An Approach to Diagnose Cardiac Conditions from Electrocardiogram Signals

LU, Yan

A Thesis Submitted in Partial Fulfillment

of the Requirement for the Degree of

Master of Philosophy

in

Mechanical and Automation Engineering

The Chinese University of Hong Kong

October 2010

Abstract

Electrocardiogram (ECG) signals can provide valuable information for physicians to diagnose cardiac conditions. In recent years, automatic diagnosis based on ECG has received more and more attention, due to the rapid development of sensor, computer and information technologies. In this thesis, an ECG-based automatic diagnosis approach is proposed, which consists of components like ECG preprocessing, feature extraction and classification.

ECG preprocessing is necessary for reliable interpretation, since ECG is usually contaminated by baseline wander and various kinds of noises, such as electrode motion, muscle artifact and so on. In this study, baseline wander is removed by an empirical mode composition (EMD) based method, which could preserve ECG feature waves from distortion. For ECG denoising, a novel method is proposed based on ECG dynamic model, EMD and instantaneous frequency. Firstly, the traditional ECG model is generalized by asymmetric Gaussians, and then it's used to pre-filter noisy ECG signals. A residue signal produced by pre-filtering is further denoised using EMD algorithm with instantaneous frequency threshold. Denoising experiments conducted on ECG data contaminated by Gaussian and real noises have demonstrated the desirable performance of the proposed method.

After preprocessing, ECG classification is conducted for automatic diagnosis. This study proposes to extract features using the generalized ECG dynamic model, since the model parameters can accurately capture the ECG morphological characteristics, which are the basic evidence for clinical diagnosis. Moreover, generalized discriminant

analysis is employed to reduce the feature dimension to dig out the most significant features in a mathematical way. Support vector machine is utilized as as the classifier for the multi-category ECG classification problem.

The Physikalisch-Technische Bundesanstalt (PTB) database provides rich ECG records corresponding to various cardiac conditions, from which six classes of ECG signals are selected for study. Experiments are carried out on the PTB database and compared with other works. The results demonstrate that the proposed method produces not only accurate but also robust classification performance, and that it has the potential for practical use.

摘 要

心電（ECG）信號可為醫師診斷心臟狀況提供重要依據。近年來，由於傳感器、計算機以及資訊科技的迅速發展，基於心電圖的自動診斷已得到越來越多的關注。本文提出了一個基於心電圖的自動診斷系統，包括信號預處理、特徵提取和分類等組成部分。

心電信號通常會受到基線漂移、電極移位、肌電信號等多種噪聲的幹擾，因此，在解讀心電圖之前，信號預處理是必需的。本文中，基線漂移是通過一個基於經驗模態分解（EMD）的方法消除的，同時還避免了特稱波形的畸變。本文提出了一個新的降噪方法，綜合利用了心電圖模型、EMD 和瞬時頻率。首先，引入非對稱高斯函數對傳統的心電模型進行推廣，然後用推廣後的模型對心電信號進行預濾波，此過程產生的殘差信號進一步通過經驗模態分解和瞬時頻率閾值法降噪。實驗結果證明該方法可實現良好的降噪效果。

預處理之後，進行心電信號分類以實現自動診斷。本文提出利用推廣的心電模型進行特徵提取，因為該模型的參數可以精確地反應心電的形態特徵。此外，本文用廣義判別分析（GDA）對特徵向量進行降維以便通過數學方法來發掘出最重要的特徵。支撐向量機（SVM）被用作該多類分類問題的分類器。

系統的性能是在 PTB 心電數據庫上進行測試的。該數據庫包含大量由人工標注過的不同心臟狀況下的心電數據。實驗結果證明，本文提出的自動診斷方法具有良好的分類準確度和魯棒性，具有實際應用的潛質。

Acknowledgement

This research work would not have been possible without the help and support of many people.

I would love to extend my gratitude to my supervisor, Prof. Yeung Yam, for his guidance, enlightening discussions and consistent encouragement. Especially, I would love to thank him for all the opportunities and instructions he has been giving me for building my academic career.

I want to thank all members in the Intelligent Control Systems Laboratory for teaching me the techniques and discussing together the problems. Thanks to all of them for giving me a joyful master study life.

I also would like to thank Prof. Yunhui Liu and Prof. Wen J. Li for being my thesis committee members and providing valuable advice for my work.

Last but not least, I would express my deepest gratitude to my parents. They have always been giving me guidance on my life as well as my career and caring about my health.

Contents

List of Figures

List of Tables

1. Introduction

This chapter introduces the background for the present study and describes the motivation, related works, and our proposed approach to the problem.

