Expert Systems With Applications 56 (2016) 268–281

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Imposed target based modification of Taguchi method for feature optimisation with application in arrhythmia beat detection

Uday Maji[a,*], Madhuchhanda Mitra[b], Saurabh Pal[b]

[a] Department of Applied Electronics and Instrumentation Engineering, Haldia Institute of Technology, Haldia, India
[b] Department of Applied Physics, University of Calcutta, Calcutta, India

ARTICLE INFO

Article history:
Received 16 October 2015
Revised 13 March 2016
Accepted 14 March 2016
Available online 17 March 2016

Keywords:
ECG
Anchor points
Taguchi
Loss function
Impact score
Bayesian

ABSTRACT

Development of an expert system for clinical application includes automation in diagnosis of abnormality and patient monitoring based on features derived from continuous data set. This paper presents a novel method for feature optimization and classification of electrocardiogram (ECG) for arrhythmia analysis. A feature set optimization technique can reduce the classification hazard by selecting few comprehensive features to cater all kind of abnormalities under consideration. Proposed work deals with ranking and selection of an optimized pair of features using Taguchi method from eleven possible features normally used for characterizing arrhythmic beats like left bundle branch (LBBB), right bundle branch (RBBB) and premature ventricular contraction (PVC) are compared to normal beats. An imposed target based modification of Taguchi method is also suggested for the systems where the output is not pre-defined as in the case of biomedical applications. The proposed method is advantageous for the expert systems in which individual identity of the features are to be stored while reducing the dimensionality of the feature set. Multiclass Navis Bayes classifier is used to classify the beats in a single run and good performance parameters are obtained as reported in the result section.

1. Introduction

Development of a knowledge based expert system for biomedical signal processing and analysis is a promising area of research in patient monitoring and healthcare technology domain. Biomedical data (ECG, EEG, PPG, EMG, respiration etc.) are acquired mostly by contact type sensors that perform continuous recording of analog clinical information. Digitization makes the dataset a huge one whose length depends on sampling frequency and the duration of the test signal. Ideal biomedical expert system for diagnostic assistance should have some inherent quality of high performance, fast processing and decision making, accuracy in classification and reliability. Database catered to the expert system should be sufficiently large to predict an event with confirmation. Moreover, data should be adequate to cover all types of clinical possibilities that may occur in reality. Apart from that, biosignals possess a number of important features clinically relevant for diagnosis and personalized monitoring. Hence huge amount of data and features may affect the response time and final decision of a knowledge based expert system. Sometimes they lead to misinterpretation if the selected features are not suitable for the particular classification. Thus optimization of features set is an important step for automatic classification of pathological events. In the present study, a knowledge based feature optimization and automatic classification system is proposed for cardiac activity monitoring. Proposed system uses a feature optimization tool based on Taguchi Method (Taguchi, 1995) to handle a number of features derived from ECG followed by a conventional Bayesian multiclass classifier (Karraz & Magenes, 2006; Russell & Norvig, 1995) to detect arrhythmic events occurring in ventricle. The study is made on arrhythmic condition of heart.

Cardiac arrhythmia (irregular heartbeats) is such a disorder in which the electrical activity of the heart becomes irregular. This irregularity of heartbeat may lead to sudden cardiac arrest (SCA) that accounts large number of deaths in different countries (Li, Bisera, Weil, & Tang, 2012). Fast and automatic diagnosis of cardiac arrhythmia at the initial stage may increase the probability of survival from a SCA incident. For last few decades enormous study has been done on authentic detection of arrhythmic changes, but development of a reliable algorithm is still an open challenge. In this paper three ventricular arrhythmia viz. LBBB, RBBB and PVC are characterized and detected against normal rhythm.

The muscle mass of the atria is small compared to that of the ventricles, and the electrical change accompanying the contraction

* Corresponding author: Tel.: +919143098701.
E-mail addresses: udaymaji@gmail.com (U. Maji), mmitra@ieee.org (M. Mitra), spal76@gmail.com (S. Pal).

http://dx.doi.org/10.1016/j.eswa.2016.03.026
0957-4174/

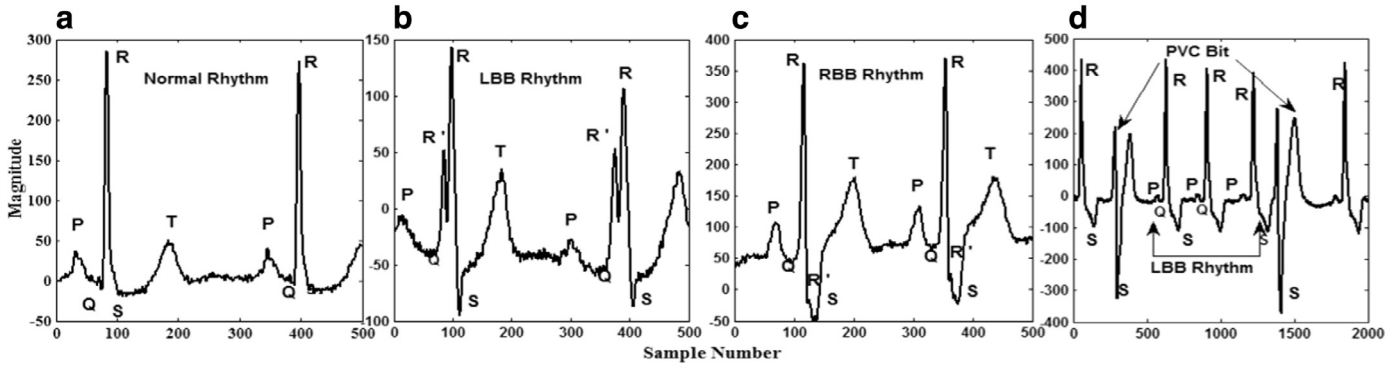

Fig. 1. Different ventricular rhythms.

