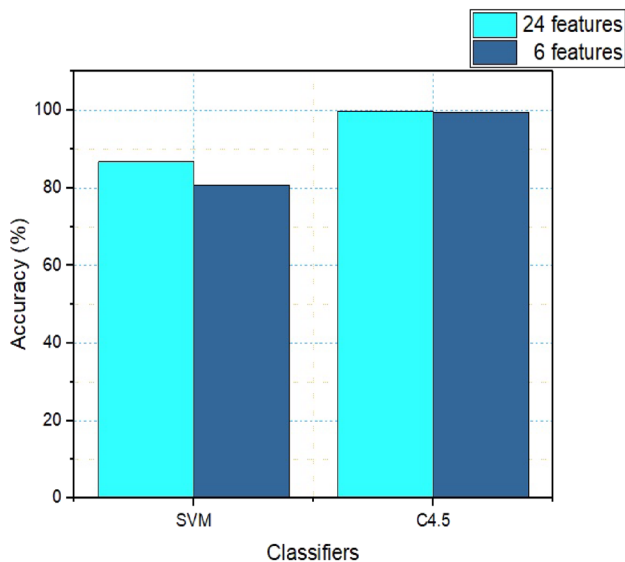


Fig. 5 Plot of accuracy for both classifiers using all 24 and 6 selected features

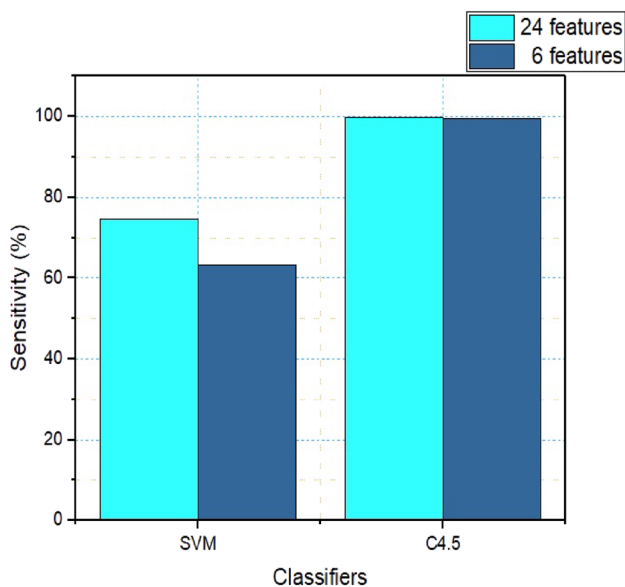


Fig. 6 Plot of sensitivity for both classifiers using all 24 and 6 selected features

selected high ranked 6 features and the results are as shown in Table 7. By considering all 24 features into account, the hybrid feature set was evaluated by both SVM and C4.5 classifiers. The accuracy rate was found to be 99.83% and 86.99% respectively for C4.5 and SVM classifiers. Figures 5 and 6 have been plotted taking accuracy and sensitivity values for both classifiers using all 24 and 6 selected features respectively. Figure 7 has been plotted for comparative analysis of evaluated parameters of C4.5 classifier using all 24 feature and a reduced set of 6 features. The mean absolute

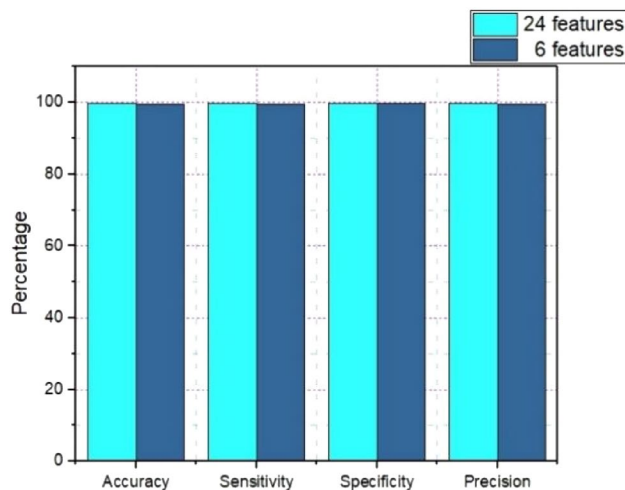


Fig. 7 Comparative analysis of evaluated parameters of C4.5 classifiers using all 24 and 6 selected features