1.1 Electrocardiogram

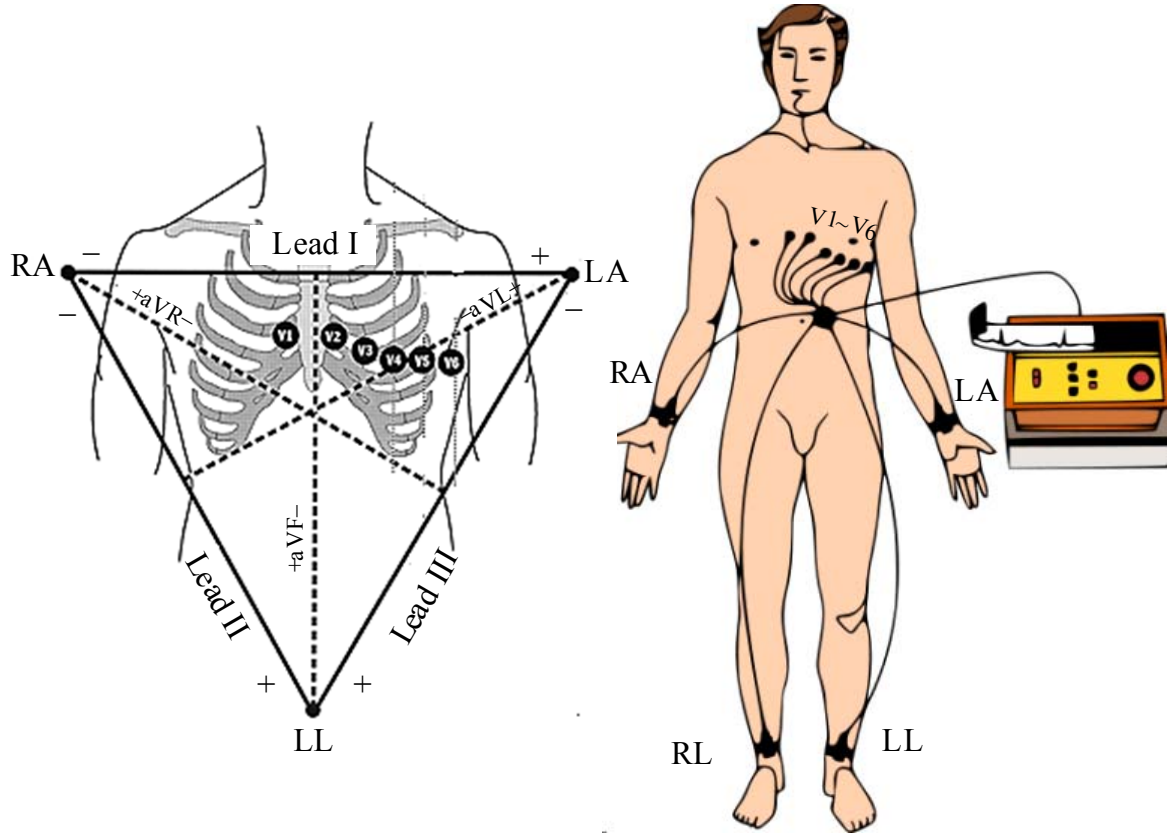
Before performing any signal processing of the Electrocardiogram (ECG), it is of importance to first understand the physiological basis of the ECG, and to review the measurement conventions of a standard ECG and how a clinician uses the ECG for patient care [1].

The heart is comprised of muscle (*myocardium*) that is rhythmically driven to contract and drives the circulation of blood throughout the body. Before every normal heartbeat, a wave of electrical current passes through the entire heart and triggers myocardial contraction. The pattern of electrical propagation is not random, but spreads over the structure of the heart in a coordinated pattern which leads to an effective, coordinated systole. This results in a measurable change in potential difference on the body surface of the subject. The resultant amplified (and filtered) signal is known as an ECG. An ECG can be affected by many factors, including abnormalities of cardiac conducting fibers, metabolic abnormalities (including a lack of oxygen, or *ischemia*) of the myocardium, and macroscopic abnormalities of the normal geometry of the heart. ECG analysis is nowadays a routine part of any complete medical evaluation, due to the heart's essential role in human health and disease, and the relative ease of recording and analyzing the ECG in a noninvasive manner.

