Biomedical Signal Processing and Control 41 (2018) 242–254

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier

Kandala N.V.P.S. Rajesh*, Ravindra Dhuli

School of Electronics Engineering, VIT University, Vellore 632014, India

ARTICLE INFO

Article history:
Received 10 June 2017
Received in revised form
14 November 2017
Accepted 10 December 2017

Keywords:
ECG signal
Higher order statistics
Class imbalance
Classification

ABSTRACT

Computer-aided heartbeat classification has a significant role in the diagnosis of cardiac dysfunction. Electrocardiogram (ECG) provides vital information about the heartbeats. In this work, we propose a method for classifying five groups of heartbeats recommended by AAMI standard EC57:1998. Considering the nature of ECG signal, we employed a non-stationary and nonlinear decomposition technique termed as improved complete ensemble empirical mode decomposition (ICEEMD). Later, higher order statistics and sample entropy measures are computed from the intrinsic mode functions (IMFs) obtained from ICEEMD on each ECG segment. Furthermore, three data level pre-processing techniques are performed on the extracted feature set, to balance the distribution of heartbeat classes. Finally, these features fed to AdaBoost ensemble classifier for discriminating the heartbeats. Simulation results show that the proposed method provides a better solution to the class imbalance problem in heartbeat classification.

1. Introduction

Cardiovascular diseases (CVDs) are one of the primary causes of the global increase in the fatality rate. According to [1], 30% of the global mortality is due to CVDs. They are notably higher in countries with relatively low-income levels. Approximately half of the cardiovascular deaths are sudden cardiac deaths (SCDs), and cardiac arrhythmias cause most of these. Arrhythmia refers to any disturbance that alters the normal rhythmic functioning of the heart. The chance for SCD is higher in patients having a history of stroke or patients at cardiovascular risk [2]. Therefore, continuous monitoring of heart activity is becoming alarmingly inevitable. Moreover, detection of arrhythmia is essential for proper therapy, to resist the deterioration in heart functioning.

1.1. Aim of the work

Electrocardiogram (ECG) is an inexpensive and noninvasive diagnostic tool used to study the electrical activity of the heart. An ECG is an electrical signal representing the action potentials of various cardiac tissues, derived from the electrodes placed on different parts of the body [3]. A portable ECG recorder called Holter monitor is a very useful tool to analyze the electrical activity of the heart for longer durations. However, investigating various abnormal rhythmic changes from the long ECG record is very exhausting, even for an expertized clinician. Hence computer-aided diagnosis plays a vital role in arrhythmia identification, owing to its effectiveness and robustness. Arrhythmia detection follows the identification of successive heartbeat classes in the given ECG. Therefore, an important step in recognizing arrhythmia is heartbeat classification.

1.2. State-of-the-art

Numerous algorithms have been developed for computer-aided heartbeat classification in the last two decades [4–33]. Feature extraction and classification are the important stages in heartbeat characterization which are widely explored in literature. Features can be extracted directly from the morphology of the ECG signal (time domain methods) or after applying a transformation. R–R intervals, amplitude and duration of the QRS complex are the features which gained major attention in the literature [6–13]. However, these features are sensitive to morphology and dynamics of the ECG. Then, transformation techniques appeared as a solution. The advantage with transformation based feature extraction is it avoids the calculation of fiducial points for heartbeats. This approach will alleviate the problem of exact time alignment of ECG waveforms which is a merit of computer-aided diagnosis [34]. In this category spectral coefficients [14,15], subband coefficients [16] based features are well explored. However, lack of sufficient vari-

* Corresponding author.
E-mail addresses: kandala.rajesh@vit.ac.in (K.N.V.P.S. Rajesh), ravindradhuli@vit.ac.in (R. Dhuli).

https://doi.org/10.1016/j.bspc.2017.12.004
1746-8094/

ations in the spectral features in accordance with the pathological conditions posed a limitation [34].

The nonlinear, non-stationary behavior of the heartbeats is the major reason behind the difficulties faced by the time-domain as well as frequency-domain approaches [34–36]. The non-stationarities arise due to the irregularities in the electrical-conduction formation and traverse. They result in inter-beat inconsistent rhythms and waveform changes in the ECG. Especially these non-stationarities are more in abnormal cardiac cycles. On the other hand, ECG signals are an outcome of nonlinear biological systems. Hence, time-frequency methods gained considerable interest in heartbeat discrimination.

Short-time Fourier transform (STFT) and Wigner–Ville distribution (WVD) are used in [17] for discriminating shockable rhythms from non-shockable cardiac rhythms. Wavelet-based feature extraction schemes are explored extensively in time-frequency methods [4,5,18–23] owing to their efficiency in analyzing non-stationary signals. An important difficulty encountered in wavelets is in selecting mother wavelet and fixing the level of decomposition. Therefore, in recent years adaptive nonlinear decomposition method called empirical mode decomposition (EMD) gathered much attention in heartbeat signal analysis. The main advantage of the EMD is, it decomposes the signal into the AM–FM oscillatory modes from the local characteristics of the signal without any presumptions. The modes derived from the EMD process are complete and partially orthogonal. It is the first adaptive and local type of approach in time-frequency analysis [37]. Some algorithms have exploited EMD for effective feature extraction and arrhythmia recognition [24–28].

In recent years, nonlinear features such as higher order cumulants, entropy measures, higher order spectra and Lempel–Ziv (LZ) complexity are often utilized for heartbeat classification [29–31]. The improved versions of EMD are also used for bio-signal analysis. In [32] EEMD based features are used for discriminating ECG heartbeat signals. We can observe that nonlinear and non-stationary decomposition along with nonlinear features play a crucial role in successful classification.

There lies an additional challenge in dealing with medical data termed as class imbalance. It is due to the limited availability of rare classes results in non-uniform distribution of abnormal data. Here, the results will be biased towards the majority class thus challenging supervised machine learning and increasing the misdiagnosis rate [38]. Works done so far, hardly explored heartbeat classification by considering this limitation.

1.3. Contribution

In this paper, we considered heartbeat classification along with class imbalance. We performed simple statistical tests to understand the nature of ECG signals. From this, we noticed that ECG is a non-stationary and non-Gaussian signal stemming from nonlinear systems. Therefore, we employed a nonlinear feature extraction scheme to classify five groups of ECG heartbeats based on the recommendation of association for the advancement of medical instrumentation (AAMI). We used an adaptive nonlinear decomposition method called improved complete ensemble empirical mode decomposition (ICEEMD) [39]. It is a data-driven method used to analyze non-stationary signals originating from a nonlinear system. It decomposes the signals into approximately orthogonal individual oscillatory modes. Later, we calculated higher order statistics (HOS) and sample entropy measure from these modes to extract the concealed information from the ECG. Next, we performed some data level sampling techniques such as random sampling, synthetic minority oversampling technique (SMOTE) and distribution based balancing to understand the impact of class imbalance on the training level. This feature set is then subjected to an ensemble classifier called AdaBoost [40] which was less explored for heartbeat classification. Furthermore, we employed a feature selection scheme to reduce the number of irrelevant features without decreasing the classification performance.

Rest of the paper is ordered as follows: the ECG data set, experimental setup and theoretical background of the methodology are presented in Section 2. Section 3 discusses about the simulation results and compares with existing works. Future challenges are presented in Section 4. The conclusion of the work is presented in Section 5.

2. Materials and method

A typical pattern classification system includes data collection, pre-processing, feature extraction, feature selection, and classification. Fig. 1 illustrates the block diagram of the proposed methodology. In this section, the theoretical background of the employed techniques is discussed.

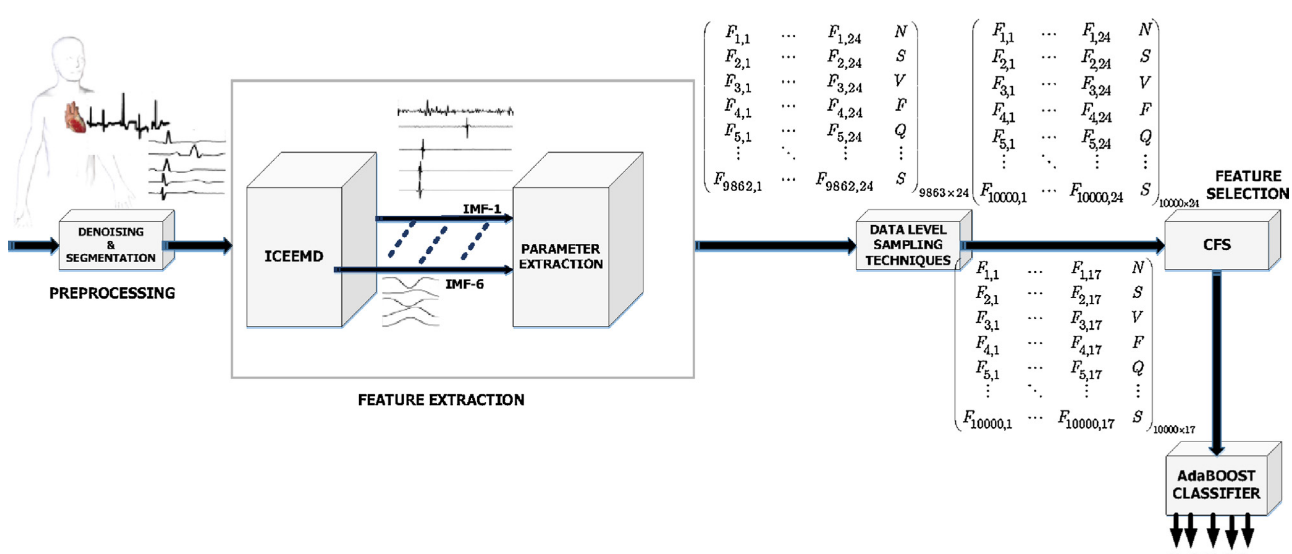

Fig. 1. Block diagram of the proposed methodology.

Table 1
Relation between MIT-BIH heartbeats and AAMI standards.