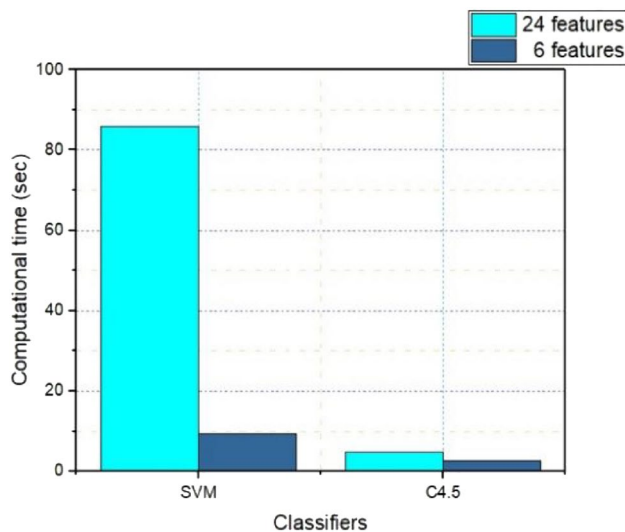


Fig. 8 Plot of computational time for both classifiers using all 24 and 6 selected features

error (MAE) value was found to be 27.48% and 0.31% for SVM and decision tree (C4.5) algorithm respectively. The computational time for both classifiers was also measured and it is found to be 85.9 s and 4.82 s for SVM and C4.5 classifiers respectively for all 24 hybrid features. From the results, it has been observed that the SVM classifier has a very high computational time of 85.9 s for all 24 hybrid features and computational time of 9.35 s for the selected feature set. Figure 8 shows the plot for computational times of both classifiers with all 24 and 6 number of reduced features. Similar observations have also been made that the

Table 8 Performance analysis of VF and VT algorithms

References	Database	Techniques	Results Acc (%)
Ciaccio et al. [33]	MIT-BIH arrhythmia and CU databases	MEDLINE search	93.18
Tripathy et al. [34]	BIH-MIT, CUDB, VFDB	VMD and RF classifier	97.23
Khadra et al. [35]	MITDB and YUDB	raised cosine wavelet transform (RCWT)	97.23
Gauna et al. [36]	MIT-AHA	continuous wavelet transform (CWT)	94.74
Acharya et al. [28]	MIT-BIH arrhythmia and CU databases	Convolutional neural network model	94.03
Balasundaram et al. [29]	MITDB	CWT, and LDA classifier	75
Anas et al. [30]	MIT-BIH arrhythmia database	Mean signal strength and EMD	99.21
Proposed method	CUDB, VFDB	Hybrid features and decision tree classifier	99.62

classification using a reduced feature set gives an accuracy of 99.62% for C4.5 classifier which is lesser than the feature set having all 24 features only by 0.21% with respect to accuracy value, but the computational time was reduced significantly to 2.71 s which earlier was a very high value of 4.82 s in all 24 hybrid features. The computational time is reduced by 43.77% using the selected features. As the computational time is an important parameter in detecting the VAS, therefore the reduced feature sets having important information is a useful approach for diagnosis of heart condition. Therefore, the reduce feature approach is appropriate in detecting VA conditions with faster rate and comparable accuracy.

Discussion

Precise recognition and classification of ventricular arrhythmias are very much crucial for the application of automatic external defibrillation in case of extreme emergency. The main aim of this work was to determine the accuracy to detect VT/ VF using hybrid dataset and a decision tree classifier. Ciaccio et al. [33] have proposed a quantitative method for atrial fibrillation and ventricular tachycardia analyses using MEDLINE search tool. A convolutional neural network approach has been used classification and achieves an accuracy of 93.18% and sensitivity of 91.04% [34]. A raised cosine wavelet transform (RCWT) has been used for the recognition of VAs which is found to be useful for distinguishing between VF, VT, and atrial fibrillation [35]. Precise discrimination of VT and VF rhythms was done with an accuracy of 94.74% [36]. Balasundaram et al. [29] discriminated the VT, Organized VF and disorganized VF with an accuracy of 93.7% using wavelet analysis and binary classifier. Anas et al. [30] suggested a novel method based on EMD that significantly improved the recognition of the life-threatening cardiac arrhythmias. Tripathy et al. [24] used ECG signals to detect the shockable/non-shockable ventricular arrhythmias by using VMD and random forest (RF) classifier and they achieved an accuracy of 97.23%. In

