Automated diagnosis of Coronary Artery Disease affected patients using LDA, PCA, ICA and Discrete Wavelet Transform

Donna Giri [a], U. Rajendra Acharya [b,c,*], Roshan Joy Martis [b], S. Vinitha Sree [d], Teik-Cheng Lim [a], Thajudin Ahamed VI [e], Jasjit S. Suri [f,1]

[a] SIM University, School of Science and Technology, Singapore 599491, Singapore
[b] Department of Electronics & Communication Engineering, Ngee Ann Polytechnic, Singapore 599489, Singapore
[c] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia
[d] Global Biomedical Technologies, Inc., Roseville, CA 95661, USA
[e] Department of Electronics & Communication Engineering, Government Engineering College, Wayanad, Kerala 670 644, India
[f] Department of Biomedical Engineering, Idaho State University (Aff.), Idaho, USA

ARTICLE INFO

ABSTRACT

Coronary Artery Disease (CAD) is the narrowing of the blood vessels that supply blood and oxygen to the heart. Electrocardiogram (ECG) is an important cardiac signal representing the sum total of millions of cardiac cell depolarization potentials. It contains important insights into the state of health and nature of the disease afflicting the heart. However, it is very difficult to perceive the subtle changes in ECG signals which indicate a particular type of cardiac abnormality. Hence, we have used the heart rate signals from the ECG for the diagnosis of cardiac health. In this work, we propose a methodology for the automatic detection of *normal* and *Coronary Artery Disease* conditions using heart rate signals. The heart rate signals are decomposed into frequency sub-bands using Discrete Wavelet Transform (DWT). Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) were applied on the set of DWT coefficients extracted from particular sub-bands in order to reduce the data dimension. The selected sets of features were fed into four different classifiers: Support Vector Machine (SVM), Gaussian Mixture Model (GMM), Probabilistic Neural Network (PNN) and K-Nearest Neighbor (KNN). Our results showed that the ICA coupled with GMM classifier combination resulted in highest accuracy of 96.8%, sensitivity of 100% and specificity of 93.7% compared to other data reduction techniques (PCA and LDA) and classifiers. Overall, compared to previous techniques, our proposed strategy is more suitable for diagnosis of CAD with higher accuracy.

1. Introduction

Coronary Artery Disease (CAD) has reached nearly epidemic proportions in our society and is the cause of more deaths, disability and economic loss than any other group of diseases [11]. CAD is a condition where the arteries which supply blood to the heart muscles are hardened and narrowed. This generally occurs due to the deposition of plaque within the arteries. When the deposits increase in quantity, the amount of blood that flows reduces and the availability of oxygen also decreases. CAD can also weaken the heart muscles and contribute to heart failure which affects the pumping action of the heart, and to arrhythmias which affects the normal beating rhythm of the heart (CAD) [12].

There are several diagnostic tools for CAD. Diagnosis generally begins with the onset of one of the common symptoms of the disease which is heart attack or a sudden cardiac arrest. Some of the general diagnostic tests include physical examination, lab tests, Electrocardiogram (ECG), echocardiogram, stress test, electron beam computed tomography, coronary angiography and cardiac catheterization. Symptoms of angina are generally diagnosed through a tread mill stress test. Though ECG is used during several analyses, one of the major limitations of this technique is the undiagnosed symptoms of CAD [12]. ECG recording can show a normal

reading for those patients suffering from CAD. Hence, the only alternative is the angiogram which is an invasive methodology. This technique is painful to the patients and causes discomfort to them. Moreover, the above mentioned procedures are specialized processes which involve a lot of time, effort and cost. They can only be conducted by trained people [12].

Computer aided diagnostic methods which extract relevant features and use them in classifiers for automated detection of diseases can overcome these difficulties. Such techniques are non-invasive and provide reproducible and objective diagnoses, and hence, can prove to be valuable adjunct tools in clinical practice. Hence, in this work, we propose one such new technique that uses Heart Rate Variability (HRV) analysis. Also, the subtle variations in the P-QRS-T wave may be difficult to identify by morphological changes [3]. Hence, we have used the HRV in this work to identify the CAD and normal classes.

Recently, analysis methods based on nonlinear dynamics have been developed to describe the complex behavior of cardiac abnormalities [1–4,9,14,23]. The studies using these new methods have reported typical self-similar or fractal like features in the R–R interval dynamics of cardiac rhythm [1–4,14] and have also implicated abnormal Heart Rate (HR) behavior of patients with structural heart disease [9,23,35,36]. Typical spectral pattern under normal conditions indicates three frequency bands: a very low frequency band from 0.00 to 0.03 Hz, a low frequency (LF) band from 0.03 to 0.15 Hz and a high frequency (HF) band in respiratory range generally more than 0.25 Hz [19]. The power of LF component is related to the sympathetic activation. HF power provides a quantitative index of the influence of respiration and may be connected to the vagal activity. LF/HF ratio depicts sympatho-vagal balance [27].

Heart rate variability signals are complex and vary in an irregular manner. Linear statistical methods such as mean values, variability measures, and power spectra of such signals may not depict directly their complexity and thus may miss useful subtle information. Numerous statistical and frequency domain analysis methods have been used to detect the CAD using HRV signals. It was shown that CAD patients had a reduced circadian rhythm of HRV than the normal subjects. It was found that angiographic severity of CAD and reduction in high-frequency power were correlated. It was observed that CAD subjects had reduction in high frequency power. Bigger et al. [9] observed that both time and frequency measures of HRV were lower in CAD patients. However, both time and frequency domain analyses are sensitive to noise, and may be prone to more errors. Hence, wavelet analysis may be better suited for the analysis of CAD HRV signals. Since the underlying mechanisms involved in the control of heart rate is nonlinear and non-stationary, the use of nonlinear techniques may be more appropriate [1,3,14,15]. Recently, new nonlinear methods have been applied to HRV analysis. They are: Lyapunov exponents [1,3], correlation dimension [40,1], fractal dimension [2], 1/f slope [31], approximate entropy (ApEn) [41,1] and Detrended Fluctuation Analysis [39]. Nonlinear parameters like correlation dimension, entropy features were extracted from HRV signals to detect CAD [32,37,42]. It was found that Detrended Fluctuation Analysis (DFA) features and fractal properties of RR interval dynamics were reduced in CAD patients [28].

Baseline wander, muscle tremors and other artifacts may influence measurement in the time domain methods. There can be many biological signals of different rhythms with similar *means* and standard deviations. Hence, time domain methods may not be used to extract relevant features for Heart Rate Variability (HRV) signals.

In order to obtain a meaningful analysis using frequency domain analysis, the HRV signal need to be assumed as stationary and periodic in nature, having positive and negative alterations. The assumption of stationarity may not hold good when the HRV signal of long durations are being analyzed. Also, the spectral analysis methods are more sensitive to the presence of artifacts than time domain methods. Due to the limited frequency resolution of the wavelet transforms were found to be more suitable to HRV analysis and able to extract dynamical information from these signals.

