MYOCARDIAL INFARCTION CLASSIFICATION USING POLYNOMIAL APPROXIMATION AND PRINCIPAL COMPONENT ANALYSIS

P.-C. Chang
Department of Information Management
Yuan Ze University
Chung-Li, 32026, Taiwan, R.O.C.
iepchang@saturn.yzu.edu.tw

J.-J. LIN
Department of Information Management
Yuan Ze University
Chung-Li, 32026, Taiwan, R.O.C.
s969203@mail.yzu.edu.tw

Y.-C. WU
Department of Information Management
Yuan Ze University
Chung-Li, 32026, Taiwan, R.O.C.
angus751203@gmail.com

ABSTRACT
A rapid and accurate diagnosis in patients with acute myocardial infarction is vital, since expeditious reperfusion therapy can improve prognosis in most patients. Myocardial infarction occurred when the blood supply to part of the heart was interrupted. In ECG monitoring, ST segment means the change of electric potential in the period which from the end of ventricular depolarization to the origin of repolarization, therefore, in myocardial infarction, ST segment change is the only criterion by which a doctor can determine if myocardial ischemia has occurred in the heart. This study combines the advantages of polynomial approximation and principal component analysis. Polynomial approximation can describe the morphology of ST segment change by coefficients and principal component analysis can generate the significant features for classification. The accuracy achieves 98.07% and the proposed approach is stable for the 12-lead ECG collected from PTB database.
KEY WORDS
12-lead ECG, Myocardial Infarction, Principal Component Analysis, Polynomial Approximation, Support vector machine.

1 INTRODUCTION

In clinical medicine, Electrocardiogram (ECG) is one of the most widely used non-invasive diagnostic tools for cardiopulmonary diseases. It is a recording of the surface potential created by the electrophysiological processes of the cardiac cycle and used diagnostically by cardiologists and general practitioners. The basic components of an ECG complex are P wave, which represents atrial depolarization, QRS complex, which represents ventricular depolarization and T wave corresponding to the period of ventricular repolarization. One normal cardiac cycle is started at the sinus node with the depolarization of the right atrium and spreads toward the entire atria in a well-ordered manner. Atrial depolarization defines the P-wave in the ECG. In the next, the depolarization impulse reaches the ventricles and the fast contraction in there produces the QRS complex of the ECG. Finally, ventricular repolarization makes the T-wave complex and the cardiac cycle of one heart beat is terminated [1].

Clinical 12-lead ECG data is now available in most hospitals and it includes more detailed information about cardiac disease. The standard 12-lead ECG is composed of six leads called the limb leads, which corresponds to the subject's four extremities and the central terminal is the average of the potentials from the limb leads, and six horizontal leads, which are also called chest leads [2]. Those leads offer 12 different angles for visualizing the activities of the heart and are named Lead I, II, III, aVL, aVF, aVR, V1, V2, V3, V4, V5 and V6, respectively.

The clinical value for Myocardial Infarction (MI) of 12-lead ECG were discovered by Van 'T Hof et al. [3]. A rapid and accurate diagnosis in patients with acute myocardial infarction is vital, since expeditious reperfusion therapy can improve prognosis in most patients. Myocardial infarction occurred when the blood supply to part of the heart was interrupted. This is mostly due to occlusion of a coronary artery following the rupture of a vulnerable atherosclerotic plaque.

The key in treating ECG complex is using the morphology in time detection and it results in concentrating researches about ECG process analysis in last decade [4]. In ECG monitoring, ST segment means the change of electric potential in the period which from the end of ventricular depolarization to the origin of repolarization, therefore, in myocardial infarction, ST change is the only criterion by which a doctor can determine if myocardial ischemia has occurred in the heart. In normal conditions, ST segment shows in horizon line but if some heart disease occurs, the ST segment may show in various forms as shown in Fig. 1 [5].

