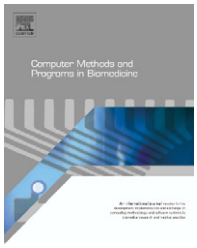
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 95 (2009) 270–279

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Discrimination of myocardial infarction stages by subjective feature extraction

Dingfei Ge[a,*], *Lihui Sun*[a], *Jiayin Zhou*[b], *Yuquan Shao*[c]

[a] *School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China*
[b] *Biomedical Engineering Research Centre, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore*
[c] *Sir Run Run Shaw Hospital, Affiliated Hospital of Medical College, Zhejiang University, Hangzhou 310016, China*

ARTICLE INFO

Article history:
Received 12 July 2008
Received in revised form
6 December 2008
Accepted 22 March 2009

Keywords:
Myocardial infarction
ECG
Feature extraction
Subjective-classification
Hyper-dimensional time-series

ABSTRACT

Lots of studies on myocardial infarction (MI) computer assisted diagnosis are based on certain important ECG components which only account for local information. 12-Lead ECG signals which were regarded as hyper-dimensional time-series data were utilized to extract features from global information in this study. Existing feature extraction techniques for classification attempt to classify all the classes included. However sometimes it is more important to better recognize certain specific classes rather than to discriminate all the classes. A feature extraction method based on subjective-classification was proposed to discriminate the specific classes, which the classification priority was given subjectively, and each of the other classes was separated at the same time. The method includes data reduction by principal component analysis (PCA), data normalization by whitening transformation and derivation of projecting vectors for subjective-classification, etc. The data in the analysis were collected from PTB diagnostic ECG database. The results show that the proposed method can obtain a small number of effective features from 12-lead ECGs to better classify classes with priority, and the other classes can be classified at the same time.

1. Introduction

Automatic detection and classification of abnormalities in electrocardiogram (ECG) will be of great help in medical examinations or monitoring of critical ill patients. Lots of existing studies on myocardial infarction (MI) computer assisted diagnosis are based on certain important ECG components which only account for local information, such as VCG and ECG based MI detection [1], Neural Network aided MI diagnosis [2], time-domain threshold methods for MI classification [3], detecting MI using ECG parameters [4], Hermite expansions [5], multiple logistic regression method [6]. Actually, clinical physicians diagnose patient's diseases are usually based on 12-lead ECG and entire ECG cycle. In order to obtain global information, 12-lead ECG and its entire ECG cycle were used for the analysis in current study. The key problem of the classification is how to obtain essential information from the hyper-dimensional data. Feature selection and feature extraction are two major approaches to dimensionality reduction. Feature selection seems to lose some of significant information for the classification due to selecting a subset of the original input variables in the measurement space, while feature extraction involves a transformation of the original variables and the features provided are a set of new variables in the transformed space. Generally, feature extraction provides more efficient representation of patterns [7]. Thus, feature extraction technique for dimensionality reduction is adopted in this study. The ECG

* *Corresponding author*. Tel.: +86 13186950314.
E-mail address: gedingfei@hotmail.com (D. Ge).
0169-2607/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2009.03.008

patterns were represented by hyper-dimensional time-series data.

On the other hand, although it is always important to discriminate as much as possible classes with as good, as possible accuracy, sometimes it is more important to better recognize certain important classes rather than to classify all the classes. In many cases, we are more interested in the classification of a particular set of classes, not the all classes.

Myocardial infarction (MI) may happen even without any symptoms. Typical MI ECG can evolve through 3 stages including MI in acute stage (MIAS, only lasts several minutes to several hours), MI in subacute stage (MISAS, can last days) and MI in recovery stage (MIRS). If patients can obtain the thrombolytic therapy during MIAS, the heart tissue that is close to death can be saved and continue to survive, otherwise the irreversible heart tissue damage will occur and sometimes even lead to inevitable sudden death in later stages. The various stages of MI are directly related to the size of the damaged myocardium supplied by the culprit coronary artery distal to the occlusion [2,5,8]. The outcome of early prognosis depends on the size of the damaged myocardium [9]. Consequently the detection of MIAS is more important and significant than that of MISAS and MIRS for patients suffering from coronary heart disease (CHA) in their daily life. In hospitals, besides MIAS, physicians also need to recognize MISAS for patient's urgent treatment and management. In this case, next highest priority is given to the detection of MISAS. In clinical practice, physicians regard the classes which have highest probability belonging to a class as the most important one or interesting one according to their experiences.

