Biomedical Signal Processing and Control 8 (2013) 437–448

Contents lists available at SciVerse ScienceDirect

Biomedical Signal Processing and Control

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

Technical note

ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform

Roshan Joy Martis [a,*], U. Rajendra Acharya [a,b], Lim Choo Min [a]

[a] Ngee Ann Polytechnic, Singapore
[b] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia

ARTICLE INFO

Article history:
Received 20 August 2012
Received in revised form
12 December 2012
Accepted 18 January 2013
Available online 12 February 2013

Keywords:
Electrocardiogram (ECG)
Discrete Wavelet Transform (DWT)
Association for Advancement of Medical Instrumentation (AAMI)
Principal Component Analysis (PCA)
Linear Discriminant Analysis (LDA)
Independent Component Analysis (ICA)
Support Vector Machine (SVM)

ABSTRACT

Electrocardiogram (ECG) is the P-QRS-T wave, representing the cardiac function. The information concealed in the ECG signal is useful in detecting the disease afflicting the heart. It is very difficult to identify the subtle changes in the ECG in time and frequency domains. The Discrete Wavelet Transform (DWT) can provide good time and frequency resolutions and is able to decipher the hidden complexities in the ECG. In this study, five types of beat classes of arrhythmia as recommended by Association for Advancement of Medical Instrumentation (AAMI) were analyzed namely: non-ectopic beats, supra-ventricular ectopic beats, ventricular ectopic beats, fusion betas and unclassifiable and paced beats. Three dimensionality reduction algorithms; Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) were independently applied on DWT sub bands for dimensionality reduction. These dimensionality reduced features were fed to the Support Vector Machine (SVM), neural network (NN) and probabilistic neural network (PNN) classifiers for automated diagnosis. ICA features in combination with PNN with spread value (σ) of 0.03 performed better than the PCA and LDA. It has yielded an average sensitivity, specificity, positive predictive value (PPV) and accuracy of 99.97%, 99.83%, 99.21% and 99.28% respectively using ten-fold cross validation scheme.

1. Introduction

The incidence and prevalence of cardiovascular diseases (CVD) have increased in recent years [1]. As per the recent WHO report, the overall death rates due to CVD have declined, however the burden of the disease still remains high [2]. In 2007, the overall death rate due to CVD was 251.7 per 100,000 population [2]. The death rate was different for male and female genders and also for black and white races. The rate was 294.0 per 100,000 among white males, 405.9 per 100,000 among black males, 205.7 per 100,000 white females and 286.1 per 100,000 black females [2]. This increase in death rates due to CVD in the modern world is due to the epidemiological transition due to obesity, diabetes mellitus, smoking habit and other lifestyle changes. One of the complication of CVD among many others is atrial and ventricular arrhythmias which occur due to cardiac rhythm disturbances. Arrhythmia is a collective term for a heterogeneous group of conditions in which there would be abnormal electrical activity. There are many causes for arrhythmias, majority of them are related to CVD. Arrhythmias like ventricular fibrillation and flutter are life threatening medical emergencies which result in cardiac arrest, hemodynamic collapse and sudden cardiac death [3]. The electrical activity of the heart could be non-invasively monitored using ECG. It will be very difficult to decipher the hidden information present in the ECG data due to its small amplitude and duration. Therefore a computer assisted tool could help the physicians in their diagnosis. This computer assisted diagnosis can be used as an adjunct tool for the physicians in their practice and interpretation and can play a major role in the management of cardiovascular diseases [4].

Using local fractal dimension, six types of ECG beats (normal, left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature contraction (APC), ventricular premature contraction (VPC) and paced beats) were classified with more than 97% of sensitivity for normal, LBBB, RBBB, paced and VPC beats; and more than 86% sensitivity for APC beats [5]. Five types of ECG beats (normal, LBBB, RBBB, APC and VPC) were classified with 93.97% of accuracy [6]. Wavelet transform and particle swarm optimization technique were used to classify six types of ECG beats (normal, LBBB, RBBB, APC, VPC and paced beats) and obtained 88.84% of accuracy [7].

A minicomputer system was designed for analyzing 24 h ambulatory ECG and automatically classified normal, supra-ventricular ectopic beats and ventricular ectopic beats using time intervals between different characteristic peaks of ECG [8]. Using linear prediction method, the VPC was detected with a sensitivity of 92% [9]. Using Hidden Markov Model (HMM), various segments of ECG were modeled and subsequently classified the ECG beats into normal, supra-ventricular ectopic beats (SVEB) and ventricular

* Corresponding author. Tel.: +65 6460 8393.
E-mail address: roshaniitsmst@gmail.com (R.J. Martis).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.01.005

Table 1
MIT-BIH arrhythmia database beats classified as per ANSI/AAMI EC57:1998 standard [17].