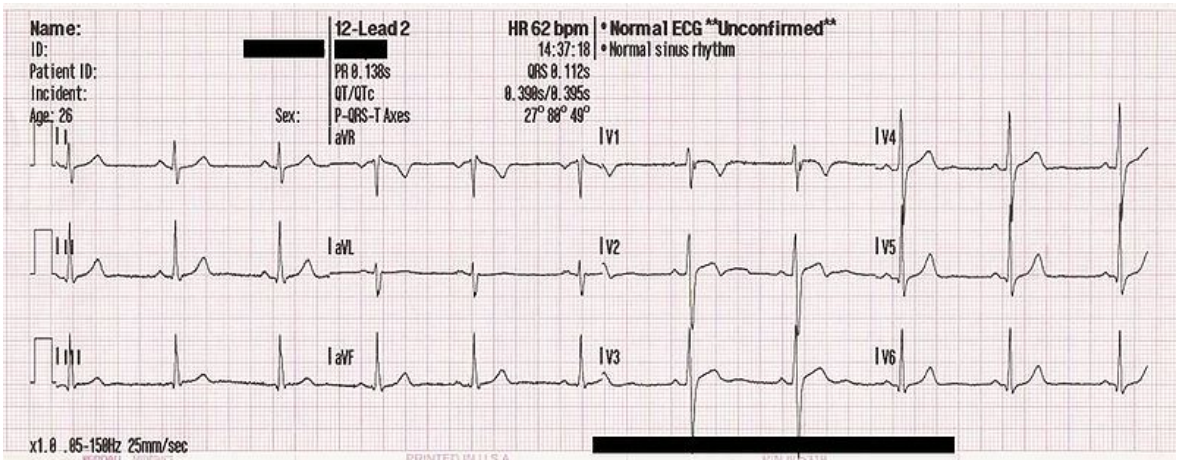
1.1.1 ECG Measurement

ECG is a transthoracic interpretation of the electrical activity of the heart over time captured and externally recorded by skin electrodes. It is a noninvasive recording produced by an electrocardiographic device. The etymology of the word is derived from *electro*, because it is related to electrical activity, *cardio*, Greek for heart, and *graph*, a Greek root meaning "to write".

The electrodes are attached to the patient's body, usually with very sticky circles of thick tape-like material (the electrode is embedded in the center of this circle), onto which cables clip. In ECG, the word *lead* refers to the voltage between two electrodes. ECG leads use different combinations of electrodes to produce various signals from the heart, and each lead will have a specific name. For example, Lead I (lead one) is the voltage between the right arm (RA) electrode and the left arm (LA) electrode, whereas Lead II (lead two) is the voltage between the RA electrode and the left leg (LL) electrode. The traditional ECG contains 12 leads: I, II, III, aVF, aVL, aVR, V1, V2, V3, V4, V5, and V6, which are obtained from 10 electrodes placed at different positions, as illustrated in Fig. 1(a). A segment of 12-lead ECG plotted on graph paper is also shown in Fig. 1 (b).



(a)



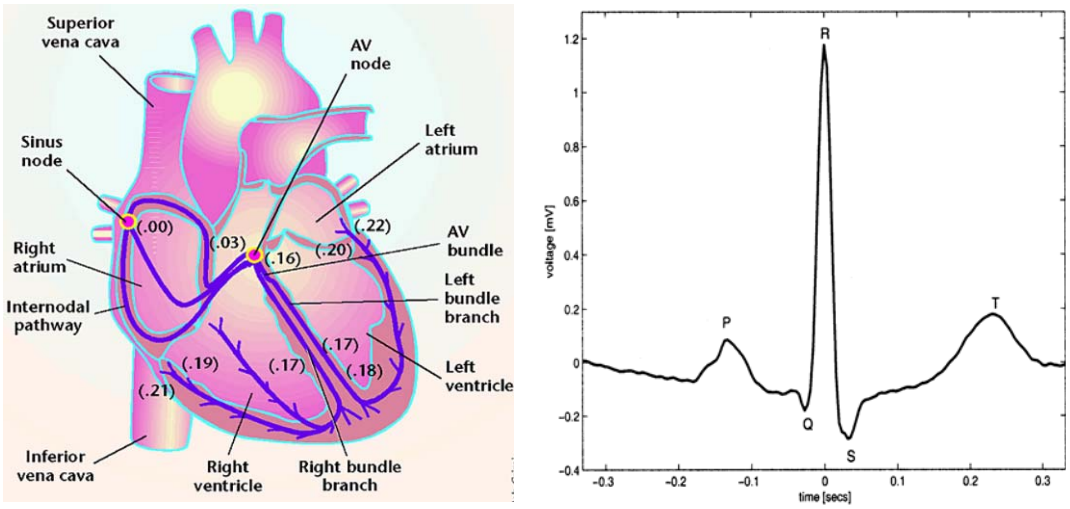
(b)