AAMI classes	MIT-BIH heartbeats
Non-ectopic beats (N)	Normal beats Left bundle branch block beats Right bundle branch block beats Nodal (junctional) escape beats Atrial escape beats
Supra ventricular ectopic beats (S)	Aberrated atrial premature beats Supraventricular premature beats Atrial premature Contraction Nodal (junctional) premature beats
Ventricular ectopic beats (V)	Ventricular flutter wave Ventricular escape beats Premature ventricular contraction
Fusion beats (F)	Fusion of ventricular and normal beat
Unknown beats (Q)	Paced beats Unclassifiable beats Fusion of paced and normal beats

2.1. Data collection

Our proposed method is evaluated on standard baseline database titled, the MIT-BIH arrhythmia database [41]. This database contains 48 Holter recordings from both male and female patients. Each record is of 30 min duration and is sampled at 360 Hz. This database possesses over one lakh heartbeats with fifteen label types. However, AAMI has recommended categorizing these heartbeats into five different groups based on their physiological origin [9] (detailed in Table 1).

2.2. Pre-processing

Pre-processing step involves denoising and segmentation. ECG signals are often mixed with artifacts such as baseline wander (due to muscle movement), power line interference (due to poor contact between electrodes and the body surface, and undesired electrical interference). They degrade the quality of ECG and mask important features that are significant in clinical monitoring. Denoising is a common practice used to enhance the quality of ECG signal by suppressing these artifacts. For this, we employed a filtering routine proposed by Amann [42] with minimal modification. Filtering operation comprises the following:

1. Mean separation from noisy ECG.
2. Moving average filter of order five.
3. High-pass filter with cut-off frequency 1 Hz (for baseline wander suppression).
4. Low-pass Butter worth filter with cut-off frequency 45 Hz (To suppress any left out high-frequency noise).

We make use of the annotation file information for segmenting ECG signals to obtain individual heartbeats. The length of each heartbeat is 830 ms. Table 2 presents the numerical details of the heartbeats used in our work.

Table 3
Test for Gaussianity and linearity.

AAMI class	χ^2	Pfa	R (estimated)	R (theory)
N	3387.5654	0	63.0173	32.2007
S	8359.595	0	322.8324	52.9599
V	3881.732	0	281.625	36.5808
F	4073.6071	0	360.9918	36.2447
Q	1156.847	0	84.225	19.3309

2.3. Feature extraction

Feature extraction is a process of representing a large volume of data with a few samples. These samples are often called features. A generic way of feature extraction is based on the characteristics of data. Proper choice of features will enhance the efficiency of the system. From Fig. 2 it is observed that there are many morphological variations among and within the heartbeat groups. Hence, it is difficult to fix the simple time or frequency domain features as per the state-of-the-art. Therefore, the features that capture the morphology of the signal can reveal the hidden information that lies in the heartbeats. Thus the morphological features of the heterogeneity in the heartbeat groups will aid in the classification. As aforementioned in the Section 1.2 ECG is a nonlinear, non-stationary signal. To support the statement we conduct two statistical tests to analyze ECG signal.

First, Hinich test [43] was carried out to check the linearity and Gaussianity of the data. The basic idea behind the test is, if the third order cumulant of a process is zero, then its bispectrum is zero consequently bicoherence is zero. The process is non-Gaussian if the bispectrum is non-zero. Also, if the process is linear and non-Gaussian, the bicoherence is a constant. Here, the null hypothesis condition is bispectrum of the process is zero, and the alternative hypothesis is bispectrum of the process is non-zero. If the null hypothesis rejects, then we can test for linearity. For linearity test, the null hypothesis is the bicoherence of a process is constant. The bicoherence of a process can be calculated as [43]

$$|\hat{bic}_{xxx}(f_1, f_2)|^2 = \frac{|\hat{bis}_{xxx}(f_1, f_2)|^2}{bis_{2x}(f_1 + f_2) bis_{2x}(f_1) bis_{2x}(f_2)}$$

If $\hat{bis}_{xxx}$ (bispectrum) follows the Gaussian distribution, then $|\hat{bis}_{xxx}|^2$ follows the chi-squared (χ^2) distribution with 2-degrees of freedom. Hinich performed a statistical test to test the consistency of '*bis*' with central χ^2 distribution using the probability of false alarm (Pfa). If the Pfa is close to zero then the assumption of zero bispectrum is rejected there by the non-Gaussianity.

Later, test for linearity of a non-Gaussian process is based on the assumption that the bicoherence needs to be constant for all frequencies f_1 and f_2. The theoretical value of an interquartile range (R) of $|\hat{bic}|^2$ can be estimated and compared with the theoretical R range of χ^2 and non-centrality parameter λ. The linearity condition will be rejected if the difference is more. For further details refer [43]. This test was implemented in a MATLAB routine **glstat** developed by Swami et al. [44]. From Table 3 we can observe that the value of Pfa is equal to zero for all the classes hence we can reject the Gaussianity condition. Also, the difference between $R(estimated)$ and $R(theory)$ is more, then we can reject the linearity condition.

Table 2
Total number of beats, type of beats and train and test samples used in this work.

AAMI class	Number of beats	Training samples (approx. each fold)	Test samples (approx. each fold)
N	3240	324	2916
S	2506	250	2256
V	2298	230	2068
F	467	46	421
Q	1351	135	1216

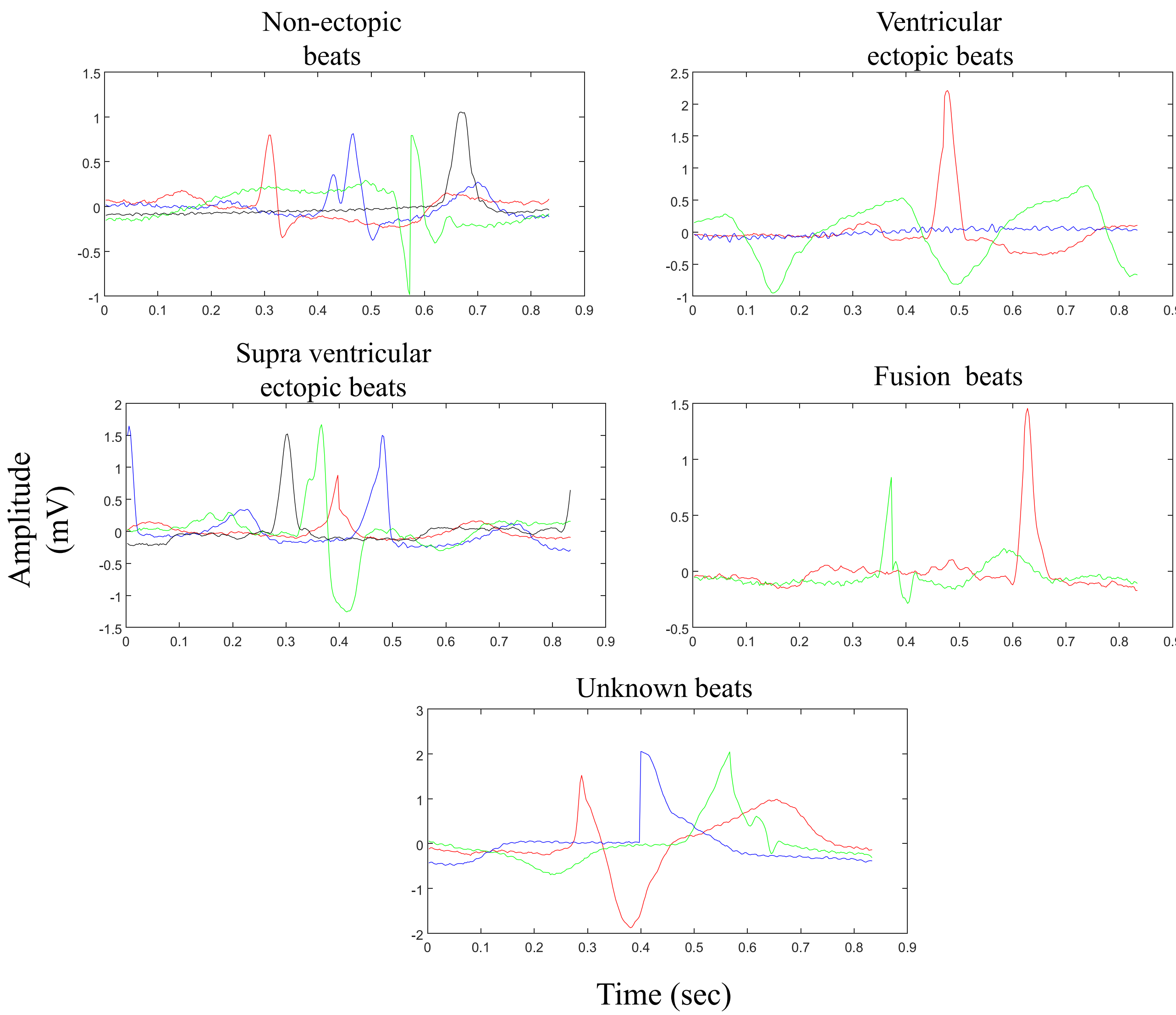

Fig. 2. Samples from AAMI classes.

Table 4
Test for stationarity.

AAMI class	*h*	*p* value
N	1	0.01
S	1	0.01
V	1	0.01
F	1	0.01
Q	1	0.01

Next, we consider KPSS test [45] for checking stationarity of the process based on the criterion of the unit root. It was implemented using by Matlab routine **kpsstest**. Here, a logical variable h is used. If a stochastic process has unit root then h takes a value zero favoring the stationarity hypothesis. In Table 4 the value of h returns 1 for all the classes, that means the alternative hypothesis of unit root is rejected. Hence, the group of heartbeats exhibits the non-stationary behavior. Therefore, we can infer that ECG heartbeat groups are nonlinear, non-Gaussian and non-stationary from these experiments. Therefore, it is worthy to extract the features from the nonlinear decomposition method can divulge the implicit (nonlinear, non-stationary) information from ECG signals.

The proposed feature extraction scheme performs an adaptive nonlinear decomposition method named ICEEMD on ECG segments. It gives oscillatory modes, which reveals subtle information lying in non-stationary signals stemming from nonlinear systems [39]. Later, parameters namely HOS and sample entropy measures are computed from the obtained modes.