the proposed method, a set of all 24 hybrid features gives the highest accuracy rate of 99.83% with a sensitivity of 99.75% in C4.5 classifier. By taking a threshold value of 0.2, a set of only first 6 high ranked features was evaluated by decision tree classifier. The accuracy rate was of 99.62% along with the sensitivity of 99.41% as described in Table 7. It has been observed that in the reduced feature set the classification accuracy is lesser by only 0.21%, but the computational time has been reduced significantly by 43.77%. Hence, the method with a few selected important features can be a solution in the fastest detection of cardiac arrhythmias. It has been proven that the proposed hybrid method has high detection accuracy with less computational time. Furthermore, this method has significantly low computational complexity and achieves the best performance so far (Table 8).

Conclusions

The automatic arrhythmia classification system is of great significance for the diagnosis of the cardiac diseases. 24 parameters are taken to form a hybrid feature set that includes time–frequency, and statistical domain for classification of normal, VF and VT arrhythmias. The pre-processed ECG signals are decomposed into different levels to extract features using three different transform techniques i.e. wavelet transform, ensemble EMD, and VMD. The proposed system achieves a high classification accuracy of 99.83% in decision tree classifier a short duration (5 s) of ECG data. In the selected high ranked features, the classification accuracy was of 99.62% which is less by only 0.21% with respect to all features. However, this compromise in the accuracy value has been achieved with lesser computational time. The reduced feature set has the computational time about 2.71 s, in case of all 24 parameters based feature was 4.82 s which is reduced by 43.77%. Therefore the proposed method standouts in terms of computational time, as well as in detection success for recognizing the ventricular arrhythmias. The main limitation of this study is the application of deep learning

techniques for prediction of VAs which is the future scope of the work using image datasets [37].

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics statement The paper does not contain any studies with human participants or animals performed by any of the authors.