Using wavelet analysis and neural network, a system was designed to classify normal and CAD subjects [29]. Linear and non linear parameters were extracted and classified using SVM classifier [33]. Using empirical mode decomposition and Teager energy operator, on normal and CAD heart rate variability signals were classified using back propagation neural network [46]. PCA was used for data reduction on normal and CAD heart rate variability ECG signals using SVM classifier [8].

In this work, we have proposed a novel efficient method to automatically diagnose CAD using Discrete Wavelet Transform (DWT) – and dimensionality reduction techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) and classifiers. The layout of the paper is as follows. Section 2 explains the data acquisition and pre-processing. Section 3 discusses the methodology used for this work. Section 4 of the paper presents the results of the proposed method. The results are discussed in Section 5, and finally, the paper concludes in Section 6.

2. Methodology

This section comprises of data acquisition, pre-processing, methods used for feature extraction DWT, PCA, LDA and ICA and various classifiers are briefly explained. They are briefly explained below.

2.1. Data acquisition

The Electrocardiograms of ten *CAD* patients and fifteen *normal* subjects were recorded using the BIOPAC™ equipment (BIOPAC) [10]. The ECG signal was sampled at 500 Hz. The age of both the *CAD* and *normal* groups varied from 40 to 70 years with mean age of 55 years. The *CAD* group were in-patients of Iqraa Hospital, Calicut Kerala, India. The cardiologist identified them, their medications were identical, and the effect of drug on the HRV signal was also found to be similar. In total there were 96 data files comprising both normal and CAD subjects. Each data file consisted of 15 min ECG. From each ECG file, heart rate variability data consisting of 1000 intervals is generated as explained in the next section. The typical heart rate variability signal of the *CAD* and *normal* subjects is shown in Fig. 1. We assume that relevant features can be extracted within this duration [1–4]. We have used threefold cross-validation with 32 files (observations/patterns) per fold. Hence in each fold there will be 16 data files of each classes (normal and CAD). The 16 patterns are sampled on the pattern space such that entire pattern space is covered.

Due to the limitation of obtaining data, we have acquired the ECG data from 10 CAD subjects and 15 normal subjects for a duration of 15 min each on different days under supine condition. We have used 1000 samples (almost 15 min of data) for analysis. We assume that relevant features can be extracted within this duration. We have used threefold cross-validation with 32 files (observations /patterns) per fold. Hence in each fold there will be 16 data files of each classes (normal and CAD). The 16 patterns are sampled on the pattern space such that entire pattern space is covered.

2.2. Pre-processing

Original ECG signals were passed through a low-pass filter (second order Butterworth filter) with a cut-off frequency of 15 Hz to
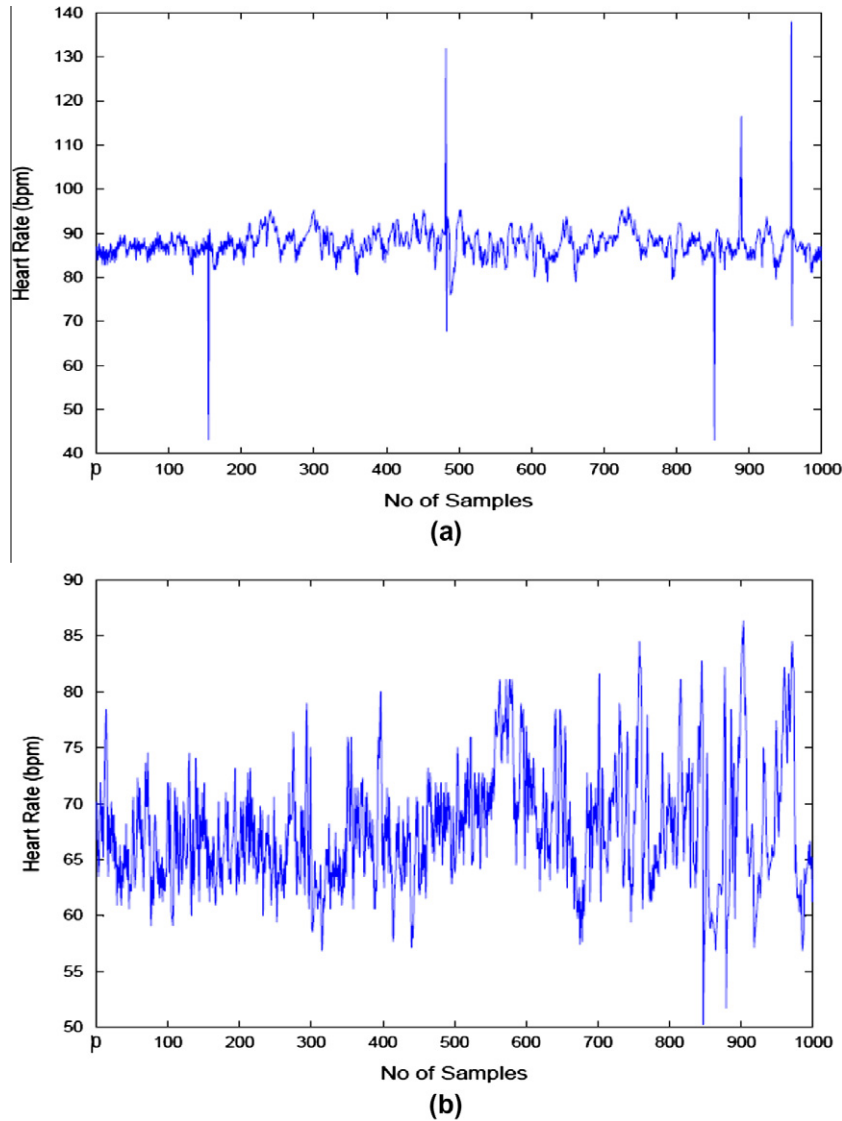
Fig. 1. Typical heart rate signal: (a) CAD and (b) normal.

remove the unwanted high frequency noise. The baseline wander was removed by passing through a high-pass filter of cut-off frequency of 0.3 Hz. Power-line interference noise was removed using a band-reject filter of cut-off frequency of 50 Hz. Finally, the R peaks of ECG signal were detected. RR interval is defined as the interval between two successive QRS complexes. Heart rate (beats per minute) is calculated from the RR interval (in seconds) using the equation:

$$HR_{bpm} = \frac{60}{RR} \tag{1}$$

2.3. Feature extraction

Fig. 2 shows the proposed system of our work. The pre-processed heart rate data segments (instances) are subjected to DWT in order to obtain wavelet coefficients. These coefficients are fed to three dimensionality reduction techniques: PCA, LDA and ICA. Significant features (coefficients) are selected using the Student's t-test. These significant features are used to train and evaluate the performance of classifiers. The classifier output will

be two classes (normal/CAD). The techniques employed are described in this section.