2.3.1. Improved complete ensemble empirical mode decomposition (ICEEMD)

Empirical mode decomposition [37] is a data-driven method used to analyze non-stationary signals. It produces fast and slow oscillation modes called intrinsic mode functions (IMFs). The original signal can be reconstructed by summing these IMFs and a monotonic trend. The main advantage of EMD is its capability to segregate stationary and non-stationary components from any complex data [46]. Hence, it gained popularity in medical diagnosis [33,47,48]. However, EMD has a serious problem called “mode mixing,” i.e., information of different modes coexists in a single mode. Various noise assisted algorithms were proposed to alleviate this issue [49,50]. Recently, Colominas et al. [39] proposed a noise-assisted data analysis method namely ICEEMD to address the mode mixing problem. The other advantages of this technique are: less residual noise and minimum reconstruction error in the

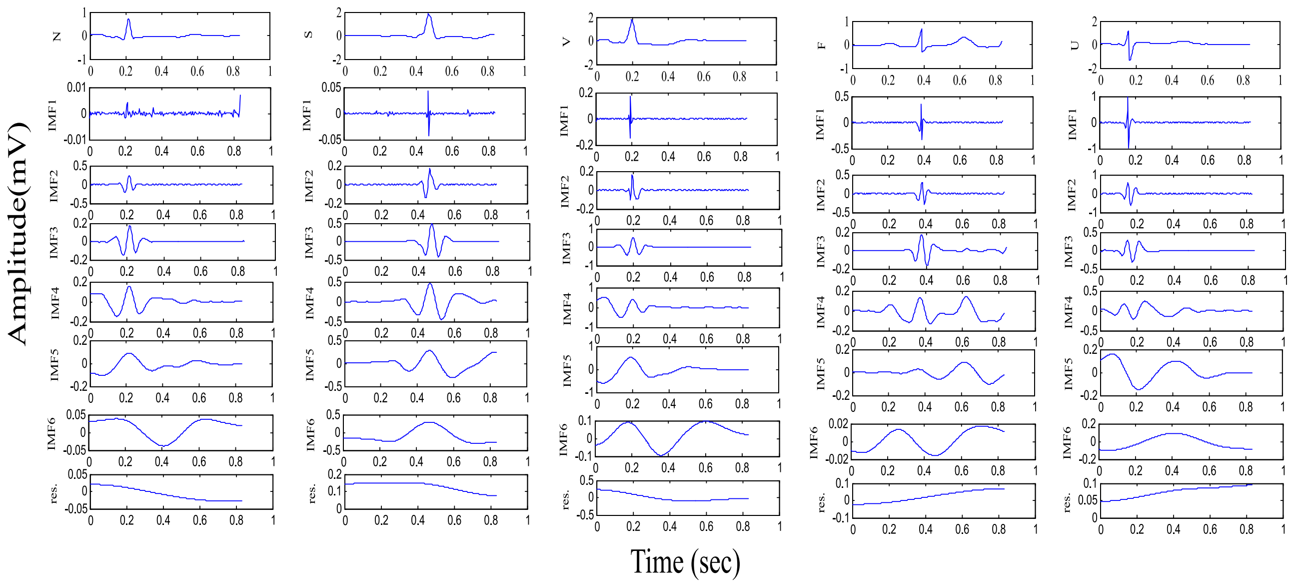

Fig. 3. Decomposition of non-ectopic, supra ventricular, ventricular ectopic, fusion and unknown beats into IMFs using ICEEMD.

modes and ensure completeness of the algorithm. According to [39], this ICEEMD is well suitable for biomedical signal processing. IMF generation using ICEEMD is detailed as follows.

Notation used in algorithm: $E_l(\cdot)$ = lth EMD mode, $M(\cdot)$ = local mean of the signal, $<\cdot>$ = averaging operator, $w^{(j)}$ = realization of white Gaussian noise with zero mean and unit variance and x = input signal.

The algorithm steps:

1. Compute the local means of J realizations $x^{(j)} = x + \beta_0 E_1(w^{(j)}), j = 1, 2, \ldots, J$ using EMD, to obtain first residue $r_1 = \langle M(x^{(j)})\rangle$.
2. At the first stage (l = 1) compute the first IMF:

$$C_1 = x - r_1. \tag{1}$$

3. For l = 2, ..., L calculate r_l as

$$r_l = \langle M(r_{l-1} + \beta_{l-1}E_l(w^{(j)}))\rangle. \tag{2}$$

4. Calculate the lth mode as

$$C_l = r_{l-1} - r_l. \tag{3}$$

5. Go to step 3 for next l

Here $\beta_l = \epsilon_0\sigma_{(r_l)}$ is used to obtain the desired SNR at each stage. We choose ϵ_0 = 0.2. The resultant IMFs are more regular and occupy same scale in the whole time span. It provides a better spectral separation. Each IMF contains useful information regarding individual ECG heartbeat. Fig. 3 depicts the decomposition of heartbeats using ICEEMD. Now few parameters will be extracted from each mode separately.

2.3.2. Higher-order statistics (HOS)

The main idea behind the use of HOS is to derive the information which caused deviation from Gaussianity and linearity. ECG signals are mostly non-Gaussian and present quadratic and higher order nonlinearities [51]. The ability of HOS is to suppress the noise with symmetric probability density functions (pdfs) such as Gaussian, uniform, etc. In this process, ECG signals are preserved because they have non-symmetric pdfs [51]. Second, third and fourth order cumulants are used in our work for feature extraction. Cumulants for a zero mean discrete time random process are defined [44] as:

$$k_{2y}(k_1) = \mathbf{E}[y(n)y(n+k_1)], \tag{4}$$

$$k_{3y}(k_1, k_2) = \mathbf{E}[y(n)y(n+k_1)y(n+k_2)] \tag{5}$$

$$\begin{aligned} k_{4y}(k_1, k_2, k_3) &= \mathbf{E}[y(n)y(n+k_1)y(n+k_2)y(n+k_3)] \\ &-k_{2y}(k_1)k_{2y}(k_2-k_3) \\ &-k_{2y}(k_2)k_{2y}(k_1-k_3) \\ &-k_{2y}(k_3)k_{2y}(k_1-k_2), \end{aligned} \tag{6}$$

where $\mathbf{E}[\cdot]$ is expectation operator and k_1, k_2, k_3 are time lags. In our work cumulants are calculated from selected modes by taking zero lags.

2.3.3. Sample entropy

Sample entropy is a nonlinear measure used to find irregularities of a time series [52]. It is a useful tool for exploring the dynamics of heartbeats [53]. The high value of sample entropy represents more irregularity in time series.

2.4. Class imbalance

Class imbalance refers to non-uniformity in the class distribution. It implies many examples in one class and a few in other classes [54]. Class imbalance problem is an important issue experienced by a large number of domains including bio-informatics [55], text classification [56], medical diagnosis [57], etc. It influences the performance of the standard learning algorithms which presume a balanced distribution of classes [54]. In recent years, research communities are paying proper attention to this problem. There are mainly three approaches for handling class imbalance issues namely: data level methods, algorithm level methods and hybrid methods.

In this work, we study the effect of various data level methods on skewed data with the combination of ensemble classifiers empirically. Three data level methods used in this work:

1. Re-sampling.
2. Synthetic minority oversampling technique (SMOTE).
3. Distribution based data sampling.

2.4.1. Re-sampling

Re-sampling is a widely used method for balancing class distributions. It alters the class distributions by using two well-known strategies called random oversampling and random undersampling [58].

Random oversampling. It increases the number of instances in the minority class via random replications of the instances of the same class.

Random undersampling. It generates random subsample of majority class instances.

2.4.2. SMOTE

It is a popular oversampling method proposed by Chawla in 2002 [59]. The main objective of this approach is to produce "synthetic" minority samples instead of duplicating the samples. The operation of SMOTE is as follows:

- Chose k-nearest neighbors for a selected minority sample (feature vector) based on the requirement of oversampling.
- Take a difference between selected sample and its nearest neighbor of each.
- Multiply the difference vectors by a random number between 0 and 1.
- A new "synthetic" sample will be generated by adding this vector to the selected minority sample.

These synthetic vectors are formed along the line segment joining the selected minority samples and the chosen k-nearest neighbors. This approach generalizes the decision region of the minority class.

In our work, we choose SMOTE along with random undersampling to obtain a more balanced class distribution.

2.4.3. Distribution based data sampling

Imbalanced data sets severely suffer from class oversampling and disjuncts. This problem is referred as "between class imbalance." This issue is addressed by pre-processing through distribution based balancing [60]. The steps involved are as follows:

- Find the prior probability distribution of each feature f_i, $i = 1, 2, \ldots, r$ given class label c_j as $p\left(f_i/c = c_j\right)$.
- Sample new b instances of each class with the prior probability distribution knowledge.

In this way, artificially generated balancing set is formed by sampling equal number of instances in each class.

2.5. AdaBoost ensemble classifier

In recent times, ensemble classifiers are playing a predominant role in machine learning algorithms. Especially, these classifiers are used to address the class imbalance problem in various applications [61,62]. The main aim of ensemble classifier is to reduce the misclassification rate (error rate) of a weak classifier by introducing and aggregating several classifiers. The basic idea is to get predictions of several classifiers on the original data and make it a strong classifier by combining all the various predictions. The foremost strategies in ensemble classifiers are bootstrap aggregating (Bagging) [63] and Boosting [64]. The decision tree based learners are commonly used. In our work, we used adaptive boosting (AdaBoost) ensemble implementation with the random forest as base learner. The AdaBost algorithm was developed by Freund [40]. It is the first practical boosting algorithm. The basic idea of AdaBoost can be written as a weighted combination of the weak learners as follows:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_i h_t(x)\right),$$

where $h_t(x)$ is a weak learner, α_i-weight chosen such that error rate is minimum. Steps given below will give more clarification.

AdaBoost algorithm:

- Given data in the training group along with class labels represented as $D_t : (x_1, y_1), \ldots, (x_n, y_n)$ where $x_i \in \mathbb{R}^n$ and $y_i \in \left\{-1, +1\right\}$.
- Initialize: $w_i^t(i) = \frac{1}{n}, i = 1, 2, \ldots, n$.
- For $t = 1, 2, \ldots, T$.
  - Weak learner classifier: $h_t(x), x \to \left\{-1, +1\right\}$.
  - Find weighted error rate $\epsilon^t = \sum_p w_i$, with constraint $\sum_i w_i = 1$. Here p refers to wrongly classified instances.
  - Chose $\alpha_i = \frac{1}{2}\ln\left(\frac{1-\epsilon^t}{\epsilon^t}\right)$.
  - Update weights for $i = 1, 2, \ldots, n$
    $w_i^{t+1} = \frac{w_i^t}{z_t}\exp(-\alpha_i^t h^t(x_i) y_i)$
    where $z_t = 2\sqrt{\epsilon^t\left(1-\epsilon^t\right)}$, is the normalization factor.
- Final hypothesis:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_i h_t(x)\right) \tag{7}$$

We have used all the tools discussed so far and performed some simulations, to verify the effectiveness of the proposed method.