Fig. 1. ECG measurement. (a) ECG measurement by electrodes, (b) Sample of12-lead ECG

1.1.2 Cardiac Conduction Pathway and ECG Morphology

In the cardiac conduction system illustrated in Fig. 2(a), the sinus node (sometimes called the sinus-atrial node) serves as the heart's pacemaker, emitting an impulse that results in an action potential. The cells in the node have almost no contractile elements but are connected directly to the atrial fibers, so that the action potential spreads immediately into the atrial cardiomyocytes and is transmitted through the entire atrial muscle mass. However, propagation occurs more rapidly through three specialized bundles of atrial muscle called the internodal pathways, which contain, in addition to the atrial cardiomyocytes, specialized conductive cells. This has the net effect of transporting the conductive impulse to the atrial-ventricular (A-V) node within 30 ms. There is a delay of about 130 ms in the A-V node and bundle system during which time the atria can contract, filling the ventricles. The conducting impulse is then propagated through the specialized Purkinje cells of the conduction system. These are large cells that are able to transmit the action potential at 2−4 ms, a rate that is six times that of the normal ventricular cardiomyocyte. The impulse is transmitted through the entire Purkinje fiber system within 30 ms, ensuring that a rhythmic and concerted ejection of blood from the ventricles take place as the endocardial cells and finally the epicardial cells are stimulated.

Each beat of the heart can be observed as a series of deflections away from the baseline on the ECG. These deflections reflect the time evolution of electrical activity in the heart which initiates muscle contraction. A single sinus (normal) cycle of the ECG, corresponding to one heartbeat, is traditionally labeled with the letters P, Q, R, S,



(a) (b)

Fig. 2. Cardiac conduction system and ECG morphology. (a) Cardiac conduction system, (b) Morphology of a mean PQRST-complex of an ECG recorded from a normal human.

and T on each of its turning points as illustrated in Fig. 2 (b). The ECG may be divided into the following sections.

• P-wave: A small low-voltage deflection away from the baseline caused by the depolarization of the atria prior to atrial contraction as the activation (depolarization) wavefront propagates from the SA node through the atria.

• PQ-interval: The time between the beginning of atrial depolarization and the beginning of ventricular depolarization.

• QRS-complex: The largest-amplitude portion of the ECG, caused by currents generated when the ventricles depolarize prior to their contraction. Although atrial repolarization occurs before ventricular depolarization, the latter waveform (i.e. the QRS-complex) is of much greater amplitude and atrial repolarization is therefore not seen on the ECG.

• QT-interval: The time between the onset of ventricular depolarization and the end of ventricular repolarization. Clinical studies have demonstrated that the QT-interval

increases linearly as the RR-interval increases [2]. Prolonged QT-interval may be associated with delayed ventricular repolarization which may cause ventricular tachyarrhythmia leading to sudden cardiac death [3].

• ST-interval: The time between the end of S-wave and the beginning of T-wave. Significantly elevated or depressed amplitudes away from the baseline are often associated with cardiac illness.

• T-wave: Ventricular repolarization, whereby the cardiac muscle is prepared for the next cycle of the ECG.

1.1.3 A Basic Clinical Approach to ECG Analysis

In analyzing the clinical electrocardiogram, it is important to use a systematic approach. The following overview, which illustrates a clinical approach, should not be considered as completely thorough. Rather, it should be considered as simply a guide to understanding how clinicians identify abnormalities in the ECG.

1. *Identify the QRS complexes.* The following observations should be made:
   • What is the ventricular rate?
   • Are the QRS complexes spaced at regular intervals? If not, what is the nature of the irregularity?
   • Are the QRS complexes identical in shape in a given lead? Are they of normal size and morphology?

2. *Identify the P waves*. In some cases this will require careful observation, and more than one lead axis may be necessary. The following questions should be explored:
   • Is there a one-to-one relationship between P-waves and QRS complexes? If not, is there a definable pattern?

• Is the PR interval of normal duration?

• What is the atrial rate?

• Are the P waves identical in shape in a given lead? Are they of normal size and shape?

Based on the above analysis, it should be possible to identify the mechanism of the rhythm in most cases.