2.3.1. Wavelet transform

Wavelet analysis provides both time and frequency localization and the resultant wavelet coefficients can be used as features in classifiers [17]. The wavelet representation of a signal is sparse compared to time domain representation due to energy compaction property of wavelet transform. In order to decompose the heart rate variability signal into time–frequency components, a basis function at scale a and location b is defined as,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

Eq. (1) defines continuous wavelet transform. It is sampled on a dyadic grid to obtain the DWT. The basis function of DWT at scale 2^{-m} and the time instant n is given by,

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \tag{3}$$

Using the dyadic wavelets, the DWT of the signal $x(t)$ is given by,

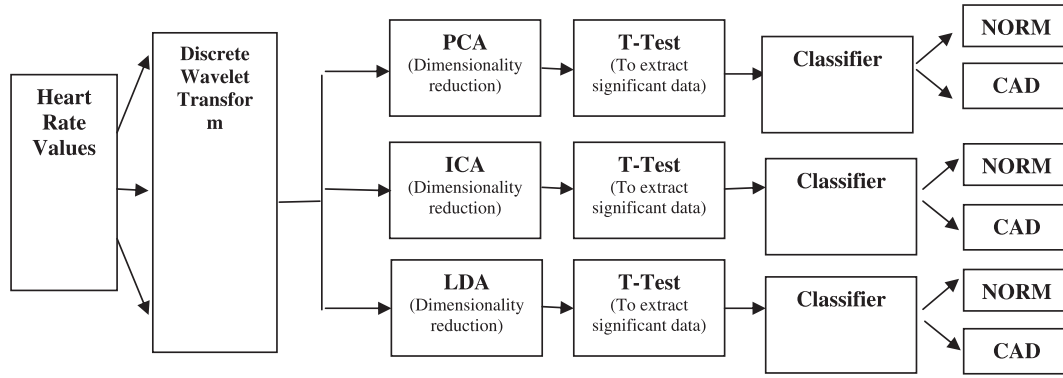$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt \tag{4}$$

Fig. 2. The proposed system.

The inverse DWT is given by,

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T_{m,n} \psi_{m,n}(t) \qquad (5)$$

To decompose a signal, the basis function shape must be similar to that of the signal. The heart rate variability signal is decomposed using 54 basis functions. They are Haar, Db2 to Db20, Symlet, Coiflet and Legendre wavelets.

2.3.2. Dimensionality reduction techniques

In this work, we have used three dimensionality reduction techniques, namely, Principal Component Analysis (PCA), Linear discriminant Analysis (LDA) and Independent Component Analysis (ICA). Based on the classification performance the best dimensionality reduction method is identified. The details of each of these techniques are mentioned in depth in the following sections. By trial and error method we have chosen the ten features (principal components, linear discriminant scores and independent components). After ten, the features will be very small in magnitude.

2.3.2.1. Principal Component Analysis (PCA).
PCA is a linear dimensionality reduction technique that seeks projection in the directions of maximum variability [18,44]. It involves computation of Eigen values and Eigen vectors of the data covariance matrix, sorting of Eigen vectors in the descending order of Eigen value and finally projecting the actual data into the directions of sorted Eigen vectors. The step by step procedure of PCA is provided below:

- *Step 1*: Calculate the covariance matrix of order $N \times N$ for the N dimensional signal samples as,

$$\Sigma = \frac{1}{N}\{(x - \bar{x})(x - \bar{x})^T\} \qquad (6)$$

where x is the given signal matrix consisting of M data points each of dimension N, and $\bar{x}$ is its mean vector.
- *Step 2*: Find matrix V of Eigenvectors and diagonal elements of matrix D as eigenvalues of covariance matrix Σ as given by,

$$V^{-1}\Sigma V = D \qquad (7)$$

- *Step 3*: Sort the eigenvectors in the descending order of Eigenvalues in D.
- *Step 4*: Project the data into the directions of sorted Eigenvectors by taking the dot product between the given data and Eigenvectors.
- *Step 5*: Choose first few principal components depending on containment of a given percentage of variability (like 95% or 98% depending on the problem).

2.3.2.2. Independent Component Analysis.
ICA is another feature extraction method that transforms multivariate random signal into a signal having components that are mutually independent. It is assumed that each measured signal is a linear combination of each of the independent signals, and that there are an equal number of measured signals and independent signals. Assume that we observe n linear mixtures of the form $\boldsymbol{x}_1, \ldots, \boldsymbol{x_n}$. The mixtures are a linear combination of $\boldsymbol{n}$ independent components.

$$\boldsymbol{x_j} = \boldsymbol{a_{j1}}\boldsymbol{s_1} + \boldsymbol{a_{j2}}\boldsymbol{s_2} + \cdots + \boldsymbol{a_{jn}}\boldsymbol{s_n} \qquad (8)$$

where s_k ($k = 1$ to n) represent the independent components we are trying to find. Without loss of generality we can assume that both the mixture variables and the independent components have zero mean. Using vector notation we can let $\boldsymbol{x}$ denote the mixtures $\boldsymbol{x}_1$, $\boldsymbol{x}_2, \ldots, \boldsymbol{x_n}$. Here "$\boldsymbol{s}$" denotes the independent components. If the mixing matrix is denoted as "$\boldsymbol{A}$", the mixing model is written as; $\boldsymbol{x} = \boldsymbol{As}$. This model is known as Independent Component Analysis. After estimating the matrix A, its inverse W can be computed and the independent components can be determined as $\boldsymbol{s} = \boldsymbol{Ws}$. The steps of ICA are as follows: [25,45]

- *Step 1*: **Centering of the data**: This is the process of subtracting the mean from the data. This is to ensure that the components have a zero mean.
- *Step 2*: **Whitening of the data**: This is another pre-processing method. During whitening, the mixture is transformed so that its components are uncorrelated and its variances are unity.
- *Step 3*: **Selection of the independence criteria**: The independence criteria chosen depend on the data to be analyzed. There are several methods which include ICA by maximization of non-Gaussianity, ICA by Kurtosis Maximization or Minimization, Negentropy and the FastICA algorithm. In this study, we used Negentropy method.
- *Step 4*: **Data Reconstruction**: The output obtained on applying the independence criteria is multiplied with the whitening output. This result is multiplied with the transpose of the input data.