3. Results and discussion

This section contains a discussion on the simulated results of the proposed methodology illustrated in Fig. 1. Our focus is mainly on class imbalance and its impact on heartbeat classification. We perform an empirical study on, how various data level pre-processing algorithms along with an ensemble classifier are contributing to alleviate the class imbalance problem. The proposed method is evaluated on the MIT-BIH arrhythmia database. In computer-aided diagnosis, feature extraction is of paramount interest. In this work, significant features are extracted from the individual ECG segments using ICEEMD. HOS and sample entropy parameters are the features computed from the selected IMFs. HOS parameters measure the deviation of distribution Gaussianity where as sample entropy measures the irregularity of the bio-medical signals. For understanding the capability of the features we considered the box-whisker plots in Fig. 4, as a matrix of figures. Here, it can be observed that each row corresponds to an IMF, and each column corresponds to feature. For instance, second row fourth column is the box-whisker plot of sample entropy feature of IMF2. In this, X-axis represents the heartbeat type and Y-axis represents the value of the feature. Here, the red color dots are feature samples. The extreme horizontal lines are termed as whiskers, samples beyond this whiskers are called outliers. The horizontal lines constituting the boxes are generated from the median values. In some of the figures boxes are very thin and not visible. For additional details on box-whisker plot refer [65].

In the fourth column by observing the sample entropy values of all IMFs, it is evident that the box and whiskers of all heartbeat groups are almost distinct to each other. Hence sample entropy can help in discriminating heartbeats. Similarly, when we observe plots of the cumulant values, the spread and outlier distributions vary

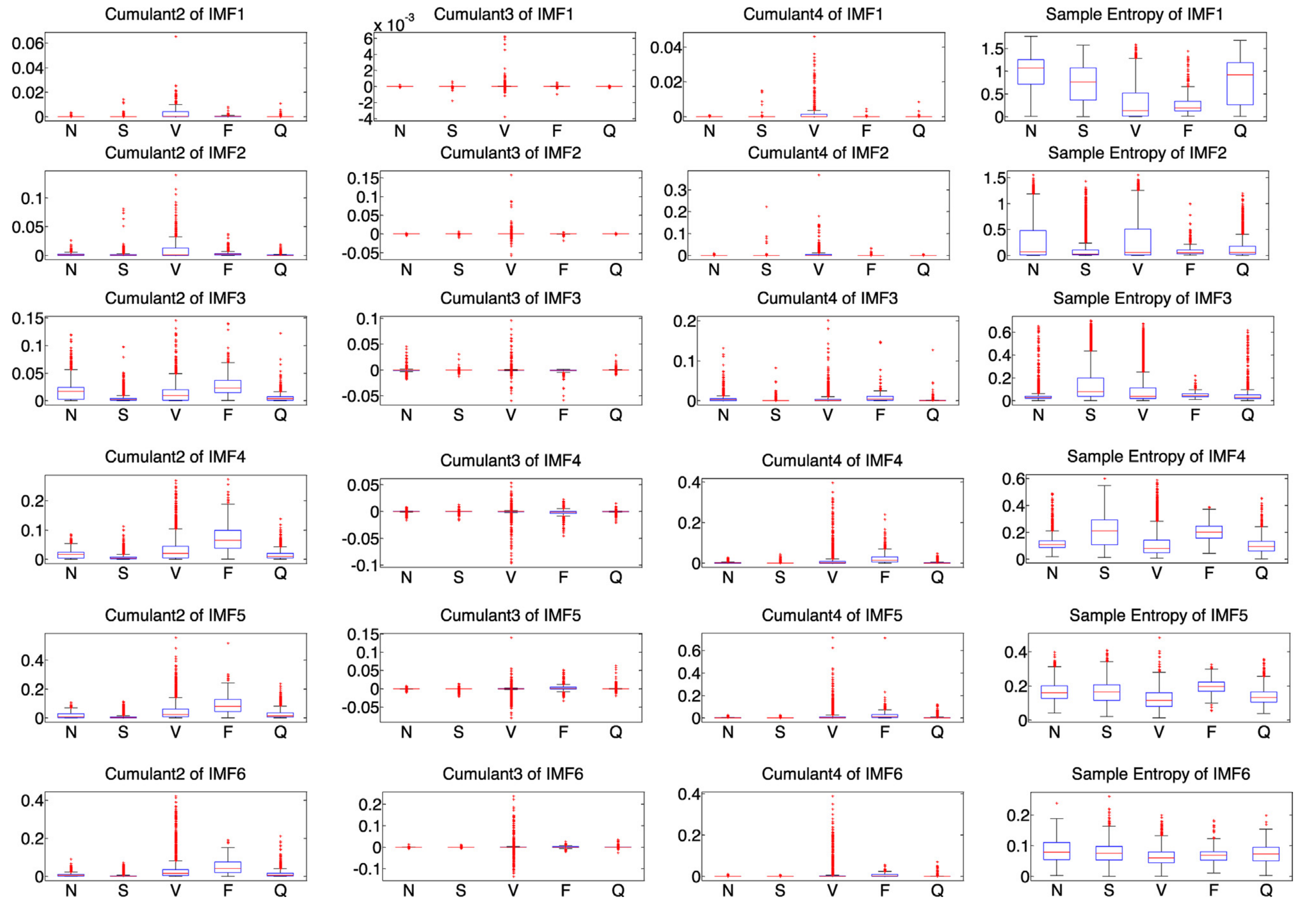

Fig. 4. Box-whisker plots manifest the superior discriminatory capability and efficacy of the features.

Table 5
Confusion matrices for LDA, SVM, *k*-NN, and AdaBoost classifiers.

	LDA					SVM					*k*-NN					AdaBoost				
	N	S	V	F	Q	N	S	V	F	Q	N	S	V	F	Q	N	S	V	F	Q
N	2799	345	31	6	59	2825	271	74	4	66	2768	248	63	8	153	3007	180	33	5	15
S	346	1771	210	7	172	324	1851	181	9	141	305	1820	184	23	174	150	2219	81	9	47
V	243	325	1456	121	153	183	251	1781	38	45	111	252	1808	26	101	26	131	2054	21	66
F	16	22	16	377	36	14	23	71	336	23	15	29	19	377	27	7	17	33	396	14
Q	311	226	95	47	672	313	263	197	13	565	250	177	74	24	826	36	80	83	8	1144

from beat to beat. In cumulant plots the spreads associated with the different heartbeats are distinct to each other. So cumulants can distinguish heartbeats. Thus the entire illustration in Fig. 4 supports our choice of features in heartbeat classification.

To understand the affect of class imbalance, we feed the feature vectors to different classifiers namely support vector machine (SVM) [66], *k*-nearest neighbor (*k*-NN) [67] and linear discriminant analysis (LDA) [68] along with AdaBoost classifier individually without applying data level pre-processing algorithms. These feature vectors are then cross-validated using 10-fold validation. In 10-fold cross-validation, data set is divided into ten subsets. One of the ten subsets is used for testing, and the remaining subsets are used for training. This process is repeated for ten times. The advantage of this method is every sample will be involved in the testing exactly once. Performance of the proposed methodology is analyzed using statistical measures namely sensitivity (SEN), specificity (SPE), receiver operating characteristics (ROC) and accuracy (ACC) [69]. Resulting confusion matrices for the learning algorithms are presented in Table 5. Diagonal values in the confusion matrix represent correctly classified heartbeats. Statistical measures calculated from these matrices are given in Table 6.

From these results, ensemble classifier can be preferred while dealing with imbalanced data sets. In fact, we consider a small amount of non-ectopic beats available in the database. It is to maintain balance with other classes having fewer examples. Still, data is not balanced since fusion, and unknown class labels contain very few examples. Hence, we need to search for an alternative approach.

Data level pre-processing algorithms are one of the good choices to address class imbalance problem in various fields [58]. In this work, we performed three different data level pre-processing techniques on the feature set, and the results are processed through AdaBoost classifier individually. This kind of approach provides more generalization and reduces false alarm rates. Tables 7 and 8 present confusion matrices and performance measures calculated for each class. Random oversampling and undersampling (ROU), re-balance the heartbeat examples and provides uniformly distributed classes, each one having 1972 examples. In the same way, SMOTE + random undersampling (RU) and distribution based balancing (DBB) methods balance all classes by generating new samples. Each class is balanced, with approximately 2000 samples. From Table 8, we can understand that AdaBoost classification on

Table 6
Average statistical performance measures of ten-folds (mean ± standard deviation): LDA, SVM, *k*-NN and AdaBoost.

	LDA				SVM			
	SEN %	SPE %	ROC %	ACC %	SEN %	SPE %	ROC %	ACC %
N	86.4 ± 0.09	86.1 ± 0.08	92.8 ± 0.05	86.2 ± 0.05	87.2 ± 0.06	87.4 ± 0.09	89.5 ± 0.04	87.3 ± 0.06
S	70.5 ± 0.2	87.5 ± 0.18	87.6 ± 0	83.4 ± 0.97	73.9 ± 0.09	89.0 ± 0.04	85.5 ± 0.06	85.2 ± 0.04
V	63.1 ± 0.15	95.3 ± 0.04	90.7 ± 0.03	88.2 ± 1.01	77.5 ± 0.14	93 ± 0.03	90.4 ± 0.07	89.4 ± 0.06
F	80.7 ± 0.3	98.0 ± 0.04	97.6 ± 0.03	97.3 ± 0.34	71.9 ± 0.3	99.3 ± 0.03	93.6 ± 0.12	93.6 ± 0.12
Q	49.9 ± 0.2	95.0 ± 0.17	88.4 ± 0.04	88.9 ± 0.09	41.5 ± 0.33	96.7 ± 0.03	80.3 ± 0.14	89.1 ± 0.06
Average	70.1 ± 0.18	92.4 ± 0.10	91.4 ± 0.03	88.8 ± 0.31	70.4 ± 0.2	93 ± 0.04	87.9 ± 0.08	89.9 ± 0.06
k-NN					AdaBoost			
	SEN %	SPE %	ROC %	ACC %	SEN %	SPE %	ROC %	ACC %
N	85.7 ± 0.21	90 ± 0.2	87.9 ± 0.1	88.6 ± 0.11	92.7 ± 0.15	96.9 ± 0.3	98 ± 0.09	95.5 ± 0.13
S	72.6 ± 0.42	90.4 ± 0.11	81.5 ± 0.2	86.2 ± 0.95	88.7 ± 0.21	94.4 ± 0.12	96.7 ± 0.12	93 ± 0.14
V	78.3 ± 0.2	95.3 ± 0.12	86.9 ± 0.14	91.4 ± 0.14	89.4 ± 0.15	97.0 ± 0.11	97.8 ± 0.07	95.1 ± 0.08
F	81.1 ± 0.47	99 ± 0	90.1 ± 0.2	98.1 ± 0	85.6 ± 0.64	98.4 ± 0.09	97.6 ± 0.16	98.8 ± 0.05
Q	61 ± 0.63	94.8 ± 0.13	77.8 ± 0.35	90.1 ± 0.15	85.5 ± 0.72	98.4 ± 0.09	97.8 ± 0.18	96.6 ± 0.11
Average	75.7 ± 0.38	93.9 ± 0.11	84.8 ± 0.19	90.8 ± 0.27	88.3 ± 0.37	97.0 ± 0.71	97.6 ± 0.12	95.8 ± 0.10

Table 7
Confusion matrices for ROU, SMOTE + RU, DBB sampling techniques with AdaBoost classifier.

