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

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.

© 2017 Elsevier Ltd. All rights reserved.

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).

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 mis-

diagnosis 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.
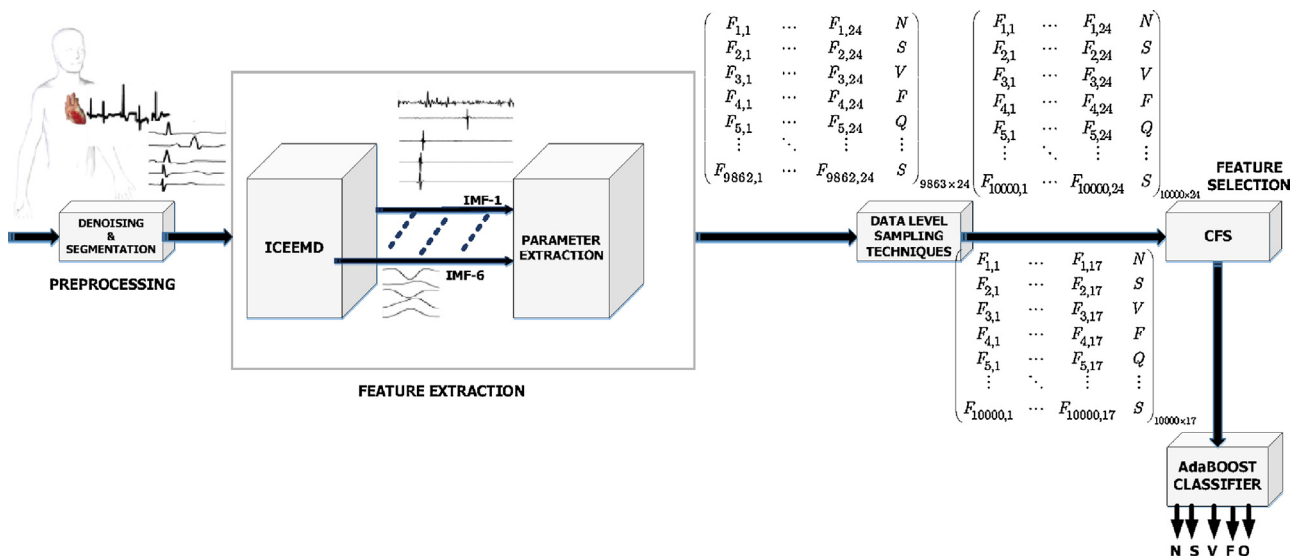

Fig. 1. Block diagram of the proposed methodology.