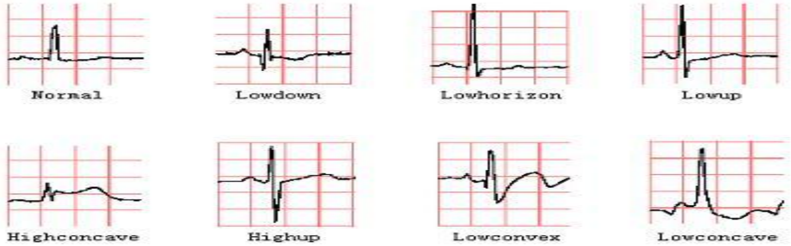

Fig. 1. Different kinds of ST waveform [5].

It is very important to identify a myocardial infarction from a patient's 12-lead ECG data in the early stage. A medical doctor can suggest a patient for expeditious reperfusion therapy if she or he is identified and the therapy can improve prognosis significantly. However, owing to the complexity and high dimensionality of the 12-lead ECG data, it is not a trivial task to accurately identify the classification of the myocardial infarction and normal data. Therefore, this study proposes a hybrid approach including polynomial approximation and Principal Component Analysis (PCA) to deal with the challenge in this field. The whole idea is based on studying the effect on extract the features in ECG data by analyzing the morphological characteristics. In next section, this paper will mention about the related literatures about how to apply these two approaches on ST segment in ECG complex and the following sections will explain the overall flow in detail.

2. LITERATURE REVIEW

A. ST Shape Change Classification by Polynomial Approximation

Because of the various shapes in ST segment, some researches try to use polynomial approximation method to extract the features in ECG. Jeong et al. [6] uses polynomial approximation to analysis ST shape change. The important parameter of polynomial approximation is the number of order within polynomial formula, e.g., 9th order of one polynomial formula and 5th order for three polynomials. The results only concern about the relative shape change of ST segment but it can be an approach to describe the variation of ST shape and give the features from coefficient of polynomials.

Because of the ST-shape change is a very important parameter for a cardiologist to diagnosis MI disease. This study uses polynomial approximation to gather the features of ST segment. For taking one MI heartbeat in a lead II ECG as an example shown in Fig. 2, one polynomial with 5th order is adopted because that it is enough to describe the variation of ST shape.

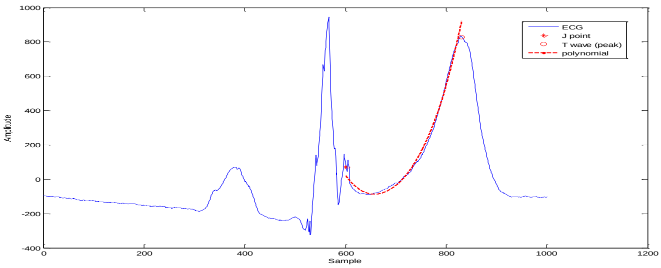

Fig. 2. The selected ST-segment and its related polynomial approximation.

In this approach, the coefficients of polynomial mean the feature vector of the ST segment in each heartbeat. Hence the feature vector for the polynomial approximation (dash red line) shown in Fig. 2 can be:

$$f(t) = \sum_{i=1}^{n} a_i \cdot t^{n-i} + b \tag{1}$$

where n is the number of orders.

Since each heartbeat has 12-Lead ECG data, therefore the number total features of a heartbeat complex is 12 x (n+1).

B. ECG Feature Extraction by Principle Component Analysis

Principal component analysis is extensively used in feature extraction to reduce the dimensionality of the original data by a linear transformation. PCA extracts dominant features, or so called principal components, from a set of multivariate data. The dominant features retain most of the information, both in the sense of maximum variance of the features and in the sense of minimum reconstruction error.

PCA in ECG signals processing takes its starting point from the samples of a segment located in some suitable part of the heartbeat. PCA utilize a representation of the data in a statistical domain rather than a time or frequency domain. PCA is often used to reduce dimensionality and for feature extraction. In ECG signals, the information can be also separated from the noise or baseline shift by PCA analysis [7].

Some researchers also use PCA to analysis multi-lead ECG. A common way to convert the multi-lead ECG into suitable data is to concatenate it. Ge et al. [8] proposed the research to concatenate the 12-lead in the order: Lead I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6. The dimension of the data will be higher than single-lead but the information of 12-lead can be solved by PCA at one time. The researches provide the evidence that PCA is useful in feature extraction stage for classification.