	ROU + AdaBoost					SMOTE + RU + AdaBoost					DBB + AdaBoost				
	N	S	V	F	Q	N	S	V	F	Q	N	S	V	F	Q
N	1868	66	19	7	12	1816	132	26	8	17	1974	18	0	0	8
S	49	1841	33	10	39	105	1758	57	22	57	32	1958	0	0	10
V	7	49	1861	17	38	19	112	1756	32	80	1	5	1965	21	18
F	0	2	2	1967	1	5	8	13	1965	8	0	0	3	1992	5
Q	3	22	20	9	1918	22	49	58	15	1855	55	27	0	1	1917

Table 8
Average statistical performance measures of ten-folds (mean ± standard deviation): ROU, SMOTE + RU, DBB sampling techniques with AdaBoost classifier.

	ROU + AdaBoost				SMOTE + RU + AdaBoost			
	SEN %	SPE %	ROC %	ACC %	SEN %	SPE %	ROC %	ACC %
N	94.8 ± 0.19	99 ± 0.10	98.9 ± 0.06	98.2 ± 0.09	91.2 ± 0.39	98.3 ± 0.09	98.3 ± 0.13	96.4 ± 0.35
S	93.4 ± 0.22	98.1 ± 0.11	98.6 ± 0.08	97.1 ± 0.09	88 ± 0.36	96.2 ± 0.15	97.7 ± 0.26	94.7 ± 0.34
V	93.9 ± 0.37	99.0 ± 0.05	98.9 ± 0.09	98 ± 0.08	87.8 ± 0.35	97.8 ± 0.07	97.9 ± 0.12	95.8 ± 0.09
F	99.6 ± 0.08	99.6 ± 0.11	99.8 ± 0.04	99.5 ± 0.04	98.5 ± 0.18	99.0 ± 0.05	99.90.27	98.9 ± 0.18
Q	97 ± 0.32	98.5 ± 0.21	99.5 ± 0.06	98.4 ± 0.08	93.4 ± 0.32	97.8 ± 0.36	98.9 ± 0.18	97.0 ± 0.06
Average	95.7 ± 0.23	98.8 ± 0.11	99.1 ± 0.06	98.2 ± 0.07	91.7 ± 0.32	97.8 ± 0.14	98.5 ± 0.19	96.6 ± 0.2

DBB + AdaBoost				
	SEN %	SPE %	ROC %	ACC %
	98.4 ± 0.22	98.7 ± 0.08	99.7 ± 0.05	98.7 ± 0.08
	97.6 ± 0.28	99.3 ± 0.09	99.6 ± 0.19	99.0 ± 0.12
	98.2 ± 0.25	99.8 ± 0.28	99.9 ± 0.04	99.5 ± 0.07
	99.3 ± 0.2	99.7 ± 0.04	99.9 ± 0.04	99.6 ± 0.06
	95.8 ± 0.4	99.4 ± 0.11	99.6 ± 0.17	98.7 ± 0.1
Average	97.9 ± 0.27	99.4 ± 0.12	99.7 ± 0.09	99.1 ± 0.08

DBB based balanced set gives superior results when compared to other methods. The highest average performance measures with DBB + AdaBoost are 97.9% of SEN, 99.4% of SPE, 99.7% of ROC, and 99.1% of ACC. This approach provides a reliable solution to the class imbalance problem in heartbeat classification in a generalized way.

Another essential criterion to build a more generalized predictive model is feature selection. Feature selection is a process of acquiring an informative subset of features from the original feature set. It is required since redundant and irrelevant features will diminish the performance of a classifier. Moreover, an increase in the number of features leads to curse of dimensionality.

In this work, we used a filter based approach for features selection. Here, significance of a feature is measured with a rank. For this we utilized the minimum redundancy maximum relevance (mRMR) criterion based mutual information developed by Peng et al. [70]. The mRMR method gives ranks to features based on the scores obtained from the statistical dependency between features and the class label. The feature having maximum relevance to the class and minimum redundancy with other features attain high score and top rank. For implementation we used the feature selection toolbox developed by [71]. Table 9 presents scores and the corresponding ranks of the features.

Based on the feature ranking, we carry out classification in an incremental order. Each iteration form a subset of features. Initially, rank one feature is only used for classification. Following that rank 2 feature is added for classification. In this way, we tested all features in the subset. In our work, the minimum subset of size five is required to construct a decision tree. In Fig. 5 number of features in subset against average values of SEN, SPE, and ACC are illustrated. It can be observed that after 17 features (F13, F21, F16, F9, F4, F1, F17,

Table 9
Feature ranking using mRMR.

S. No	Scores	Feature rank	Features
1	0.4998	F13	C2(IMF4)
2	0.4762	F21	C2(IMF6)
3	0.4435	F16	SmpEn(IMF4)
4	0.4399	F9	C2(IMF3)
5	0.4377	F4	SmpEn(IMF1)
6	0.4039	F1	C2(IMF1)
7	0.3904	F17	C2(IMF5)
8	0.3730	F15	C4(IMF4)
9	0.3437	F5	C2(IMF2)
10	0.3227	F11	C4(IMF3)
11	0.3075	F12	SmpEn(IMF3)
12	0.3037	F23	C4(IMF6)
13	0.2650	F2	C3(IMF1)
14	0.2546	F22	C3(IMF6)
15	0.2406	F14	C3(IMF4)
16	0.2404	F3	C4(IMF1)
17	0.2376	F19	C4(IMF5)
18	0.2333	F18	C3(IMF5)
19	0.2173	F10	C3(IMF3)
20	0.2108	F8	SmpEn(IMF2)
21	0.1818	F7	C4(IMF2)
22	0.1756	F20	SmpEn(IMF5)
23	0.0708	F24	SmpEn(IMF6)
24	0.0577	F6	C3(IMF2)

Note: C2–second order cumulant, C3–third order cumulant, C4–fourth order cumulant, SmpEn–sample entropy

F15, F5, F11, F12, F23, F2, F22, F14, F3, F19) there is no significant improvement.

Confusion matrix and statistical performance measures with selected 17 features are shown in Tables 10 and 11. The overall highest average performance measures of the proposed method are 96.5% of SEN, 99.1% of SPE, 99.5% of ROC, and ACC of 98.6%. It justifies that this feature subset along with AdaBoost classifier containing random forest as a base learner yields a preferable solution to the class imbalance problem in heartbeat classification.

In addition to this analysis, we presented the 3D scatter plots using some significant features as axes in Figs. 6–9. Scatter plot using features F13, F21 and F16 is illustrated in Fig. 6. Here, each point in the space represents the feature vector of a heartbeat and the color corresponds to the type. The clusters of N, S, and V groups are prominent than F and Q groups due to minimal number of examples. Similarly, we presented the scatter plot using F21, F16 and F9 in Fig. 7. Here also noticeable clusters of N, S and V groups are formed. The scatter plots after applying DBB resampling are presented in Figs. 8 and 9. Now, we can spot the clusters of minority classes namely F and Q along with N,S, and V. It is implied that all the significant features together give notable differentiable clusters in the feature space. Thus we can justify that these features are capturing the morphological information of the different arrhythmias effectively. Along with the features, AdaBoost ensemble classifier will formulate a meaningful hypothesis for classification. We developed these scatter plots with the help WEKA 3.9 version [72].

Table 10
Confusion matrix of 17 selected features: DBB sampling techniques with AdaBoost classifier.

DBB + AdaBoost	N	S	V	F	Q
N	1945	31	0	0	24
S	57	1923	0	0	20
V	1	4	1933	49	13
F	0	1	12	1979	8
Q	67	44	4	5	1880

Table 11
Average statistical performance measures of ten-folds: AdaBoost (mean ± standard deviation) on selected 17 feature subset.

DBB + AdaBoost	SEN %	SPE %	ROC %	ACC %
N	97.3 ± 0.32	98.3 ± 0.11	99.5 ± 0.05	98.1 ± 0.14
S	96.1 ± 0.36	99 ± 0.09	99.5 ± 0.09	98.4 ± 0.09
V	96.8 ± 0.27	99.7 ± 0.04	99.7 ± 0.06	99.1 ± 0.06
F	98.4 ± 0.36	99.4 ± 0.08	99.8 ± 0.04	99.2 ± 0.07
Q	94.1 ± 0.35	99.1 ± 0.06	99.2 ± 0.12	98.1 ± 0.12
Average	96.5 ± 0.33	99.1 ± 0.07	99.5 ± 0.72	98.6 ± 0.1

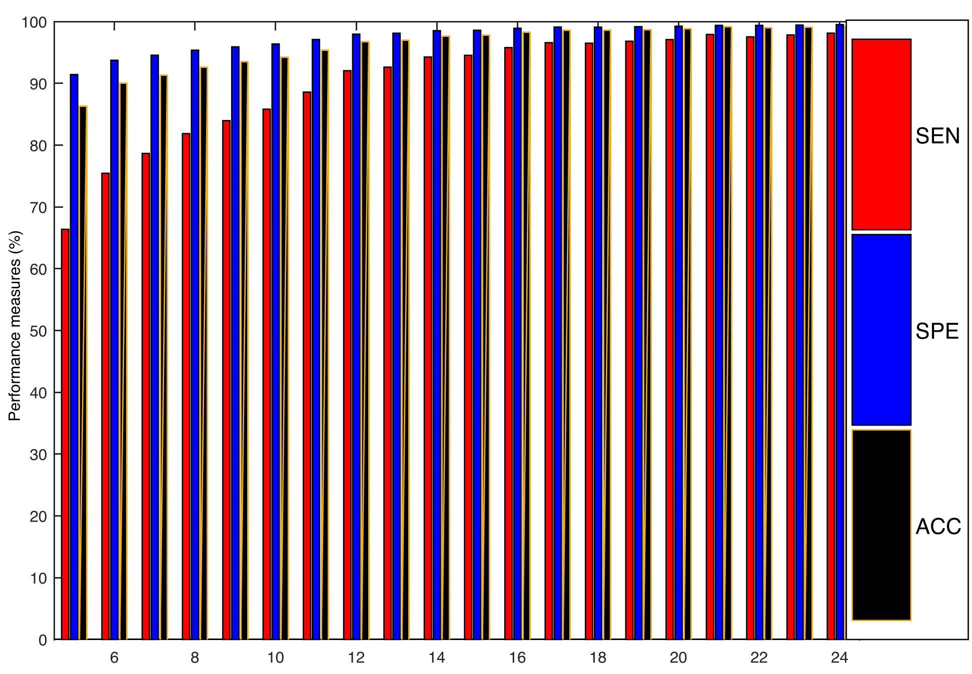

Fig. 5. Performance measure values for different number of features in a subsets selected using mRMR based feature selection.