Non ectopic beat (*N*)	Supra-ventricular ectopic beats (*S*)	Ventricular ectopic beats (*V*)	Fusion beat (*F*)	Unknown beat (*Q*)
Normal beat	Atrial Premature (AP) beat	Premature ventricular contraction (PVC) beat	Fusion of ventricular and normal (fVN) beat	Paced beat
Left bundle branch block (LBBB) beat	aberrated Atrial Premature (aAP)	Ventricular escape (VE) beat		Fusion of paced and normal (fPN) beat
Right bundle branch block (RBBB) beat	Nodal (junctional) Premature (NP) beat			Unclassifiable (U) beat
Atrial escape (AE) beat	Supra-ventricular premature beat (SP) beat			
Nodal (junctional) escape (NE) beat				

ectopic beats (VEB) using the time domain features [10]. The ECG morphology and RR interval features were used for the classification of normal and five types of arrhythmia beats using particle swarm optimization and obtained average accuracy of 93.27% [11]. Using auto-regressive model and generalized linear model classifier, normal, APC, supra ventricular tachycardia (SVT), ventricular tachycardia (VT), VPC and ventricular flutter (VF) ECG beats were classified with 93.2% of accuracy [12]. Using Hermite coefficients of ECG beats and neuro-fuzzy technique, the ECG beats were classified with 96% of accuracy [13]. A hardware system was developed with four signals (ballistocardiogram, ECG, lower body impedance plethysmogram and lower body electromyogram) for monitoring the cardiovascular health at home [14]. Two kinds of abnormalities in the ECG were classified using Gaussian Mixture Model (GMM) and reported more than 94% accuracy [15]. Using wavelet transform and PCA, normal and abnormal beats in ECG were classified using different classifiers and reported 95.6% of accuracy with SVM classifier [16]. However all these methods have following drawbacks.

(i) All these methods use time domain features where subtle changes in the ECG signal and the hidden complexities cannot be clearly deciphered.
(ii) Most of these methods were tested only on limited data sets and the generalization performance of these methods on large databases was not tested.
(iii) All these methods were tested only on few classes of ECG beats and there is a need to test the methods and algorithms on a standard classification scheme of arrhythmia beats such as ANSI/AAMI EC57:1998 [17].

The subtle changes in the amplitude and duration of ECG represent the function of the heart [21,23]. In our previous work, we have computed the principal components of time domain ECG signal and subjected them for statistical Student's t-test. Also in parallel the principal components of DWT coefficients of ECG beats were computed and subjected to Student's t-test. It was seen from our experiments that the later method provided higher value for the t-statistic, which implies that the features were more discriminating in DWT domain than time domain for normal and arrhythmia classes. So the subtle changes in the amplitude and duration in the ECG as it was in the time domain did not provide good discrimination, hence the DWT transform domain features were used. There are a large number of coefficients, and hence a dimensionality reduction algorithm needs to be applied. In our proposed method the ECG beats were transformed using DWT and subsequently three dimensionality reduction methods were applied independently to extract the features. Three dimensionality reduction methods were used: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Components Analysis (ICA). These features with reduced dimensionality were fed to neural network (NN), SVM with different kernel functions and probabilistic neural network (PNN) for automated classification.

The paper is organized as follows: Section 2 presents the materials used, Section 3 deals with the methods and classifiers used, Section 4 provides experimental results and the results are discussed in Section 5. Finally the paper concludes in Section 6.

2. Materials used

In this study, the MIT-BIH arrhythmia database [18] is used, where the signals were sampled at 360 Hz. The database consists of 48 signals, each of half an hour duration of Holter recording. In this analysis, we have used the *entire data of MIT-BIH arrhythmia database* as recommended by ANSI/AAMI EC57:1998 standard [17]. The data used consisted of 90, 580 non-ectopic beats, 2973 supra-ventricular ectopic beats, 7707 ventricular ectopic beats, 1784 fusion beats and 7050 unknown and paced beats. In each of the 48 signal files, if insufficient samples were present to the left of the first detected QRS complex, the corresponding beat is neglected. Also if there were insufficient samples after the last detected QRS complex in any of the signal files, then also the corresponding beat is neglected. Table 1 shows the details of the different beats of MIT-BIH database grouped into five main classes.

3. Methodology

Fig. 1 shows the proposed automated system for classification of 5 types of beats in ECG of arrhythmia as recommended by ANSI/AAMI EC57:1998 standard [17]. The working of each block is explained in detail in the following sections.

3.1. Discrete Wavelet Transform (DWT) and denoising of ECG

Unlike Fourier transform, the wavelet transform offers resolution in both time and frequency domains. The DWT is obtained from the continuous wavelet transform by sampling it on a dyadic grid. The DWT decomposes a signal successively into low frequency and high frequency components. The low frequency component is called approximation, and the high frequency component is called the detail. Fig. 2 depicts the DWT decomposition using filter banks. The ECG signal $x(n)$ is passed through a low pass filter $h(n)$, and then down sampled by a factor of two to obtain the approximation coefficients at level one. The high pass filter is derived from the low pass filter as,

$$g(L-1-n) = (-1)^n h(n) \tag{1}$$

where L is the length of the filter in number of points. The detail coefficients were obtained by passing the signal through $g(n)$ and then down sampling by a factor of two. The two filters $h(n)$ and $g(n)$ were called quadrature mirror filters. The DWT filtering along with sub sampling were given by,

$$y_{\text{low}}(k) = \sum_n x(n)h(-n+2k) \tag{2}$$