3. METHODOLOGY

This paper uses two approaches to extract features for each heart-beat in order to increase the accuracy for MI classification. One is concatenated ST-T segments in 12-lead ECG and then using PCA to extract the features. Another one is applying PCA on coefficients of polynomial approximation. Fig. 3 shows the overall flowchart in this study.

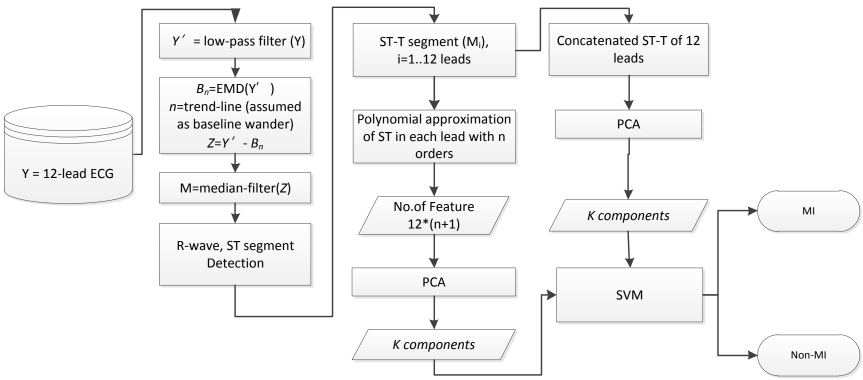

Fig. 3. The proposed framework

In the stage of pre-processing, this study adopts a simple way to decompress the effect of noise and baseline drift. First, a low-pass filter is applied. The threshold of frequency is set as 40. Then each lead will be separated into several signals by Empirical Mode Decomposition (EMD). EMD is proposed by Huang et al. [9] as a tool to decompose a signal into a collection of components. It is especially suited for nonlinear and nonstationary signals.

The result of EMD produces *n* intrinsic mode functions (IMFs) and a residue signal. The residue signal can be regarded as a trend line in original signal. Therefore, the residue signal is assumed as the baseline wander in this research due to the low frequency. After subtract the assumed baseline wander, a median-filter is also used to make the signals more stable and clear.

After filtering the 12-lead ECG, QRS-wave location is necessary for ST-T segment identification and heartbeat isolation. This study gives the result of the detected result of lead-II ECG complex in 5000 points and shows in Fig. 4. The circles point out the start position of ST-T segment and the end position is labeled as cross point. The triangle means where the R-wave is. As shown, T wave is considered in this study because the morphology or the direction of the peak also provides essential information for classification.

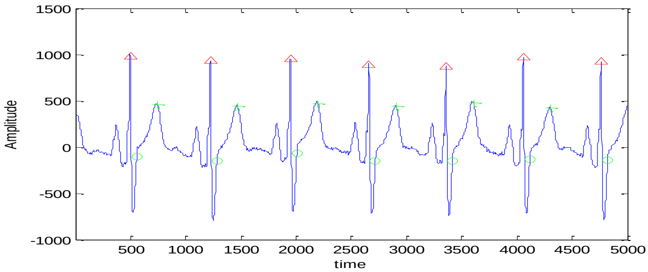

Fig. 4. The detected R-waves and ST-T segments in filtered ECG

This study focuses on analysis a single heartbeat belongs to MI or not. There are two feature extraction strategies in this research. Both of them use PCA to reduce and gather significant features. This research compares the accuracy of original (but filtered) ECG data and the performance of polynomial approximation. For first strategy, To combine the information in 12-lead ECG data, the segmented ECG data from corresponding 12 lead ECG are concatenated in the order as [8]. For second strategy, PCA is applied on the collection of coefficients from polynomial approximation of 12 lead ECG. The significant information will be selected to train the classifier. Here SVM [10] is selected to validate the relationship between extracted features and MI.

4. EXPERIMENTAL RESULT

In experimental result, testing strategy adopted in this research is four-fold cross-validation with repeating ten times. The performance measurement considers about accuracy, sensitivity (SE), specificity (SP) and positive predictive (PP). In comparison, the parameter of the number of order for polynomial is also verified. In Table II, poly(n) means n orders of polynomial approximation. The last row including ST+PCA is the first strategy with concatenated ST-T segments of 12-lead ECG.

