Biomedical Signal Processing and Control 44 (2018) 200–208

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Efficient classification of ventricular arrhythmias using feature selection and C4.5 classifier

Monalisa Mohanty[a], Santanu Sahoo[a], Pradyut Biswal[b], Sukanta Sabut[c,*]

[a] Dept. of Electronics & Communication Engineering, ITER, SOA University, India
[b] Dept. of Electronics & Telecommunication Engineering, IIIT Bhubaneswar, India
[c] Dept. of Electronics Engineering, Ramrao Adik Institute of Technology, Navi Mumbai, India

ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form 8 March 2018
Accepted 9 April 2018

Keywords:
Sudden cardiac arrest
Ventricular arrhythmias
ECG signals
Ventricular tachycardia (VT)
Ventricular fibrillation (VF)
C4.5 classifier

ABSTRACT

The occurrence of sudden cardiac arrest (SCA) leads to a massive death across the world. Hence the early prediction of ventricular tachycardia (VT) and ventricular fibrillation (VF) becomes vital to prevent from ventricular arrhythmia. In this study, we present a process to detect and classify VT and VF arrhythmias using temporal, spectral, and statistical features. The CU Ventricular Tachyarrhythmia Database (CUDB) and MIT-BIH Malignant Ventricular Ectopy Database (VFDB) databases are used from PhysioNet database for evaluation and comparison of the proposed algorithm. Thirteen time-frequency based features were extracted for a window length of 5 s with an appropriate thresholding to make a feature set. The gain ratio attribute evaluation has been used for potential utilization of the informative features by ranking them according to their individual evaluation weightage. Classification of selected features for VF, VT, and normal sinus rhythm (NSR) is done by using cubic support vector machine (SVM) and the C4.5 classifiers. Assessment of this process is done on 57 records of electrocardiogram (ECG) signals and the result shows that the proposed method achieved a sensitivity of 90.97%, specificity of 97.86% and accuracy of 97.02% in C4.5 classifier, which was better than the obtained results of cubic SVM having an accuracy of 92.23%. This study demonstrated that by using informative features and classifying them with C4.5 algorithms, the system data could be an aid to the clinician for precise detection of ventricular arrhythmias.

1. Introduction

Cardiac tachy-arrhythmias are divided into two most important classes of ventricular tachycardia (VT) and ventricular fibrillation (VF) based on their anatomical origin [1]. VF is the common cause of cardiac arrest in which the electrical activity of the ventricles is disorganized hence the heart quivers rather than pumping. The VT is manifested by fast heart rate which occurs due to the improper electrical activity in the ventricles and dangerous if persist for a long time. It is a potential life-threatening arrhythmia which may result in VF, and finally death [2]. Cardiac arrhythmias diagnosis and classification is a vital task for cardiologists, which could be achieved by automatic diagnosis method in the ECG analysis [3].

Numerous algorithms have been developed to identify and classify the VF and VT arrhythmia conditions from normal sinus rhythm (NSR).The time and frequency domain approach has been used to distinguish ventricular arrhythmias [4–8]. However the time domain approaches are suitable for real-time execution, therefore have an advantage over frequency domain [9,10]. The support vector machines (SVMs), wavelet transforms, neural networks or knowledge-based methods are some of the useful approaches for distinguishing several arrhythmias [11–13]. A combined approach considering frequency and morphological characteristics of ECG signal is an efficient approach in detecting VF and VT conditions [14,15]. Thakur et al. [7] described a time domain method based on sequential probability ratio test (SPRT) and threshold crossing interval (TCI) to discriminate the VF and VT condition. Chen et al. [16] used short-term autocorrelation function with regression test to detect VF and other cardiac arrhythmias. In a modified method, Chen et al. [17] improved the accuracy by reducing the overlap error in the distributions of TCI with a modified SPRT using a new feature called dubbed blanking variability (BV). Zhang et al. [18] utilized complexity measure with a suitable threshold to detect VF and VT by converting the ECG signal into a 0–1 string.

The time-frequency plane has been used successfully for detecting a cardiac arrhythmia in ECG signal analysis [19–22]. In VF condition, the frequency contents are concentrated in the range 2–10 Hz, and in VT the frequency range is characterized by two

* Corresponding author.
E-mail address: sukanta.sabut@rait.ac.in (S. Sabut).

https://doi.org/10.1016/j.bspc.2018.04.005
1746-8094/

distinct bands in the range of 2–5 Hz and 6–8 Hz. Normally the frequency domain approach is used to detect VF and VF conditions by the estimation ECG signal [23]. Some of these methods are based on wavelet transform [21], neural networks [22], and support vector machines (SVMs) [24]. The wavelet analysis approach based upon time-frequency representations is applied successfully in detection of VF and VT classification [25]. In the combined approach that reflects the frequency and morphological characteristics of ECG signal were found to be an effective approach in detecting VF condition [15]. For quantitative analysis of cardiac arrhythmia, the high-order spectral approach has been used for classification of VF and VT [26].The frequency-based filtering based on machine learning techniques are successfully used to recognize for VF condition [27]. The nonlinear features such as approximation entropy and chaotic dimension produced promising results in discriminating VF from sinus rhythm [28,29], however very powerful computers required for quick analysis of signals in clinical practice. Recently, Alickovic et al. [30] proposed an algorithm for heart arrhythmia classification considering ECG beat as an input to the system. They used DWT for decomposing ECG signals into different successive frequency bands to reduce noise and classified with random forest classifier for diagnosing cardiac arrhythmias. Another similar work presented by Wiharto et al. [31] achieved a sensitivity of 74.7% in diagnosing of coronary heart disease by combining the synthetic minority oversampling technique (SMOTE) and the C4.5 classification algorithm.