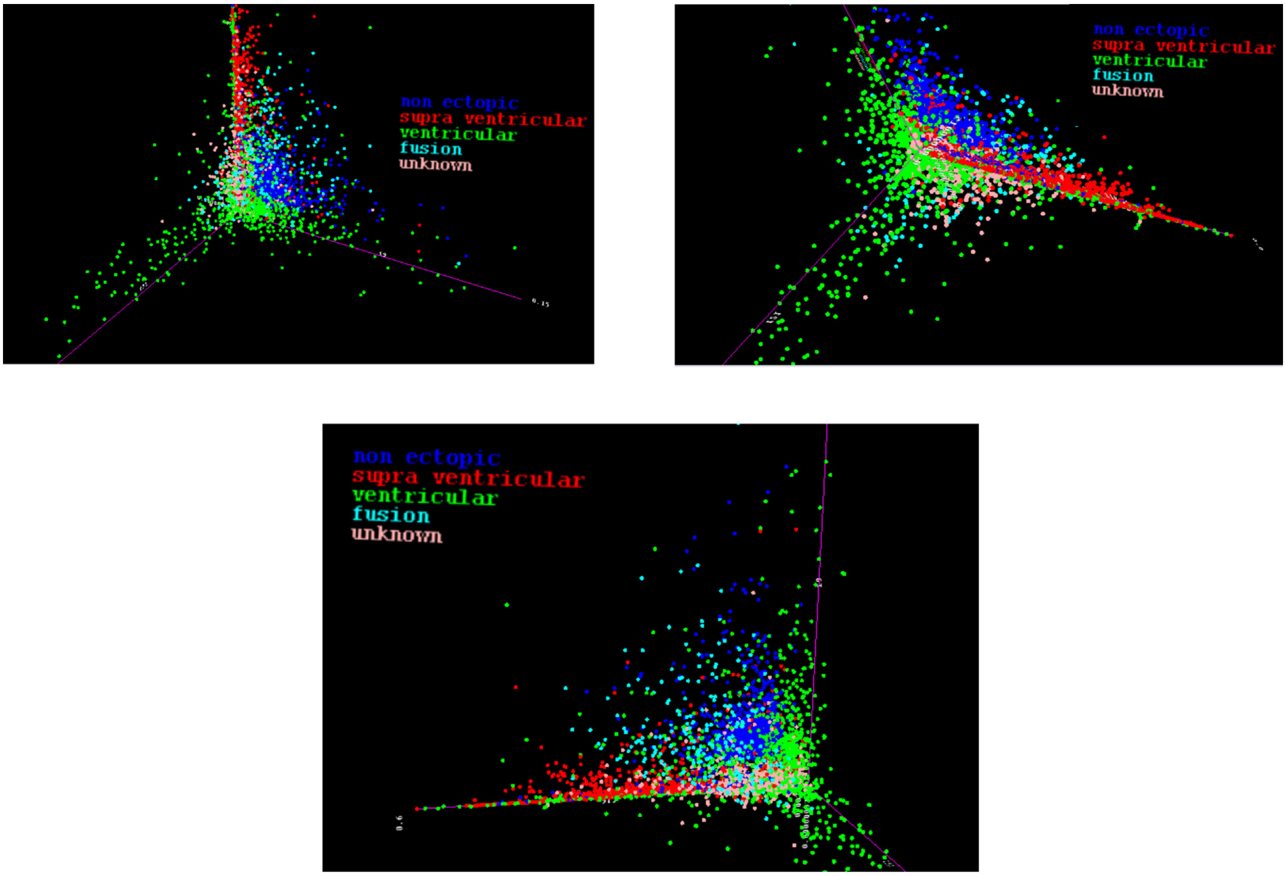

Fig. 6. 3D scatter plots of features F13, F21, F16 in different angles.

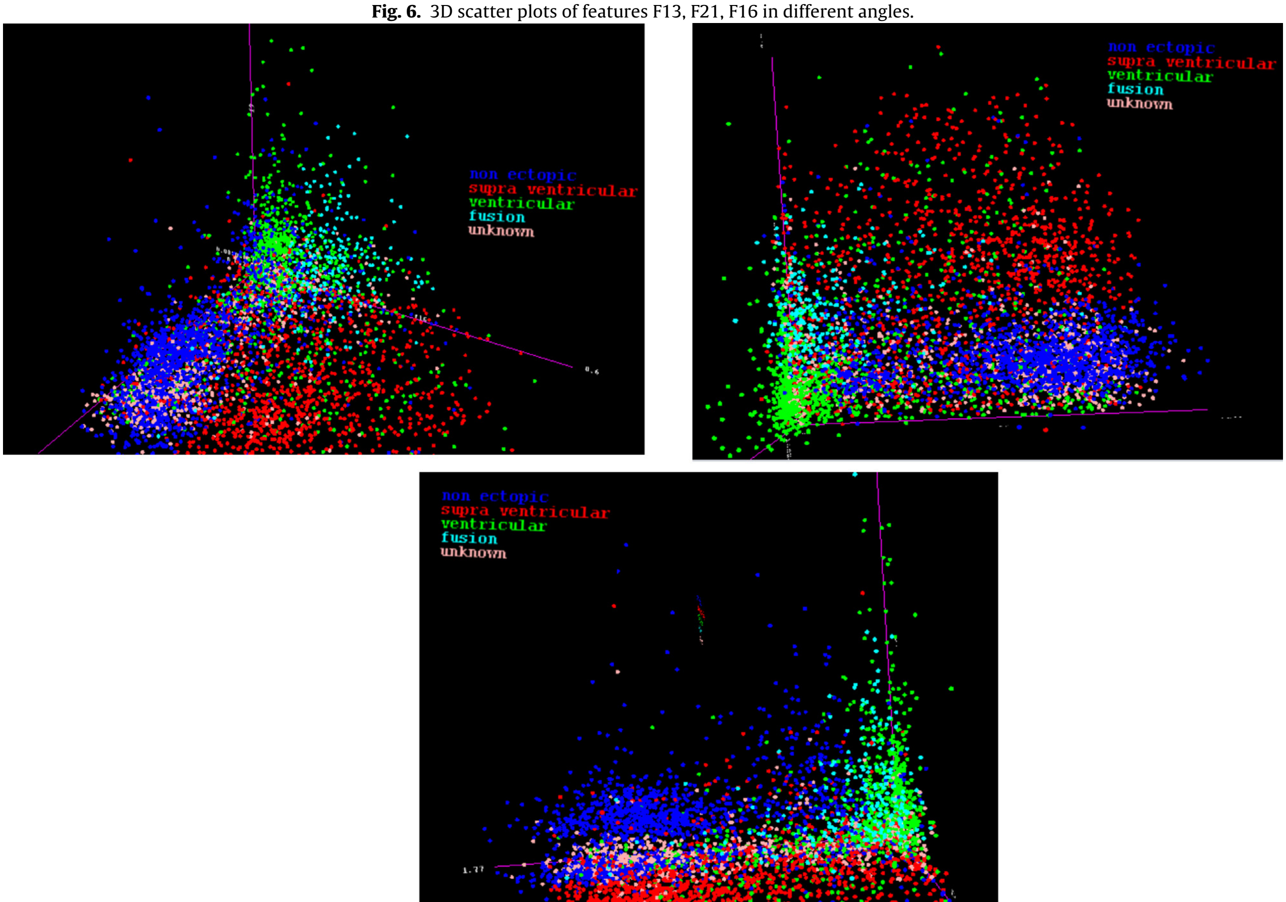

Fig. 7. 3D scatter plots of features F21, F16, F9 in different angles.