2.3.2.3. Linear Discriminant Analysis.
The PCA technique, optimally projects the data in the direction of maximum variability, but the PCA algorithm never uses the information on which signal belongs to which class. Therefore the PCA directions represent the entire data irrespective of their class labels. These directions are not guaranteed to provide maximum separation between the classes. There is a need of the method which provides maximum separation between the classes. Linear Discriminant Analysis (LDA) provides highest separation between the classes present in the data [20].
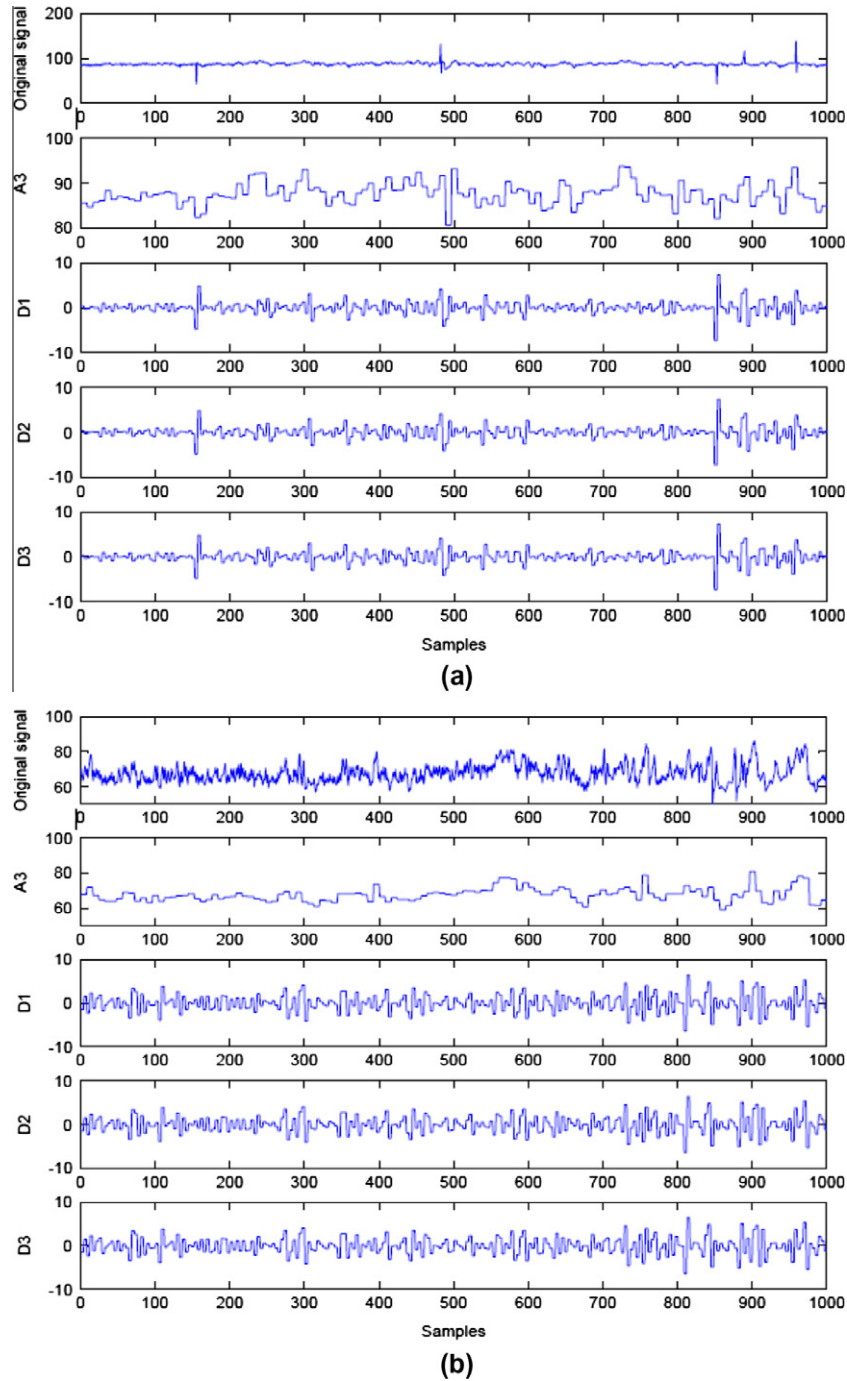
Fig. 3. DWT decomposition of HRV signal: (a) CAD and (b) normal.

LDA maximizes between class scatter measure S_b, while minimizing the within class scatter matrix S_w. The matrices S_b and S_w are given as:

$$S_w = \sum_{i=1}^{C} \sum_{x_j \in C_i} (x_j - \bar{m}_i)(x_j - \bar{m}_i)^T \qquad (9)$$

$$S_b = \sum_{i=1}^{C} n_i (m_i - \bar{m})(\bar{m}_i - \bar{m})^T \qquad (10)$$

where $\bar{m}_i$ is the ith class mean, $\bar{m}$ is the global mean, n_i is the number of samples in the ith class. The LDA space is spanned by the set of vectors W satisfying the following equation.

$$W = \arg_W \max \frac{|W^T S_b W|}{|W^T S_w W|} \qquad (11)$$

W consists of Eigenvectors corresponding to k largest Eigenvalues of the matrix, $S_w^{-1} S_b$.

2.3.3. Students t-test

This test assesses whether the *means* of two groups are statistically different from each other. We use *t*-test to assess the capability of extracted features to discriminate between *normal* and *CAD* groups. In this work, we have considered features with very low *p*-values (less than 0.0001) as significant.

2.3.4. Classifiers

In this work, we have used four classifiers namely Support Vector Machine (SVM), Gaussian Mixture Model (GMM), K-Nearest Neighbor (KNN), and Probabilistic Neural Network (PNN). They are briefly explained below.

2.3.4.1. Support Vector Machine (SVM). The SVM is a single layer and highly non linear network based on statistical learning principles. It has ability to classify unseen patterns correctly [5,16,26,38]. Unlike other classifiers, SVM minimizes the structural risk rather than empirical risk. During the training of SVM, it maximizes the distance from patterns to the class separating hyper-plane. Generally the patterns are not linearly separable, therefore non linear kernel transformation is performed. There are various kernels that can be used during SVM training. Some of them are quadratic, polynomial and Radial Basis Function (RBF) kernels. In this study RBF kernel is used for kernel transformation.

The SVM with a RBF kernel function has two training parameters: cost (C) which controls over fitting of the model, and sigma (σ) which controls the degree of nonlinearity of the model. The values of these training parameters C and sigma are determined using a "grid search" approach [24]. The final SVM parameters obtained were: cost constant C: 110, and sigma: 0.001.

2.3.4.2. Gaussian Mixture Model (GMM). The Gaussian Mixture Model (GMM) assumes that the data is drawn from a few mixtures of Gaussian distributions, where the number of mixing components is made equal to the number of classes present in the data. In this study there are two mixing components [5,21,43]. Initially some random values are defined as the initial a priori probability. Some random mean and covariance matrix are initialized. Based on initial model parameters (the apriori, mean and covariance matrix), the class conditional probability density is computed. Based on this probability, the posterior density is computed using Baye's theorem. The model is re-estimated from the data and new values for a priori probability, class mean and covariance matrix are computed. The process is iterated until the difference between the new parameters of the model and the old parameters will be less than a threshold. The model is said to be converged and the testing data is applied to the GMM model and the performance is noted. In this classifier, the clusters centers were initialized to 4 for normal and CAD groups.