Numerous methods have been discussed for recognizing cardiac arrhythmias with merits and demerits in implementation, therefore more robust detection methods are needed to classify arrhythmia ventricular arrhythmias efficiently in non-stationary ECG signals. In this study, the main focus was to detect and classify VT and VF arrhythmias using time-frequency (temporal, spectral, and statistical) domain features obtained from processed ECG signals. The signals information was enhanced with filters and different features were extracted for a window length of 5 s with an appropriate thresholding approach. The gain ratio attribute evaluation has been for used for potential utilization of the informative features with ranking approach. The classification of arrhythmia conditions is done with selected features using cubic SVM and C4.5 classifiers. The method has been evaluated with measuring different parameters like sensitivity, accuracy and compared with some of the published results.

2. Methods

The flowchart of the proposed arrhythmia detection process based on ECG signal filtration followed by feature extraction and classification is presented in Fig. 1.

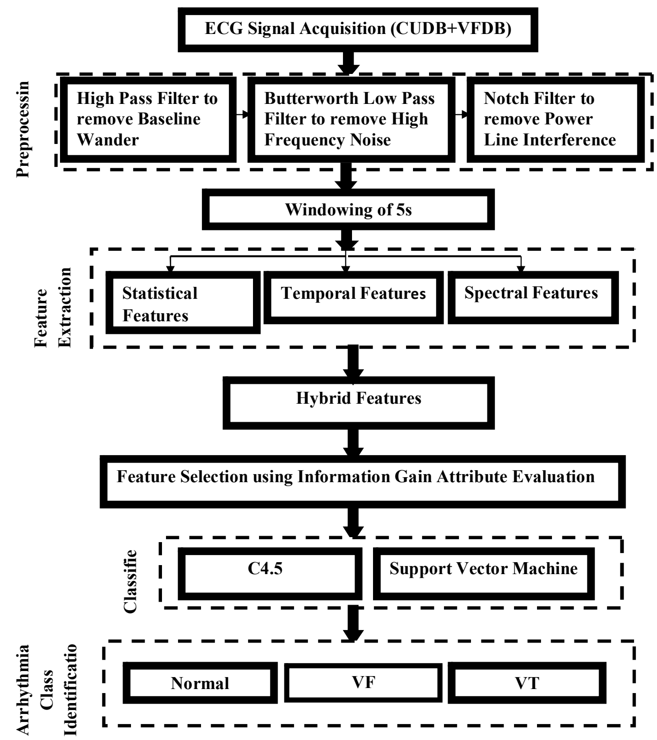

Fig. 1. Flowchart of ventricular arrhythmia detection.

2.1. Database

The VF, VT and normal signal records from CUDB and VFDB databases were used for analysis in this work [32]. The CUDB database includes 35 ECG recordings each of eight-minute duration. All records were digitized with a sampling frequency of 250 Hz with 12-bit resolution over a 10 V range. Every record consists of 127232 samples with duration of 8:28.928 min. These recordings were originally taken by Floyd M. Nolle at the Creighton University Cardiac Center. An appropriate length of 5 s has been chosen for the best window size. With a sampling frequency of 250 Hz, the number of samples taken in one window is 1250. Hence for the entire CUDB database, we have approximately 101 numbers of windows and from each window, a set of 13 features were extracted. The database contains both 'normal beats' and 'ventricular fibrillation beats' but with a large amount of data corresponds to 'normal beats'. A total window of 791 for VF and 2642 windows for normal beats was considered. Two records i.e. cu09 and cu14 consists only Normal beats, hence were not considered for the feature extraction process.

Along with CUDB, another database called MIT-BIH Malignant Ventricular Arrhythmia database was taken to improve the performance of the proposed algorithm to classify the VT, VF and Normal beats. The database contains 22 ECG recordings of different patients who have experienced episodes of sustained ventricular tachycardia, ventricular flutter, and ventricular fibrillation. All records were sampled with a cutoff frequency of 250 Hz and contain 525000 sample intervals with duration of 35 min. For a window length of 5s, the no. of samples in one window was taken as $250 \times 5 = 1250$. The total no. of the window for one record was found to be 420 and hence a total window of 1202 for VT and 7618 windows for Normal beats were found in the entire VFDB database. In our work, a large number of Normal beats of VFDB database were not considered in order to maintain a balance between the number of VT, VF, and Normal beats.

2.2. ECG signal preprocessing by filtration

A 250 Hz sampling frequency of ECG signal was taken for both the databases. The preprocessing of ECG signal is as follows [33]:

(1) To suppress residual baseline wander, a high-pass filter is used with 01 Hz cutoff frequency
(2) To reduce high-frequency noise, a second-order Butterworth low-pass filter is used with 30 Hz cutoff frequency; and
(3) To eliminate power line interference, a notch filter is used.

2.3. Extracted features

In order to classify the ventricular arrhythmia, the following thirteen features were extracted from the processed ECG records.