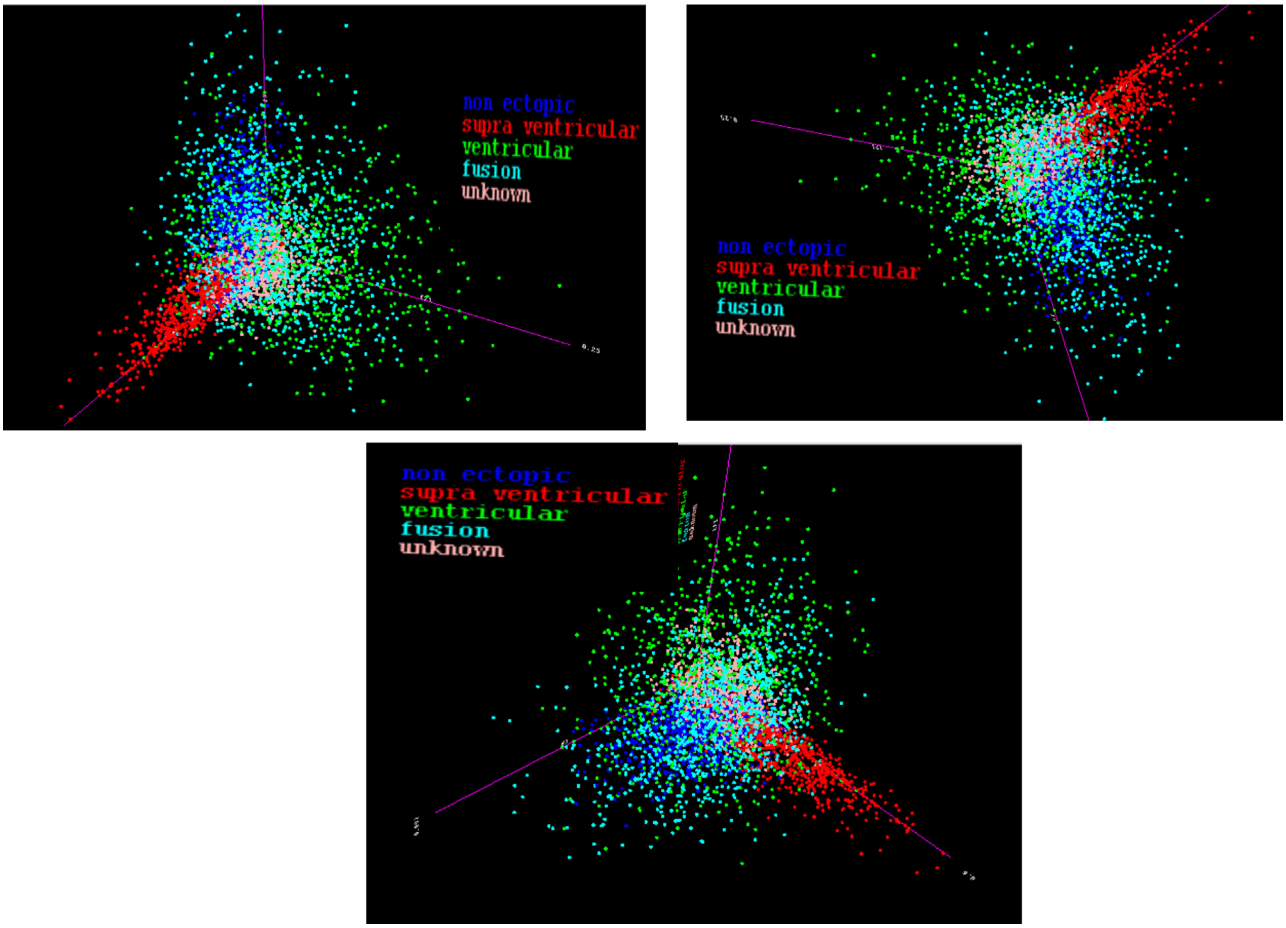

Fig. 8. 3D scatter plots of features F13, F21, F16 in different angles after applying DBB.

Table 12
Performance comparison.

Literature	Features	Classifier	Classes	Performance measures in %
Hu et al. [7]	Time domain features	Mixture of experts (MOE)	4 beat groups	94.0 ACC, 82.6 SEN, 97.1 SPE, 77.7 PPV
Engin [5]	Auto regressive (AR) coefficients + HOS + discrete wavelet transform (DWT) variances	Artificial neural network (ANN)	4 beat types	98.0 ACC
De Chazal et al. [12]	Morphological + heart beat intervals + R–R intervals of ECG	Linear discriminant	5 beat groups	73 SEN, 96 SPE, 45 PPV
Kutlu and Kuntalp [23]	HOS of wavelet packet decomposition (WPD) + morphological features + energy spectral density of DFT	kNN	5 beat groups	85.59 SEN, 99.56 SPE, 95.46 PPV
Kutlu and Kuntalp [22]	HOS of WPD	K-nearest neighbor (KNN)	5 beat groups	90 SEN, 98 SPE, 92 PPV
Ye et al. [18]	(DWT + Independent component analysis (ICA)) + R–R intervals	SVM	3 beat groups	86.4 ACC
Das and Ari [4]	S-transform (ST) + DWT + temporal	Multilayer perceptron (MLP)-NN	5 beat groups	97.5 ACC
Elhaj [19]	(DWT + PCA) + (HOS + ICA)	SVM-RBF	5 beat groups	98.91 ACC, 98.91 SEN, 97.85 SPE
Afkhami et al. [6]	Statistical features + R–R intervals	Gaussian mixture model-expectation maximization (GMM-EM)	3 beat groups	96.15 ACC, 93.3 SEN, 88.99 PPV
In our work (a) (without feature selection)	ICEEMD + (HOS, sample entropy)	AdaBoost	5 beat groups	97.9 SEN, 99.4 SPE, 99.7 ROC, 99.1 ACC
(b) (With feature selection)	ICEEMD + (HOS, sample entropy)	AdaBoost	5 beat groups	96.5 SEN, 99.1 SPE, 99.5 ROC, 98.6 ACC

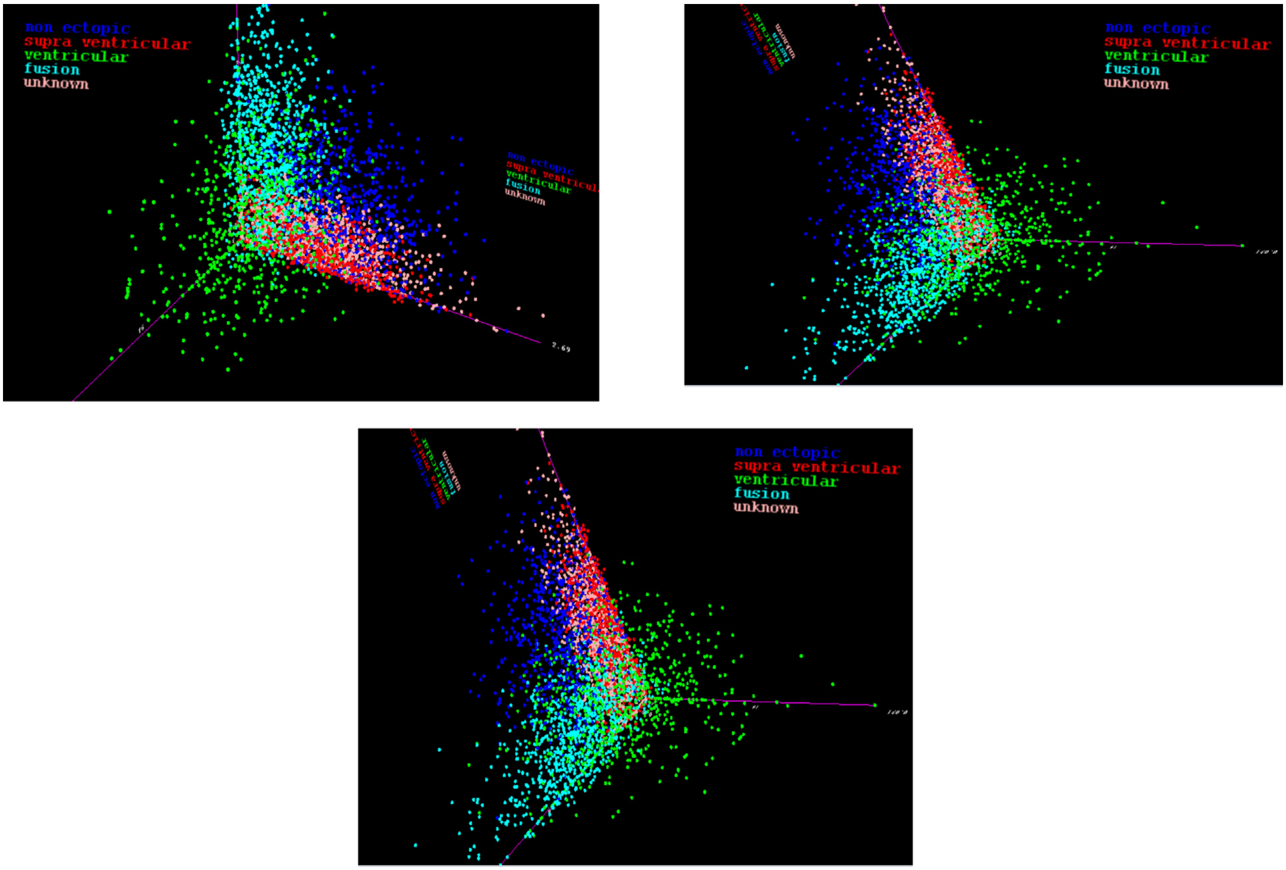

Fig. 9. 3D scatter plots of features F21, F16, F9 in different angles after applying DBB.

To assess the performance of our proposed methodology, we compared our results with the existing methods in the literature. Comparisons are presented in Table 12. From Table 12 it is evident that, the proposed method obtained significant performance measures: 97.9% of SEN, 99.4% of SPE, 99.7% of ROC and ACC of 99.1% without feature selection and 96.5% of SEN, 99.1% of SPE, 99.5% of ROC and ACC of 98.6% with feature selection.

4. Future challenges

The present work demonstrates the ECG beat classification based on the AAMI recommendations. However, still, there are some open issues. One such issue is the intra-patient paradigm, where heartbeats of the same patient are likely to appear in both training and testing data. This situation may lead to biased results. Secondly, fixed beat length segmentation is not always preferable owing to the fact of fast and slow varying heart rhythms. The study of adaptive beat size segmentation is required. Another important issue is, understanding the relationship between underlying physiology and features extracted. There is no precise theory regarding the choice of features to identify the particular disease. Therefore, feature-based disease diagnosis is an interesting research to work on.

5. Conclusion

Computer-aided heartbeat classification studied in this work. Class imbalance has a significant affect on heartbeat classification. In our work, we performed an empirical analysis on how to alleviate class imbalance in heartbeat classification. We have employed a nonlinear data adaptive decomposition method namely ICEEMD to extract features from ECG heartbeats. Later, HOS and sample entropy parameters are calculated from the selected modes obtained from ICEEMD. Next, three data level sampling techniques namely: ROU, SMOTE + RU and DBB are performed to balance the heartbeat classes. Finally, these features are fed to AdaBoost classifier individually for classification. Out of this DBB based balanced data fed to classifier yields superior classification performance. Also, we performed correlation based feature selection to extract significant features contributing an effective classification.