2.3.4.3. K-Nearest Neighbors (KNN). KNN classification is based on the closest training examples in the feature space [5,21]. This is a type of instance based learning technique also called non parametric lazy algorithm. This technique does not use any assumptions on the data distribution, and hence, is called non parametric. In this algorithm, the k-nearest neighbors are estimated and majority voting is performed. During this, the class which is found most common among the k neighbors is assigned as the class for the new data.

In this classifier, we varied the k nearest neighbors k = 1–6. We obtained the maximum accuracy for k = 2. The distance was computed using Euclidean distance. The majority rule was used for classification.

2.3.4.4. Probabilistic Neural Networks (PNN). PNN is a form of radial basis network which is used in classification problems [5,21]. This model learns to approximate the probability density function of the training samples through training processes. PNN consists of three main layers: pattern layer, summation layer and output layer [22]. The pattern layer forms a product of the weight vector with the input data set. This output is fed into the summation layer which receives the outputs associated with a given class. The final

output layer produces the classification decisions. One of the biggest advantages of this classifier is its high training speed.

In the PNN classifier, all biases in the radial basis layer were set to $\sqrt{\ln 0.5}/s$, where s = spread constant of PNN. In this work, we obtained maximum accuracy when s = 0.1.

3. Results

The heart rate data obtained as time series was transformed into time–frequency sub-bands using DWT (Fig. 3). In total the third level detail coefficients will be 76 in number. We have considered 76 wavelet coefficients from 3rd level detail sub band. The DWT coefficients were used as inputs for dimensionality reduction. Three dimensionality reduction techniques were used independently to identify best representation of features. They are PCA, ICA and LDA. From each of these techniques ten significant components were chosen for subsequent pattern classification. In each of the dimensionality reduction method (PCA, LDA and ICA), the dimensionality is reduced from 76 coefficients to 10 coefficients. These 10 coefficients/features are classified using different classifiers to one dimensional binary space, where 0 indicates normal and 1 indicates CAD. So SVM grid search is from 10 dimension to one dimension. In KNN classification k = 2 is taken. Table 1 shows the mean and Standard Deviation (SD) of the first 10 clinically significant Principal Components (PCs) for *normal* and *CAD* groups. Table 2 shows the mean and Standard Deviation (SD) of the first 10 clinically significant Independent Components (ICs) for *normal* and *CAD* groups. Table 3 shows the mean and Standard Deviation (SD) of the first 10 clinically significant Linear Discriminant scores (LDs) arranged in descending order for *normal* and *CAD* groups.

The *ten* features extracted through these three techniques were used to test the *four* classifiers using threefold cross validation. We have used 1000 time samples (almost 15 min of data) per pattern for analysis. We assume that relevant features can be extracted within this duration. We have used threefold cross-validation with 32 files (observations/patterns) per fold. Hence in each fold there will be 16 data files of each classes (normal and CAD). The 16 patterns are sampled on the pattern space such that entire pattern space is covered. The two folds merged together to form 64 data files are the instances in learning. There will be 32 instances (the 32 data files in the third fold) for testing. The process is repeated another two times so that each of the threefold is participated in testing. Performance measures such as sensitivity, specificity, accuracy, and Positive Predictive Value (PPV) are calculated on testing data. This procedure is repeated two more times by using a new test set each time. The average of all the three performance measures is considered as the final performance measure value.

The number of True Positives (TP), False Negatives (FN), True Negatives (TN), and False Positives (FP) obtained using each classifier and each dimensionality reduction technique is shown in Tables 4–6. Let TN (True Negative) be the number of normal cases

Table 1
Principal components for CAD and normal classes (*p*-value < 0.0001).

Features	CAD (mean ± SD)	Normal (mean ± SD)
PC1	1.082E−03 ± 2.034E−03	1.024E−02 ± 1.460E−02
PC2	4.980E−03 ± 1.017E−02	2.190E−02 ± 2.020E−02
PC3	1.155E−02 ± 2.031E−02	3.461E−02 ± 2.605E−02
PC4	1.611E−02 ± 2.330E−02	4.643E−02 ± 3.032E−02
PC5	2.200E−02 ± 2.650E−02	6.057E−02 ± 3.571E−02
PC6	3.122E−02 ± 3.219E−02	6.973E−02 ± 4.073E−02
PC7	3.856E−02 ± 3.373E−02	8.324E−02 ± 4.764E−02
PC8	4.621E−02 ± 3.589E−02	9.652E−02 ± 5.395E−02
PC9	5.720E−02 ± 3.800E−02	0.108 ± 5.815E−02
PC10	6.498E−02 ± 3.694E−02	0.121 ± 6.634E−02

Table 2
Independent components for CAD and normal classes (p-value < 0.0001).

Features	CAD (mean ± SD)	Normal (mean ± SD)
IC11	1.41 ± 0.496	1.71 ± 0.142
IC12	1.35 ± 0.501	1.64 ± 0.134
IC13	1.29 ± 0.482	1.59 ± 0.140
IC14	1.24 ± 0.466	1.53 ± 0.136
IC15	1.20 ± 0.454	1.49 ± 0.119
IC16	1.16 ± 0.445	1.44 ± 0.118
IC17	1.11 ± 0.434	1.40 ± 0.120
IC18	1.08 ± 0.419	1.37 ± 0.122
IC19	1.06 ± 0.413	1.34 ± 0.120
IC20	1.03 ± 0.401	1.31 ± 0.122

Table 3
Linear discriminant scores for CAD and normal classes (p-value < 0.0001).

Features	CAD (mean ± SD)	Normal (mean ± SD)
LD1	0.320 ± 0.282	3.227E-02 ± 0.134
LD2	0.274 ± 0.261	3.728E-02 ± 0.135
LD3	0.249 ± 0.247	3.477E-02 ± 0.133
LD4	0.232 ± 0.244	3.412E-02 ± 0.134
LD5	0.217 ± 0.233	3.503E-02 ± 0.134
LD6	0.208 ± 0.232	3.609E-02 ± 0.135
LD7	0.205 ± 0.232	3.525E-02 ± 0.135
LD8	0.196 ± 0.228	3.494E-02 ± 0.135
LD9	0.190 ± 0.228	3.500E-02 ± 0.134
LD10	0.186 ± 0.225	3.557E-02 ± 0.134

identified as normal, FN (False Negative) be the number of CAD cases incorrectly identified as normal, TP (True Positive) be the number of CAD cases correctly identified as they are, and FP (False Positive) be the number of normal cases incorrectly identified as CAD. **Sensitivity** is the probability that a classifier will produce a positive result when used on the CAD population (TP/(TP + FN)). **Specificity** is the probability that a classifier will produce a negative result when used on the normal population (TN/(TN + FP)). **Accuracy** is the ratio of the number of correctly classified samples to the total number of samples ((TP + TN)/(TP + FN + TN + FP)). **PPV** is the proportion of patients with positive results who are correctly diagnosed (TP/(TP + FP)).