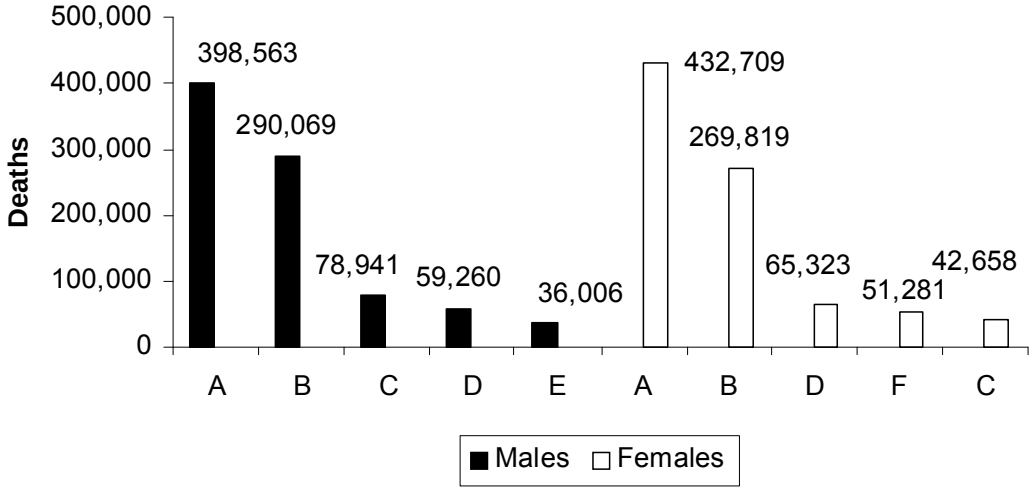
3. *Examine the QRS complex in each lead.* Is the QRS axis normal? Overall, are the QRS widths and amplitudes normal? Often, the QRS complexes are viewed in several "groups" that are specific to a particular region of the heart. The waveform patterns should also be checked for signs of intraventricular conduction block, significant amplitude Q waves and precordial R-wave pattern normality.

4. *Examine the ST-T segments*. Are there abnormalities (such as elevation or depression)? Is the abnormality suggestive of ischemia, infarction, or hypothermia?

1.2 Cardiovascular Disease

Heart disease or cardiovascular disease (CVD) is the class of diseases that involve the heart or blood vessels (arteries and veins). While the term technically refers to any disease that affects the cardiovascular system, it is usually used to refer to those related to atherosclerosis (arterial disease). These conditions have similar causes, mechanisms, and treatments. [4]

CVDs are the number one cause of death globally: more people die annually from CVDs than from any other cause. An estimated 17.1 million people died from CVDs in 2004, representing 29% of all global deaths. Of these deaths, an estimated 7.2 million were due to coronary heart disease and 5.7 million were due to stroke. By 2030, almost



A – CVD, B – Cancer, C – Accidents, D – Chronic Lower Respiratory Diseases
E – Diabetes Mellitus, F – Alzheimer's Disease

Fig. 3. CVD and other major causes of death for all males and females (United States: 2006). Source: NCHS and NHLBI.

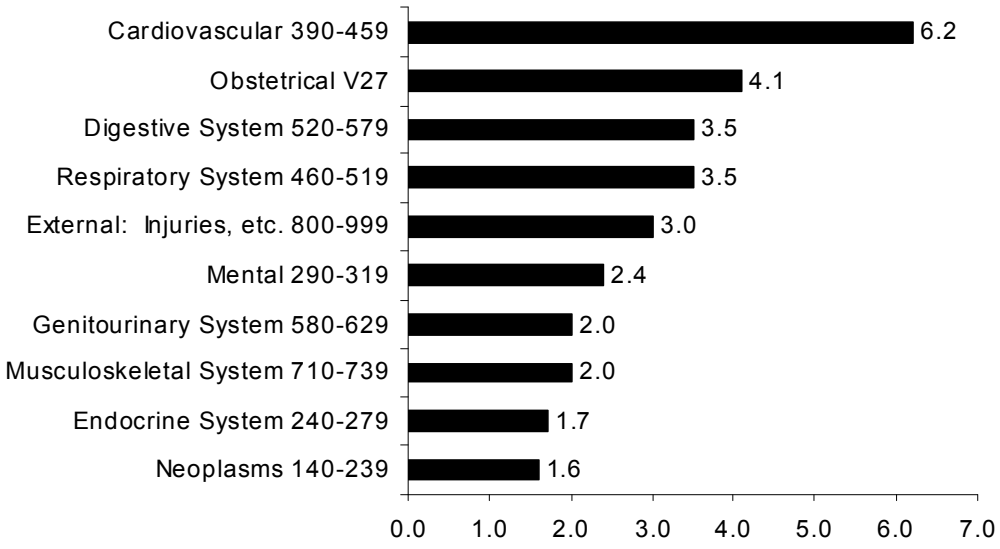


Fig. 4. Hospital discharges (in millions) for the 10 leading diagnostic groups (United States: 2006). Source: NHDS/NCHS and NHLBI.