The 12-lead ECG data are collected from PTB-database. There are total 549 cases downloaded, including 148 cases with MI and 401 non-MI cases. The classifier used here is Support Vector Machine and the kernel function selected is RBF kernel with sigma value is 2. The number of input features selected from PCA is reduced to 12 for both strategies according to the variances in experimental results and past experiments.

Tab.2 The performance results.

		accuracy	SE	SP	PP
Poly3+PCA	Mean	96.79%	**98.74%**	92.42%	96.69%
Poly4+PCA	Mean	**98.07%**	98.73%	**96.60%**	**98.49%**
Poly5+PCA	Mean	97.96%	98.71%	96.26%	98.34%
ST+PCA	Mean	96.40%	97.73%	93.44%	97.10%

The rule to judge whether the test case belongs to MI is if more than three-quarters of number of heartbeats is classified as MI, then the test case is MI, otherwise, the test case will be classified as non-MI case. In Table II, polynomial approximation can give higher performance than original 12-lead ECG data when the number of orders reaches four or five. That means the coefficients can be used to as the factors for MI detection and PCA can also improve the performance result.

5. CONCLUSION

PCA and polynomial approximation are considered as different approach for features extraction from the ST segment in one heartbeat. This study finds out that these two approaches can be combined and achieve higher performance in classifying the input case belongs to MI or not. Polynomial approximation can use coefficients which generate the fitting curve of ST segment. PCA can transform the ST segment series into a number of uncorrelated variables of original 12-lead ECG or coefficients from polynomial approximation. According to the experimental results, Because of ST segment change is a very important factor of diagnosing MI disease. Using polynomial approximation can calculate the related information, e.g. height, width, shape and elevation of ST segment in mathematical result. By through PCA and polynomial approximation, the relationship between ECG and MI disease will be more precise than using PCA directly in 12 lead ECG. In the future work, the order of polynomial approximation can be extended, for example, using polynomial formulation to calculate the coefficients of concatenated ST-segmentation may give a direct way to detect MI.

REFERENCES

[1] CHANG P. C.; HSIEH J. C; LIN J. J.; CHOU Y. H.; LIU C. H. A Hybrid System with Hidden Markov Models and Gaussian Mixture Models for Myocardial Infarction Classification with 12-Lead ECGs. IEEE International Conference on High Performance Computing and Communications, 2009, art. no. 5166983, PP: 110-116

[2] GARCIA T. B.; HOLTZ N. E. Introduction to 12-Lead ECG: The Art of Interpretation, 2001

[3] Van 'T Hof A. W. J.; LIEM A.; DE BOER M. J.; ZIJLSTRA F. Clinical value of 12-lead electrocardiogram after successful reperfusion therapy for acute myocardial infarction. Lancet, 1997.350, PP: 615-619

[4] DE CHAZAL P.; O'DWYER M.; REILLY R. B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering, 2004.51, PP: 1196-1206

[5] SHEN Z.; HU C.; LIAO J.; MENG M. Q. H. An algorithm of ST segment classification and detection, IEEE International Conference on Automation and Logistic, 2010, PP: 559-564

[6] JEONG G. Y.; YU K. H.; YOON M. J.; INOOKA E. ST shape classification in ECG by constructing reference ST set. Medical Engineering and Physics , 2010,vol. 32, PP: 1025-1031

[7] CHAWLA M. P. S. PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison. Applied Soft Computing Journal, 2011.11(2), PP: 2216-2226.

[8] GE D.; SUN L.; ZHOU J.; SHAO Y. Discrimination of myocardial infarction stages by subjective feature extraction. Computer Methods and Programs in Biomedicine, 2009.95, PP: 270-279

[9] HUANG N. E.; SHEN Z.; LONG S. R.; WU M. C.; SHIN H. H.; ZHEN G Q.; YEN N. C.; TUNG C. C.; LIU H. H. The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1998.454, PP: 903-995

[10] WANG Y. W.; CHANG P. C.; FAN C. Y.; HUANG C. H. Database classification by integrating a case-based reasoning and support vector machine for induction. Journal of Circuits, Systems and Computers, 2010.19(1), PP:31-44