Classification was done three times using components obtained through PCA, ICA and LDA as inputs. All the different classifiers gave different levels of performance when the inputs were changed. The tables below give a summary of the classification results for the three sets of inputs – PCA Classification (Table 4), ICA Classification (Table 5), and LDA Classification (Table 6). It can be seen that ICA combined with GMM classifier yielded the highest accuracy of 96.8%, sensitivity of 100% and specificity of 93.7%. PCA coupled with GMM resulted in an accuracy of 93.7%, sensitivity of 87.5% and specificity of 100%. LDA combined with KNN resulted in an accuracy of 87.5%, sensitivity and specificity of 81.2% and 93.7% respectively.

Table 4
TN, FN, TP, FP, Accuracy (Acc), Sensitivity (Sn), Specificity (Sp), PPV of all the classifiers using PCA components as input features.

Classifier	TN	FN	TP	FP	Acc	PPV	Sn	Sp
PNN	12	0	16	4	87.5	80	100	75
KNN	13	3	13	3	81.2	81.2	81.2	81.2
SVM	13	0	16	3	90.6	100	84.2	100
GMM	16	2	14	0	93.7	100	87.5	100

Table 5
TN, FN, TP, FP, Accuracy (Acc), Sensitivity (Sn), Specificity (Sp), PPV of all the classifiers using ICA components as input features.

Classifier	TN	FN	TP	FP	Acc	PPV	Sn	Sp
PNN	11	2	14	5	78.1	73.6	87.5	68.7
KNN	11	2	14	5	78.1	73.6	87.5	68.7
SVM	15	3	13	1	87.5	92.8	81.2	93.7
GMM	15	0	16	1	96.8	94.1	100	93.7

Table 6
TN, FN, TP, FP, Accuracy (Acc), Sensitivity (Sn), Specificity (Sp), PPV of all the classifiers using LDA components as input features.

Classifier	TN	FN	TP	FP	Acc	PPV	Sn	Sp
PNN	15	4	12	1	84.3	92.3	75	93.7
KNN	15	3	13	1	87.5	92.8	81.2	93.7
SVM	11	1	15	5	81.2	75	93.7	68.7
GMM	14	4	12	2	81.2	85.7	75	87.5

4. Discussion

PCA captures the second order correlation value of the series. And the correlation between the components of the data vector can be clearly viewed. PCA is captures linear relations and does not pick the nonlinear relations. The extracted features are orthogonal and variant under transformation and is a non-surveillance method. LDA method discriminates different classes and is a surveillance method. It does not perform well, if the number of samples is small. ICA captures the higher order statistics of data. It is parallel, distributed, computationally, simple, and requires little space [13].

Results of feature extraction show that, p-value is low (<0.0001) for all the three methods (PCA, LDA and ICA) giving good performance. Figs. 4–6 shows the *mean* values of the PC, LD and ICs. It can be seen from Fig. 5 that, the difference in the *mean* value between the two classes is high (about 0.3) for all the ten independent components. The difference in the *mean* of the linear discriminant values are less than 0.3 for most of the linear discriminants. Similarly the difference in the *mean* value is lower for all the principal components. Hence, the classification accuracy is higher for ICA based features as compared to the LDA and PCA based features.
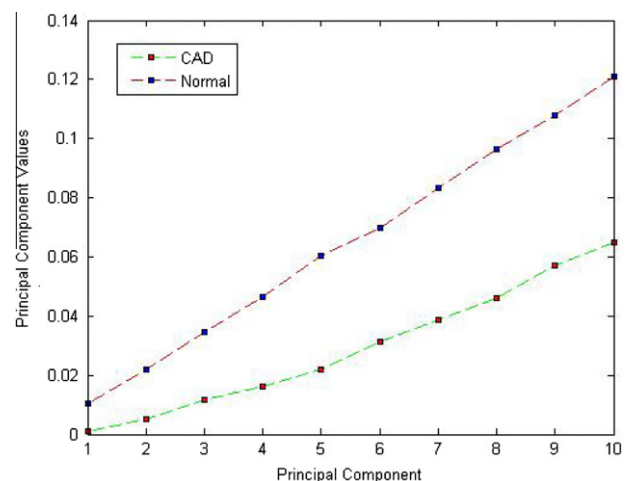

Fig. 4. Extracted features (mean of PCs) of normal and CAD classes through student T test.
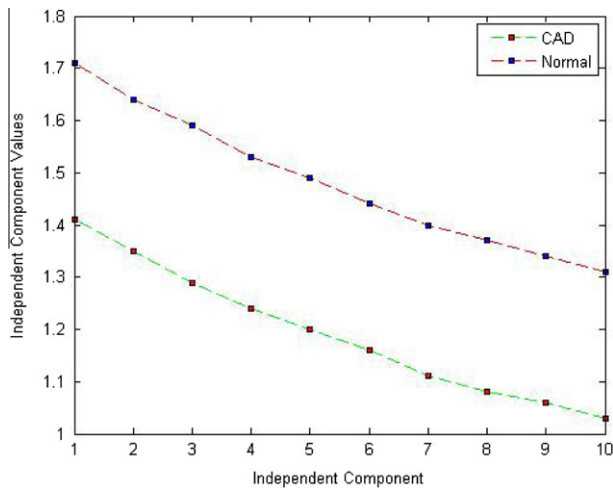
Fig. 5. Extracted features (mean of ICs) of normal and CAD classes through student *T* test.

Fig. 6. Extracted features (mean of LDs) of normal and CAD classes through student *T* test.

There are 54 types of wavelet functions including Haar wavelet, Daubechies Wavelet (1–20), Symlet, Legendre Wavelet, Coiflet to name a few. In this study all the different wavelet functions were tried and it was found that Daubechies 8 (Db8) yielded the highest accuracy. Hence, we chose Db8 as the mother wavelet for this entire study.

Wavelet features extracted from heart sounds in combination with neural network yielded an accuracy of 85% in automatically identifying CAD [29]. Arafat et al. [6] used a combination of fuzzy and probabilistic uncertainty types for CAD detection using stress ECG signals and obtained an accuracy of around 80%. However, exercise stress testing has limited use as most patients will not be able to reach the heart rate required for this test, and, hence, obtaining a good representative feature set using this test is not the best option.