23.6 million people will die from CVDs, mainly from heart disease and stroke. These are projected to remain the single leading causes of death [5]. Take the United States as an example. In 2006, CVDs not only killed the most people among all diseases and other causes (Fig. 3), but also cost the largest amount of money for treatments in hospital (Fig. 4).

1.3 Motivation

While CVD is such a terrible killer as described above, a positive fact is that CVDs are usually chronic diseases and can be alleviated or cured if timely treatments are provided. For the detection and diagnosis of CVDs, ECG is absolutely the most important and helpful tool, based on which physicians can thoroughly analyze the cardiac conditions of the subject and make diagnostic decisions.

As a result of the fast advancement of sensor technology nowadays, many household monitoring devices have been developed to sample and record human physiological signals like ECG in daily life. Although it is now easier to obtain ECG inside the home with the help of portable ECG monitors, common people are mostly short of expertise to correctly analyze ECG. In some cases, delay in detection of abnormal ECG patterns may lead to the worsening of heart conditions (an increasingly repetitive pattern of transient abnormalities) and may prevent a patient or doctor from seeking corrective treatments before a potential heart attack or stroke.

On the other hand, even if ECG signals can be analyzed by physicians in hospital, it is still a tough job for them to perform. The main difficulty originates from reading a long sequence of ECGs beat by beat. For an adult, a normal resting heart rate ranges from 60 to 100 beats per minute (bpm), which means that there would be thousands of

beats in a ECG signal of one hour. Since signs of heart disease may occur at any time and be transient, it would be necessary but tedious and time-consuming for physicians to read an ECG record thoroughly.

Considering the difficulties in ECG analysis discussed above, this thesis commits itself to exploring an automatic approach to diagnose typical CVDs based on ECG analysis by computer. The proposed system consists of components such as signal preprocessing, feature extraction, and pattern recognition. This system, if applied together with ECG monitors, can facilitate both physicians' ECG analysis in hospital and common people's ECG interpretation at home. The significance of this work is to provide fast detection and diagnosis advice of CVDs for physicians and patients with the help of computer, so that corrective treatments can be applied as early as possible, preventing the worsening of heart conditions.

1.4 Related Work

In recent years, several methods for automatic diagnosis based on ECG have been proposed in literature, most of which attempt to detect and classify various arrhythmias, such as premature ventricular contraction (PVC), ventricular fibrilation (VF), their atrial counterparts PAC and AF, and so forth. The existing methods include threshold control [5], self-organization map [6][7], spectral analysis technique [8], filter bank [9], Markov model [10][11], and neural network [12][13][14][15]. However, Inan *et al.* have pointed out that many of the highest classification accuracies (> 90%) have been achieved in experiments where training and testing sets overlapped significantly [14]. Therefore, they conducted a large scale experiment based on the total 40 files from

MIT-BIH database, achieving robust and accurate results, 95.16% over total beats and 96.82% over test beats.

However, compared with arrhythmias, the automatic detection and classification of other CVDs, such as myocardial infarction, valvular heart disease and so forth, are still lack of comprehensive studies. Zheng *et al.* investigated the classification of old myocardial infarction using naive Bayes, support vector machine (SVM) and random forest [16]. Choi applied wavelet packet decomposition and SVM to detect valvular heart disorders [17]. Ghosh *et al.* utilized wavelet aided SVM to diagnose 30 number of input patterns from 6 different CVDs [18].

1.5 Overview of Proposed Approach

Raw ECG signals are usually corrupted by different kinds of noise, such as baseline wander, power line interference, electromyographic (EMG) noise and so on. Thus, ECG signal preprocessing (i.e. denoising) is necessary before feature extraction and classification. Empirical mode decomposition (EMD), proposed by Huang *et al.* [19] decomposes a given signal into a set of intrinsic mode functions (IMFs). The lower-order IMFs capture fast oscillation modes while higher order ones represent slow oscillation modes, i.e., the trend of overall change caused by baseline wander in ECG denoising case. By discarding several higher-order IMFs, the baseline wander can be removed. In this study, we have also proposed an ECG denoising method based on EMD and instantaneous frequency, which can reduce high frequency noises effectively.