References

- Dirk H, Bernd P, Hanspeter H, et al. Non-linear coordination of cardiovascular automatic control. *IEEE Eng Med Biol.* 1998;17(6):17–21.
- Leisy PJ, Coeytaux RR, Wagner GS, et al. ECG-based signal analysis technologies for evaluating patients with acute coronary syndrome: a systematic review. *J Electrocardiol.* 2013;46:92–7.
- Brugada J. Relevance of atrial fibrillation classification in clinical practice. *J Cardiovasc Electrophysiol.* 2002;20:S27–30.
- Sun Y, Chan KL, Krishnan SM. Life-threatening ventricular arrhythmia recognition by nonlinear descriptor. *Biomed Eng Online.* 2005;4:6.
- Mathivanan P, Ganesh AB. ECG steganography based on tunable Q-factor wavelet transform and singular value decomposition. *Int J Imaging Syst Technol.* 2020;31(1):270–87.
- Arafat MA, Chowdhury AW, Hasan MK. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *SIViP.* 2011;5:1–10.
- Wasan PS, Uttamchandani M, Moochhala S, et al. Application of statistics and machine learning for risk stratification of heritable cardiac arrhythmias. *Expert Syst Appl.* 2013;40(7):2476–86.
- Miquel Alfaras M, Soriano MC, Ortín S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. *Front Phys.* 2019;7:1–11.
- Sahoo S, Das T, Sabut S. Adaptive thresholding based EMD for delineation of QRS complex in ECG signal analysis. In: *Int. Conf. Wireless Comm. Sig. Proc. and Network.* 2016.
- Maji U, Mitra M, Pal S. Characterization of cardiac arrhythmias by variational mode decomposition technique. *Biocybern Biomed Eng.* 2017;37:578–89.
- Rai HM, Trivedi A, Shukla S. ECG signal processing for abnormalities detection using multi-resolution wavelet transform and artificial neural network classifier. *Measurement.* 2013;46(9):3238–46.
- Rakshit M, Das S. An efficient wavelet-based automated R-peaks detection method using Hilbert transform. *Biocybern Biomed Eng.* 2017;37:566–77.
- Nazarahari M, Namin SG, Hossein A, et al. A multi-wavelet optimization approach using similarity measures for electrocardiogram signal classification. *Biomed Signal Process Control.* 2015;20:142–51.
- Zhang K, Aleexenko V, Jeevaratnam K. Computational approaches for detection of cardiac rhythm abnormalities: are we there yet? *J Electrocardiol.* 2020;59:28–34.
- Martis RJ, Acharya UR, et al. Computer aided diagnosis of atrial arrhythmia using dimensionality reduction methods on transform domain representation. *Biomed Signal Process Control.* 2014;13:295–305.
- Garcia M, Rodenas J, Alcaraz R, Rieta JJ. Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation. *Comput Methods Programs Biomed.* 2016;131:157–68.
- Asgari S, Mehrnia A, Moussavi M. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Comput Biol Med.* 2015;60:132–42.
- Arafat MA, Sieded J, Hasan MK. Detection of ventricular fibrillation using empirical mode decomposition and Bayes decision theory. *Comput Biol Med.* 2009;39:1051–7.
- Kaur L, Singh V. Ventricular fibrillation detection using empirical mode decomposition and approximate entropy. *Int J Emerg Technol Adv Eng.* 2013;3:260–8.
- Xu Y, Wang D, Zhang W, et al. Detection of ventricular tachycardia and fibrillation using adaptive variational mode decomposition and boosted-CART classifier. *Biomed Signal Process Control.* 2018;39:219–29.
- Abdalla FO, Wu L, Ullah H, et al. ECG arrhythmia classification using artificial intelligence and nonlinear and nonstationary decomposition. *Signal Image Video Process.* 2019;13:1283–91.
- Acharya UR, Fujita H, Sudarshan VK, et al. An integrated index for detection of sudden cardiac death using discrete wavelet transform and nonlinear features. *Knowl-Based Syst.* 2015;83:149–58.
- Zhaohua W, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal.* 2009;1:1–41.
- Tripathy RK, Sharma LN, Dandapat S. Detection of shockable ventricular arrhythmia using variational mode decomposition. *J Med Syst.* 2016;40:40–79.
- Rajesh PS, Dhuli R. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine. *Comput Biol Med.* 2017;87:271–84.
- Nguyen MT, Shahzad A, Nguyen BV, Kim K. Diagnosis of shockable rhythms for automated external defibrillators using a reliable support vector machine classifier. *Biomed Signal Process Control.* 2018;44:258–69.
- Maji U, Mitra M, Pal S. Automatic detection of atrial fibrillation using empirical mode decomposition and statistical approach. *Procedia Technol.* 2013;10:45–52.
- Acharya UR, Oh SL, Hagiwara Y, et al. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med.* 2017;89:389–96.
- Balasundaram K, Masse S, Nair K, Umamathy K. A classification scheme for ventricular arrhythmias using wavelets analysis. *Med Biol Eng Comput.* 2013;51:153–64.
- Anas E, Lee S, Hasan M. Sequential algorithm for life-threatening cardiac pathologies detection based on mean signal strength and emd functions. *Biomed Eng Online.* 2010;9(1):43–64.
- Novakovic J. Using information gain attribute evaluation to classify sonar targets. In: *Telecom. forum TELFOR.* 2009. p. 1351–4.
- Mantas CJ, Abellán J. Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Syst Appl.* 2014;41:4625–37.
- Ciaccio EJ, Biviano AB, Iyer V, Garan H. Trends in quantitative methods used for atrial fibrillation and ventricular tachycardia analyses. *Inform Med Unlocked.* 2017;6:12–27.
- Tripathy RK, Zamora-Mendez A, Serna JA, et al. Detection of life threatening ventricular arrhythmia using digital taylor fourier transform. *Front Physiol.* 2018;9:1–12.

35. Khadra L, Al-Fahoum AS, Al-Nashash H. Detection of life-threatening cardiac arrhythmias using the wavelet transformation. *Med Biol Eng Comput.* 1997;35:626–32.
36. de Gauna SR, Lazkano A, Ruiz J, Aramendi E. Discrimination between ventricular tachycardia and ventricular fibrillation using the continuous wavelet transform. *Comput Cardiol.* 2004;31:21–4.
37. Xiao W, Gao Q, Kumar R, et al. Implementation of convolutional neural network categorizers in coronary ischemia detection. *Int J Imaging Syst Technol.* 2021;31:313–26.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.