Kim et al. [30] developed a multi-parametric measure using multiple discriminant analysis of linear and nonlinear parameters extracted from the HRV signal. The ECG signals were obtained from three recumbent positions, the supine, left lateral, and right lateral positions. They reported accuracies of 75.0%, 72.5%, and 84.6% for the normal, Angina Pectoris (AP), and Acute Coronary Syndrome (ACS) groups, respectively. In a similar study, Lee et al. [33] used linear and nonlinear HRV parameters in classifiers for detecting CAD. SVM presented the highest accuracy of almost 90% using features from all three recumbent positions. Using a similar set of features, the same group [34] obtained a classification accuracy of 88.33% for the classification of normal, AP, and ACS classes. They included the carotid arterial wall thickness in addition to HRV features and observed a classification accuracy of around 85–90% for Classification based on Predictive Association Rules (CPAR) and SVM classifiers.
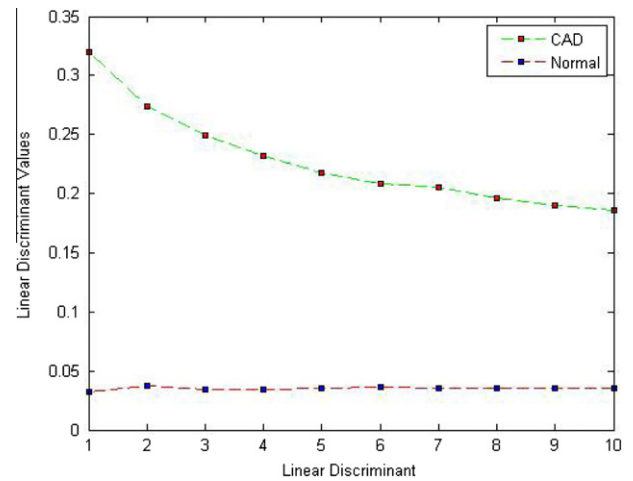
Empirical mode Decomposition – Teager Energy operator was used to extract features from the heart murmur signals and these features were fed to the backpropagation neural network to classify normal and CAD subjects with 85% accuracy [46]. Binary particle swarm optimization and genetic algorithm techniques were used on exercise stress testing data to identify CAD automatically with an accuracy of 81.46% using SVM classifier [7]. In another study, the same group used PCA features extracted from the exercise stress test data and were able to automatically diagnose CAD with an accuracy of 79.1% [8]. Table 7 provides the summary of work carried out in the detection of CAD using ECG and HRV signals.

In our present study, we extracted features from HRV instead of ECG signals. We extracted the DWT coefficients and applied PCA, ICA and LDA on the extracted coefficients for dimensionality reduction. We fed the highest ten significant coefficients obtained using each dimensionality reduction technique to four classifiers in order to determine the best dimensionality reduction technique and classifier combination that resulted in the highest classification accuracy for CAD detection. We obtained the highest accuracy of 96.8%, sensitivity of 100% and specificity of 93.7% for ICA and GMM combination. Even though this accuracy is higher than that obtained in the previously published studies, we believe that the performance of our technique can be further improved by taking more HR data, better features and robust classifiers.

The main contribution of this study is that of a new methodology to yield highest possible accuracy in classification. We have performed exhaustive experiments and shown that Independent Component Analysis features provide highest accuracy (96.8%) with Gaussian Mixture Model (GMM) classifier to discriminate normal and CAD heart rate variability signals.

Table 7
Summary of studies reporting automated detection of CAD and normal subjects.

Authors	Features/techniques used	Classifiers	Accuracy (%)
Karimi et al. [26]	Wavelet analysis	Neural network	85
Arafat et al. [6]	ECG Stress Signals with Probabilistic Neural Networks	Fuzzy Inference Systems	80
Lee et al. [33]	Linear and Nonlinear Parameters	SVM Classifier	90
Kim et al. [30]	Multiple Discriminant Analysis with linear and non linear feature	Different Classifiers	72.5–84.6
Zhao and Ma [46]	Emperical Mode Decomposition-Teager Energy Operator	Back Propagation Neural Network	85
Lee et al. [34]	HRV, carotid arterial wall thickness	CPAR and SVM	85–90
Babaoglu et al. [7]	Binary Particle Swarm Optimization	SVM	81.46
Babaoglu et al. [8]	PCA for dimension reduction	SVM	79.71
In this work	HRV signals and ICA	GMM	96.8

In the present study, the authors have identified a best dimensionality reduction method (WT + ICA) and a best classifier (GMM) among other alternatives to provide highest possible performance. This part was not present in the literature and hence is a novel contribution.

5. Conclusion

Heart rate signals can be used to discriminate *normal* and *Coronary Artery Disease* states. In this work, we extracted ten features from full time series heart rate data trained four classifiers. We have shown that the GMM classifier could differentiate between the two classes with clinically significant classification accuracy of 96.8%, sensitivity of 100% and specificity of 93.7% when trained with the ten independent components. These performance measures were relatively higher than those registered using other data reduction techniques such as PCA and LDA. Thus, the proposed methodology offers good diagnostic accuracy and automation. Hence, it is a suitable candidate for testing in a clinical environment, and could be used in routine diagnostic protocols in cardiac diagnostic centers in future.