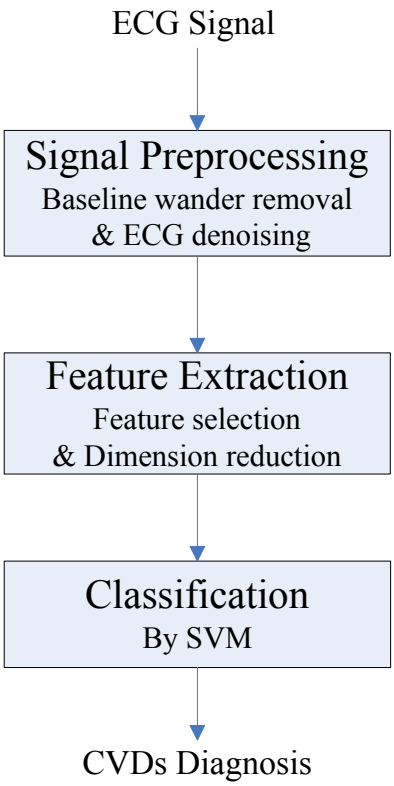


Fig. 5. System overview

Clinical experiments have demonstrated that ECG wave morphologies, including the locations, heights, widths, and other subtle features, are able to provide evidence for diagnosis of CVDs. Therefore, following the diagnostic method of physicians, this study adopts the locations, heights, and widths of ECG waves as features for pattern recognition. In order to extract these features, an ECG dynamic model first proposed by [20] is utilized which will be discussed in detail later. Then we use generalized discriminant analysis (GDA) to do feature dimension reduction. GDA firstly maps the original feature space to a larger one by producing the kernel data projection, which increases class separability of the projected training data, and then reduce the dimension to the desired number by linear discriminant analysis. Finally, support vector machine (SVM) is adopted to classify the ECG patterns. The whole system is

illustrated in Fig. 5.

1.6 Thesis Outline

This thesis is organized as follows. Chapter 2 introduces the ECG signal preprocessing, including baseline wander removal and ECG denoising, for which novel methods are proposed and experiments are conducted to verify the efficacy. Chapter 3 presents how to identify CVDs via ECG classification, including database introduction, feature extraction, dimension reduction, and classification. Finally, Chapter 4 concludes the whole study and discusses the future work.

2. ECG Signal Preprocessing

This chapter describes ECG signal preprocessing including baseline wander removal and ECG denoising. Background knowledge of ECG model and Empirical mode decomposition is introduced at first.

2.1 ECG Model and Its Generalization

2.1.1 ECG Dynamic Model

A realistic synthetic ECG generator was first proposed by McSharry *et al*. [20], using a set of 3-D state equations to produce a trajectory in the Cartesian coordinates. The dynamic equations were then transformed into the polar form for a simpler compact set by Sameni *et al*. [21]. In essence, this model describes each feature wave of one ECG cycle (e.g. P, Q, R, S and T wave) by a Gaussian with three parameters: the amplitude a_i, width b_i, and location θ_i. The vertical displacement of the ECG is determined by an ordinary differential equation

$$\dot{z}(\theta) = -\sum_{i\in\{P,Q,R,S,T\}} a_i \frac{\Delta\theta_i}{b_i^2} \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) \tag{1}$$

where $\Delta\theta_i = (\theta - \theta_i)$ and $\theta \in [0, 2\pi]$, since each cycle of ECG has been linearly mapped into the interval [0, 2π] to handle the heart rate variation.

A mathematical representation of one ECG cycle can be obtained by integrating the differential equation (1) with respect to θ

$$z(\theta) = \sum_{i \in \{P,Q,R,S,T\}} a_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) \tag{2}$$

Note that in this model, no z-offset exists since the ECG isoelectric level is assumed to be zero. It's obvious that the model approximates an ECG cycle by the sum of five Gaussians with different parameters a_i, b_i, and θ_i, describing wave amplitudes, widths and locations respectively, as illustrated in Fig 6.