of the atria is therefore small. Contraction of atria is associated with the ECG wave called 'P'. The ventricular mass is large, and so there is a large deflection of the ECG when the ventricles are depolarized. This is called the 'QRS' complex. The 'T' wave of the ECG is associated with the return of the ventricular mass to its resting electrical potential (Hampton, 2003). A typical ECG signal comprising of these waves is shown in Fig. 1a. However, if there is any abnormality in the conduction of electrical pulse generated from Sino-atrial node in either bundle branch (left and right), depolarization is delayed. Block in either bundle branch reduces the conduction through Bundle of his may cause complete heart block at extreme condition. RBBB (Fig. 1c) indicates the problem in the right side of heart and is visualised by wide QRS complex, additional R′ peak and deep S-wave in the ECG signal. Similarly LBBB arises due to similar problem in left side of heart and is observed as wide QRS complex, additional R' peak (Fig. 1b) and occasional inverted T wave (Luthra, 2012). LBB is reported as the cause of about 25% of the heart failure (HF) population (Kass, 2005). PVC rhythm (Fig. 1d) generally arises when heartbeat is initiated by Purkinje fibres in the ventricles (Sajjan, 2014). It may be observed as skipped heartbeat and appears like suction in chest. In this abnormality the QRS and T waves look very different from normal rhythm (Fig. 1a). PVC rhythms are erratic and R to R interval for a PVC cycle and its earlier cardiac cycle is smaller than usual whereas it is large for the next cycle. PVC may also cause fatigue, chest pain, a faint feeling or hyperventilation (Sajjan, 2014). So detection of arrhythmia at the earlier stage may prevent serious heart failure.

Different types of algorithms are proposed by researchers for automatic detection of arrhythmia based on the time domain or spectral analysis based features. Detection accuracy depends on obtained features and classifier type. More number of features provides better interpretation of signals which in turn leads to better classification accuracy and robustness in general. But classifiers with more number of features are mathematically cumbersome and are inherently sluggish in response. In views of this, researchers have proposed features minimization techniques prior to classification.

Almost all popular optimization techniques are tried by different researchers for ECG feature set minimization. For example, Linear Discriminant Analysis (LDA) technique is used in (Afshin et al. 2014) to find out optimal direction along which obtained features are most descriptive. This process only can reduce the dimensionality of the features but does not panel the features according to their importance, hence cannot reduce the classification complexity. Another popular feature reduction technique is Principal Components Analysis (PCA) for optimized feature set selection (Ghorbanian, Ghaffari, Jalali1, & Nataraj, 2010; Langley, Bowers, & Murray, 2010). This method basically uses large dimensional data vector to generate principal components to be employed for classification. But PCA has an inherent disadvantage that it is difficult to capture the invariance unless the training data provides the same. This may lead to inconsistent result especially in physiological signal analysis. SVM-based feature selection method (Liu & Tang, 2014) is also proposed for the mass classification of mammograms. This is basically a recurring procedure which updates the feature rank after calculating the feature weight from classifier training data, hence the process may consume huge time.

In another approach (Martis, Acharya, & Min, 2013), PCA, LDA and Independent Component Analysis (ICA) were independently applied on Discrete Wavelet Transform (DWT) coefficients for dimensionality reduction of features. Wavelet based analysis has an inherent disadvantage regarding the choice of mother wavelet and selection of level of decomposition. Different decomposition level provides different set of coefficients leading to dynamic variation of classification performance. This study also does not suggest any accuracy or other standard classification parameters for detection of different abnormal beats.

A very interesting work is reported in (Korurek & Dogăn, 2010) where Particle Swarm Optimization based ECG beat classification method is proposed with redial basis function neural network (RBFNN) which is proved to be better than K means RBFNN. This method uses PVC beat, fused Ventricular and normal beat, atrial premature beat, RBBB and fusion of normal and paced beat along with the normal one. Only four features are considered here for characterizing proposed abnormalities. PSO is used to generate the optimized RBFNN classifier. Thus this work discusses classifier development mechanism tan feature optimization. Moreover, classification of individual class is made and final classification accuracy is presented as the average of all binary classification results.

A filter-type feature selection technique with 13 features is proposed (Alonso-Atienza, Morgado, Fernández-Martínez, García-Alberola, & José Rojo-Á lvarez, 2014) to analyze the relevance of the computed parameters. This study uses temporal, spectral and signal complexity based parameters and addresses the classification problem as a pair of binary classification task with shockable vs. nonshockable and ventricular vs. non ventricular arrhythmias. But as morphological changes of biological signals signify a variety of pathological conditions, binary classification cannot cater all possibilities. A novel Genetic algorithm (GA) based features selection technique is proposed in (Li, Rajagopalan, & Clifford, 2014) to find optimum combination of features from 14 variables. In this study features are ranked only to characterize the ventricular fibrillation and tachycardia, moreover, best classification result is obtained when three features are combined.

Very recently Multi-objective Genetic Programming-Based Feature Selection technique is reported (Nag & Pal, 2016). Being a continuous learning based method, this may not perform well