However, conventional feature extraction methods are not always efficient for the classification with priority. For examples, discriminative function pruning analysis (DFPA) [7], information entropy-based method, linear discriminant analysis (LDA) [10], and discriminant based on coefficient of variance (DCV) technique [11], etc., extract features from the viewpoint of the best average separability among the classes. These techniques attempt to extract features for the classification for all classes included.

In current study, subjective-classification based feature extraction method was introduced to distinguish prior class from other classes. The basic idea of the method is: the feature extraction tends to classify the prior class for the certain intentions that we have in advance. The projecting vectors (also called weighting factors) are set subjectively following the procedures that maximize class separability between the prior class and its nearest class. The appropriate features for the prescribed purpose are extracted from hyper-dimensional ECG time-series data by projecting each pattern onto the projecting vectors. Each dimension of ECG patterns is weighted and fused according to the weighting factor. An important existing fact behind the subjective-classification method is that classification accuracy given by programmes is often lower than that given by cardiologists, almost all the annotations of existing public ECG databases for various research purposes are done by cardiologists.

During the past years, the high-resolution ECG (1000 Hz or even higher) has been widely employed for the purpose of diagnosis in clinical practice and computerized analysis of ECG [9,12]. In this study, high-resolution ECG with a frequency of 1000 Hz was utilized. The normal ECG does not exclude myocardial necrosis, especially when the necrosis of microinfarction size and detectable only with the sensitive biomakes [13]. Although ECG beats of patients with ischemic heart disease (IHD) sometimes appear normal like people who are under health control (HC), the non-ischemic ECG beats that are collected from the patients with IHD and appear normal are also called HC in this research. Thus, the analysis ECG data include HC, MIAS, MISAS and MIRS for the discrimination of MI stages. Like many other studies, MI groups include the cases with ST-elevation (STEMI) only [14–16]. The discrimination of MI stages by proposed subjective feature extraction can be embedded into some possible implementation systems, an example of such a system is shown in Fig. 1, where ST segment deviation amplitude is measured prior to subjective feature extraction using the proposed method. It is well known that severe ST depression in 12-lead ECG may occur in unstable angina (UA), non-Q-wave MI (NQMI, true non-STEMI), or as reciprocal changes in STEMI not covered by 12-lead ECG [17]. The system can be refined and used to monitor patients in hospital and remote site in telemedicine system [15,18]. If the proposed method is embedded into telemedicine systems, we can make the systems more robust so as to make a selection whether continuously send the normal ECGs to hospital center according to the patient's preliminary information, such a system is of great benefit to improve the transfer efficiency. However, in this study, we focus on the discrimination of MI stages using subjective feature extraction methods. The proposed subjective-classification feature extraction method could extract not only global features from the ECGs with entire cycles and 12 leads, but also the features to better classify the prior classes from other classes, and separate each of the classes at the same time.

LDA is one of the feature extraction methods that are from the viewpoint of the best average separability among the classes, and provides pattern representation for the classification with maximum class separability based on within-class and between-class scatter matrix. It is probably the best-known approach to supervised dimensionality reduction [7]. Therefore, the performance of the proposed method will be compared with that of LDA in this research.

In the rest of the paper, we first introduce the complete method in the order of: data preprocessing, data interpretation, feature extraction and classification. Experimental results and discussions are presented in Sections 3 and 4, respectively. Finally, the conclusions for the whole study are given in Section 5.

2. Methods

2.1. *Experimental data and preprocessing*

The standard ECG signals used in clinic consist of 12 leads including leads I, II, III, aVR, aVL, aVF, V_1, V_2, V_3, V_4, V_5 and V_6 [19]. As shown in Table 1, the standard 12-lead MI ECG signals for the analysis were collected from PTB diagnostic ECG database which is the latest public ECG database containing 549 records from 294 subjects. The data for the analysis were selected from (1) subjects under healthy control and (2)