Acknowledgement

References

[1] U.R. Acharya, N. Kannathal, S.M. Krishnan, Comprehensive analysis of cardiac health using heart rate signals, Physiol. Meas. 25 (2004) 1130–1151.
[2] U.R. Acharya, P.S. Bhat, N. Kannathal, A. Rao, C.M. Lim, Analysis of cardiac health using fractal dimension and wavelet transformation, Innov. Tech. Biol. Med. 26 (2005) 133–139.
[3] U.R. Acharya, K.P. Joseph, N. Kannathal, C. Lim, J.S. Suri, Heart rate variability: a review, Med. Biol. Eng. Comput. 44 (2006) 1031–1051.
[4] U.R. Acharya, M. Sankaranarayanan, J. Nayak, C. Xiang, T. Tamura, Automatic identification of cardiac health using modeling techniques: a comparative study, Inform. Sci. 178 (2008) 4571–4582.
[5] U.R. Acharya, S.V. Sree, S. Chattopadhyay, W. Yu, A.P.C. Alvin, Application of recurrence quantification analysis for the automatic identification of epileptic EEG signals, International Journal of Neural Systems 21 (3) (2011) 199–211.
[6] S. Arafat, M. Dohrmann, M. Skubic, Classification of coronary artery disease stress ECGs using uncertainty modeling, Congr. Comput. Intell. Methods Appl. (2005).
[7] I. Babaoglu, O. Findik, E. Ülker, A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine, Exp. Syst. Appl. 37 (2010) 3177–3183.
[8] I. Babaoglu, O. Findik, M. Bayrak, Effects of principle component analysis on assessment of coronary artery diseases using support vector machine, Exp. Syst. Appl. 37 (2010) 2182–2185.
[9] J.T. Bigger Jr., R.C. Steinman, L.M. Rolnitzky, J.L. Fleiss, P. Albrecht, R.J. Cohen, Power law behavior of RR-interval variability in healthy middle-aged persons, patients with recent acute myocardial infarction, and patients with heart transplants, Circulation 93 (1996) 2142–2151.
[10] BIOPAC^{TM}. Information available at <http://www.biopac.com/> (accessed 24.02.12).
[11] P.H. Brubaker, Coronary Artery Disease: Essentials of Prevention and Rehabilitation Programs, Human Kinetics Publishers, 2002.
[12] CAD. Information available at <http://www.nhlbi.nih.gov/health/health-topics/topics/cad/> (accessed 24.02.12).
[13] L. Cao, J. Chua, K.S. Chong, W.K. Lee, Q.M. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, Neurocomputing 55 (2003) 321–336.
[14] C.K. Chua, V. Chandran, U.R. Acharya, C.M. Lim, Cardiac health diagnosis using higher order spectra and support vector machine, Open Med. Inform. J. 3 (2009) 1–8.
[15] D. Cysarz, H. Bettermann, P. van Leeuwen, Entropies of short binary sequences in heart period dynamics (vol. 728), Am. J. Physiol. Heart Circ. Physiol. 278 (2000) H2163–H2172.
[16] V. David, A. Sanchez, Advanced support vector machines and kernel methods, Neurocomputing 55 (2003) 5–20.
[17] S. Dua, U.R. Acharya, P. Chowriappa, S. Vinitha Sree, Wavelet-based energy features for glaucomatous image classification, IEEE Trans. Inf. Technol. Biomed. 16 (2012) 80–87.
[18] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley, 2001.
[19] European Heart Journal, Task Force of the European Society of Cardiology and North American Society of Pacing and electrophysiology Heart Rate Variability: Standards of measurement, physiological interpretation and clinical use 17 (1996) 354–381.
[20] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed., Academic Press, 1990.
[21] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, second ed., Morgan Kaufmann, 2005.
[22] S. Haykin, Neural Networks and Learning Machines, third ed., Prentice Hall, 2009.
[23] K.K.L. Ho, G.B. Moody, C.K. Peng, J.E. Mietus, M.G. Larson, D. Levy, A.L. Goldberger, Predicting survival in heart failure cases and controls using fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics, Circulation 96 (1997) 842–848.
[24] C.W. Hsu, C.C. Chang, C.J. Lin, A Practical Guide to Support Vector Classification, Technical Report, Department of Computer Science, National Taiwan University, 2003.
[25] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4–5) (2000) 411–430.
[26] Jinchang, Ren, Ann Vs. Svm: Which One Performs Better in Classification of Mccs in Mammogram Imaging, Knowledge-Based Systems (in press).
[27] M.V. Kamath, D.N. Ghista, E.L. Fallen, D. Fitchett, D. Miller, R. McKelvie, Heart rate variability power spectrogram as potential noninvasive signature of cardiac regulatory system response, mechanisms, and disorders, Heart Vessels 3 (1987) 33–41.
[28] K. Karamanos, S. Nikolopoulos, G. Manis, A. Alexandridi, K. Hizanidis, S. Nikolakeas, Block entropy analysis of heart rate variability signals, Int. J. Bifurcat. Chaos. 16 (2006) 2093–2101.
[29] M. Karimi, R. Amirfattahi, S. Sadri, S.A. Marvasti, Non-invasive detection and classification of coronary artery occlusions using wavelet analysis of heart sounds with neural networks, in: 3rd IEE International Seminar on Medical Applications of, Signal Processing, 2005, pp. 117–120.
[30] W. Kim, S. Jin, Y. Park, H. Choi, A study on development of multi-parametric measure of heart rate variability diagnosing cardiovascular disease, IFMBE Proc. 14 (2007) 3480–3483.
[31] M. Kobayashi, T. Musha, 1/f fluctuations of heart beat period, IEEE Trans. Biomed. Eng. 29 (1982) 456–457.
[32] T.T. Laitio, H.V. Huikuri, T.H. Makikallio, J. Jalonen, E.S.H. Kentala, H. Helenius, O. Pullisaar, J. Hartiala, H. Seheinin, The breakdown of fractal heart rate dynamics predicts prolonged postoperative myocardial ischemia, Anesth. Analg. 98 (2004) 1239–1244.
[33] H.G. Lee, K.Y. Noh, K.H. Ryu, Mining biosignal data: coronary artery disease diagnosis using linear and nonlinear features of HRV, PAKDD Workshops 2007 (2007) 218–228.
[34] H.G. Lee, K.Y. Noh, K.H. Ryu, A data mining approach for coronary heart disease prediction using HRV features and carotid arterial wall thickness, Int. Conf. Biomed. Eng. Inform. 1 (2008) 200–206.
[35] T.H. Makikallio, T. Seppanen, M. Niemela, K.E.J. Airaksinen, J.M. Koistinen, M. Tulppo, H.V. Huikuri, Abnormalities in beat to be at complexity of heart rate dynamics in patients with a previous myocardial infarction, J. Am. Coll. Cardiol. 28 (1996) 1005–1011.
[36] T.H. Makikallio, T. Seppanen, K.E.J. Airaksinen, J.M. Koistinen, M.P. Tulppo, C.K. Peng, A.L. Goldberger, H.V. Huikuri, Dynamic analysis of heart rate may predict subsequent ventricular tachycardia after myocardial infarction, Am. J. Cardiol. 80 (1997) 779–783.
[37] G. Manis, S. Nikolopoulos, A. Alexandridi, C. Davos, Assessment of the classification capability of prediction and approximation methods for HRV analysis, Comput. Biol. Med. 37 (2007) 642–654.
[38] Ping-Feng Pai, Hsu Ming-Fu, Wang Ming-Chieh, A support vector machine-based model for detecting top management fraud, Knowl.-Based Syst. 24 (2011) 314–321.
[39] C.K. Peng, S. Havlin, J.M. Hausdorf, J.E. Mietus, H.E. Stanley, A.L. Goldberger, Fractal mechanisms and heart rate dynamics, J. Electrocardiol. 28 (1996) 59–64.
[40] P.B. Persson, C.D. Wagner, General principles of chaotic dynamics, Cardiovasc. Res. 31 (1996) 332–341.
[41] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. USA 88 (1991) 2297–2301.
[42] K. Qtsuka, G.F. Cornélissen, F. Halberg, Circadian rhythmic fractal scaling of heart rate variability in health and coronary artery disease, Clin. Cardiol. 20 (2009) 631–638.
[43] D. Reynolds, Gaussian Mixture Models, MIT Lincoln Laboratory, 1991.
[44] L.I. Smith, A Tutorial on Principal Components Analysis. Information. <http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf> (accessed 24.02.12).
[45] A. Subasi, M.I. Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, Exp. Syst. Appl. 37 (2010) 8659–8666.
[46] Z. Zhao, C. Ma, An intelligent system for noninvasive diagnosis of coronary artery disease with EMD-TEO and BP neural network, International Workshop on Education Technology and Training and International Workshop on Geoscience and Remote Sensing 2 (2008) 631–635.