References

[1] S. Mendis, P. Puska, B. Norrving, et al., Global Atlas on Cardiovascular Disease Prevention and Control, World Health Organization, 2011.
[2] R. Mehra, Global public health problem of sudden cardiac death, J. Electrocardiol. 40 (6) (2007) S118–S122.
[3] L. Sörnmo, P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, vol. 8, Academic Press, 2005.
[4] M.K. Das, S. Ari, ECG beats classification using mixture of features, International Scholarly Research Notices, 2014.
[5] M. Engin, ECG beat classification using neuro-fuzzy network, Pattern Recognit. Lett. 25 (15) (2004) 1715–1722.
[6] R.G. Afkhami, G. Azarnia, M.A. Tinati, Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals, Pattern Recognit. Lett. 70 (2016) 45–51.
[7] Y.H. Hu, S. Palreddy, W.J. Tompkins, A patient-adaptable ECG beat classifier using a mixture of experts approach, IEEE Trans. Biomed. Eng. 44 (9) (1997) 891–900.
[8] G. de Lannoy, D. François, J. Delbeke, M. Verleysen, Feature relevance assessment in automatic inter-patient heart beat classification, Biosignals (2010) 13–20.
[9] T. Mar, S. Zaunseder, J.P. Martínez, M. Llamedo, R. Poll, Optimization of ECG classification by means of feature selection, IEEE Trans. Biomed. Eng. 58 (8) (2011) 2168–2177.
[10] P. de Chazal, R.B. Reilly, A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2535–2543.
[11] G. Doquire, G. De Lannoy, D. François, M. Verleysen, Feature selection for interpatient supervised heart beat classification, Comput. Intell. Neurosci. 2011 (2011) 1–10.

[12] P. De Chazal, M. O'Dwyer, R.B. Reilly, Automatic classification of heartbeats using ECG morphology and heartbeat interval features, IEEE Trans. Biomed. Eng. 51 (7) (2004) 1196–1206.
[13] M. Korürek, B. Doğan, ECG beat classification using particle swarm optimization and radial basis function neural network, Expert Syst. Appl. 37 (12) (2010) 7563–7569.
[14] C. Mead, H. Pull, K. Clark, L. Thomas, Expanded frequency-domain ECG waveform processing: integration into a new version of ARGUS/2H, Proceedings of Computers in Cardiology (1982) 205–208.
[15] K.-i. Minami, H. Nakajima, T. Toyoshima, Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network, IEEE Trans. Biomed. Eng. 46 (2) (1999) 179–185.
[16] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, S. Luo, ECG beat detection using filter banks, IEEE Trans. Biomed. Eng. 46 (2) (1999) 192–202.
[17] V.X. Afonso, W.J. Tompkins, Detecting ventricular fibrillation, IEEE Eng. Med. Biol. Mag. 14 (2) (1995) 152–159.
[18] C. Ye, B.V. Kumar, M.T. Coimbra, Heartbeat classification using morphological and dynamic features of ECG signals, IEEE Trans. Biomed. Eng. 59 (10) (2012) 2930–2941.
[19] F.A. Elhaj, N. Salim, A.R. Harris, T.T. Swee, T. Ahmed, Arrhythmia recognition and classification using combined linear and nonlinear features of ECG signals, Comput. Methods Progr. Biomed. 127 (2016) 52–63.
[20] R.J. Martis, U.R. Acharya, K. Mandana, A.K. Ray, C. Chakraborty, Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, Expert Syst. Appl. 39 (14) (2012) 11792–11800.
[21] M. Thomas, M.K. Das, S. Ari, Classification of cardiac arrhythmias based on dual tree complex wavelet transform, in: 2014 International Conference on Communications and Signal Processing (ICCSP), IEEE, 2014, pp. 729–733.
[22] Y. Kutlu, D. Kuntalp, Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients, Comput. Methods Progr. Biomed. 105 (3) (2012) 257–267.
[23] Y. Kutlu, D. Kuntalp, A multi-stage automatic arrhythmia recognition and classification system, Comput. Biol. Med. 41 (1) (2011) 37–45.
[24] M.O.A. Omar, A.S.A. Mohamed, Application of the empirical mode decomposition to ECG and HRV signals for congestive heart failure classification, in: 2011 1st Middle East Conference on Biomedical Engineering (MECBME), IEEE, 2011, pp. 392–395.
[25] U. Desai, R.J. Martis, C. Gurudas Nayak, G. Seshikala, K. Sarika, K.R. Shetty, Decision support system for arrhythmia beats using ECG signals with DCT, DWT and EMD methods: a comparative study, J. Mech. Med. Biol. 16 (01) (2016) 1640012.
[26] E.M.A. Anas, S.Y. Lee, M.K. Hasan, Exploiting correlation of ECG with certain EMD functions for discrimination of ventricular fibrillation, Comput. Biol. Med. 41 (2) (2011) 110–114.
[27] M.A. Arafat, J. Sieed, M.K. Hasan, Detection of ventricular fibrillation using empirical mode decomposition and Bayes decision theory, Comput. Biol. Med. 39 (11) (2009) 1051–1057.
[28] I. Saini, D. Singh, A. Khosla, Electrocardiogram beat classification using empirical mode decomposition and multiclass directed acyclic graph support vector machine, Comput. Electr. Eng. 40 (5) (2014) 1774–1787.
[29] U.R. Acharya, O. Faust, V. Sree, G. Swapna, R.J. Martis, N.A. Kadri, J.S. Suri, Linear and nonlinear analysis of normal and CAD-affected heart rate signals, Comput. Methods Progr. Biomed. 113 (1) (2014) 55–68.
[30] C. Shahnaz, T. Anowar, R. Rafi, I. Ahmmed, S. Fattah, Cardiac beat classification based on wavelet analysis of empirical mode decomposed ECG signals, in: TENCON 2015–2015 IEEE Region 10 Conference, IEEE, 2015, pp. 1–6.
[31] D. Xia, Q. Meng, Y. Chen, Z. Zhang, Classification of ventricular tachycardia and fibrillation based on the Lempel–Ziv complexity and EMD, in: International Conference on Intelligent Computing, Springer, 2014, pp. 322–329.
[32] N. Rajesh, R. Dhuli, Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine, Comput. Biol. Med. (2017).
[33] A.R. Hassan, M.I.H. Bhuiyan, Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Biocybern. Biomed. Eng. 36 (1) (2016) 248–255.
[34] R. Abeysekera, B. Boashash, Time-frequency domain features of ECG signals: their application in P wave detection using the cross Wigner–Ville distribution, in: 1989 International Conference on Acoustics, Speech, and Signal Processing, 1989, ICASSP-89, IEEE, 1989, pp. 1524–1527.
[35] V. Almenar, A. Albiol, A new adaptive scheme for ECG enhancement, Signal Process. 75 (3) (1999) 253–263.
[36] M. Akay, Wavelet applications in medicine, IEEE Spectr. 34 (5) (1997) 50–56.
[37] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. 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, Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci. 454 (1998) 903–995.
[38] B. Krawczyk, Learning from imbalanced data: open challenges and future directions, Prog. Artif. Intell. 5 (4) (2016) 221–232.
[39] M.A. Colominas, G. Schlotthauer, M.E. Torres, Improved complete ensemble EMD: a suitable tool for biomedical signal processing, Biomed. Signal Process. Control 14 (2014) 19–29.
[40] Y. Freund, An adaptive version of the boost by majority algorithm, Mach. Learn. 43 (3) (2001) 293–318.
[41] G.B. Moody, R.G. Mark, The impact of the MIT-BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.
[42] A. Amann, R. Tratnig, K. Unterkofler, Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators, Biomed. Eng. Online 4 (1) (2005) 60.
[43] M.J. Hinich, Testing for Gaussianity and linearity of a stationary time series, J. Time Ser. Anal. 3 (3) (1982) 169–176.
[44] A. Swami, J.M. Mendel, C.L. Nikias, Higher-Order Spectral Analysis Toolbox, Technical Support and Product Enhancement Suggestions, 1998.
[45] D. Kwiatkowski, P.C. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? J. Econom. 54 (1–3) (1992) 159–178.
[46] J. Gilles, Empirical wavelet transform, IEEE Trans. Signal Process. 61 (16) (2013) 3999–4010.
[47] K. Biju, H.A. Hakkim, M. Jibukumar, Ictal EEG classification based on amplitude and frequency contours of IMFs, Biocybern. Biomed. Eng. 37 (1) (2017) 172–183.
[48] R. Djemili, H. Bourouba, M.A. Korba, Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals, Biocybern. Biomed. Eng. 36 (1) (2016) 285–291.
[49] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (01) (2009) 1–41.
[50] M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2011, pp. 4144–4147.
[51] W.A. Zgallai, Second- and Third-Order Statistical Characterization of Non-Linearity and Non-Gaussianity of Adult and Fetal ECG Signals and Noise, INTECH Open Access Publisher, 2013.
[52] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (6) (2000) H2039–H2049.
[53] D.E. Lake, J.S. Richman, M.P. Griffin, J.R. Moorman, Sample entropy analysis of neonatal heart rate variability, Am. J. Physiol. Regul. Integr. Comp. Physiol. 283 (3) (2002) R789–R797.
[54] N. Japkowicz, The class imbalance problem: significance and strategies, Proc. of the Int'l Conf. on Artificial Intelligence (2000).
[55] P. Radivojac, N.V. Chawla, A.K. Dunker, Z. Obradovic, Classification and knowledge discovery in protein databases, J. Biomed. Inf. 37 (4) (2004) 224–239.
[56] Z. Zheng, X. Wu, R. Srihari, Feature selection for text categorization on imbalanced data, ACM SIGKDD Explor. Newsl. 6 (1) (2004) 80–89.
[57] Y.-S. Chen, An empirical study of a hybrid imbalanced-class DT-RST classification procedure to elucidate therapeutic effects in uremia patients, Med. Biol. Eng. Comput. 54 (6) (2016) 983–1001.
[58] R.C. Prati, G.E. Batista, M.C. Monard, A study with class imbalance and random sampling for a decision tree learning system, in: IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, 2008, pp. 131–140.
[59] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[60] P. Bermejo, J.A. Gámez, J.M. Puerta, Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets, Expert Syst. Appl. 38 (3) (2011) 2072–2080.
[61] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) 42 (4) (2012) 463–484.
[62] N.C. Oza, K. Tumer, Classifier ensembles: select real-world applications, Inf. Fusion 9 (1) (2008) 4–20.
[63] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[64] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227.
[65] R. McGill, J.W. Tukey, W.A. Larsen, Variations of box plots, Am. Stat. 32 (1) (1978) 12–16.
[66] V.N. Vapnik, V. Vapnik, Statistical Learning Theory, vol. 1, Wiley, New York, 1998.
[67] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.
[68] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K.-R. Mullers, Fisher discriminant analysis with kernels, in: Neural Networks for Signal Processing IX, Proceedings of the 1999 IEEE Signal Processing Society Workshop 1999, IEEE, 1999, pp. 41–48.
[69] D.M. Powers, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, 2011.
[70] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[71] J. Pohjalainen, O. Räsänen, S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Comput. Speech Lang. 29 (1) (2015) 145–171.
[72] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.