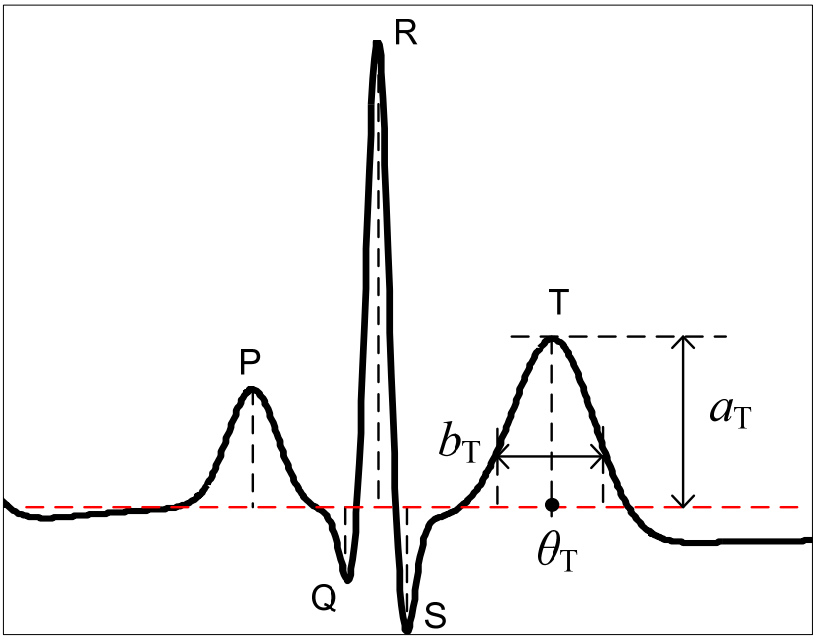


Fig. 6. ECG model with 5 Gaussians. The red dashed line denotes the isoelectric level. Note that only the parameters for T wave are illustrated. Parameters for the other waves are similar.

2.1.2 Generalization of ECG Model

If ECG feature waves (P, Q, R, S and T waves) can be assumed to have totally or nearly symmetric waveforms, then the above ECG dynamic model (EDM) can serve as a close approximation to the original signal. However, in reality, the waveforms of ECG signals are often not symmetric. Thus, it's difficult for the traditional EDM to accurately model them using symmetric Gaussians, which would further affect the

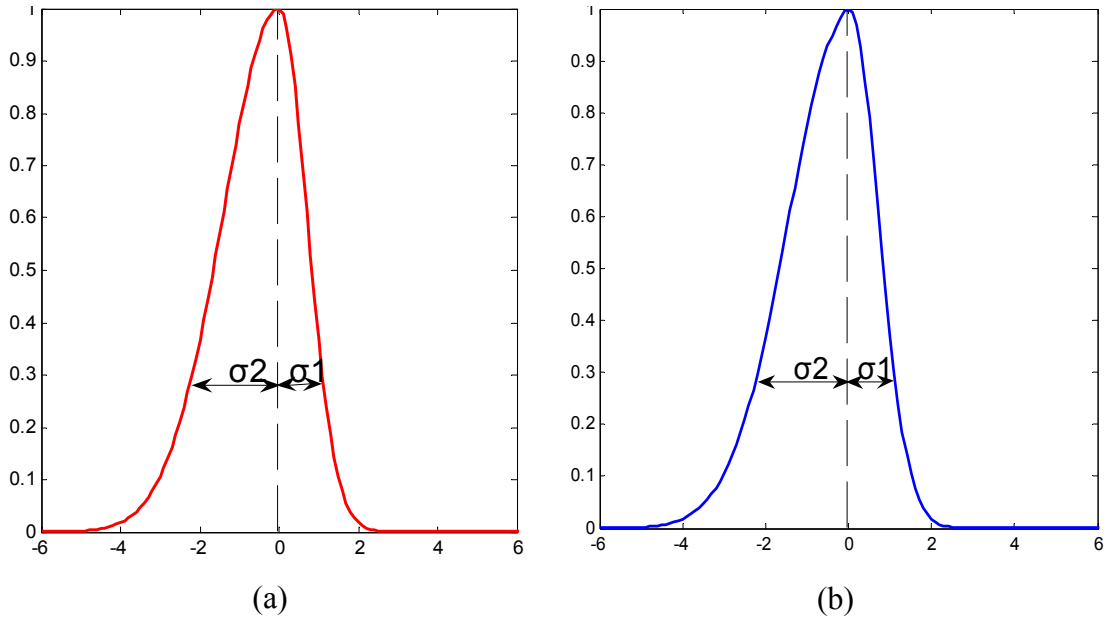
performance of model-based methods for denoising or classification. Therefore, it's necessary to generalize the EDM to represent a wider range of ECG morphologies.

As an extension of conventional Gaussian function, asymmetric Gaussian (AG) can capture spatially asymmetric distribution better [22]. Traditionally, an AG is defined as:

$$\begin{aligned} &AG(x,a,\mu,\sigma_1,\sigma_2) \\ &= a\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma_1^{\,2}}\right)\cdot\chi(x-\mu) \\ &+a\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma_2^{\,2}}\right)\cdot\left(1-\chi(x-\mu)\right) \end{aligned} \tag{3}$$

where $\chi(u)=\begin{cases}0 & u<0\\ 1 & u\geq 0\end{cases}$ is a unit step function.

As shown in Fig. 7 (a), it is the difference between σ_1 and σ_2 that determines the asymmetry degree: the larger the difference, the more asymmetric the waveform.



(a) (b)

Fig. 7. Asymmetric Gaussians: (a) with the step function, and (b) with the sigmoid function

However, the AG defined in (3) is not guaranteed to be derivable everywhere because of the step function. In order to make the AG more applicable to framework, it's necessary to make it derivable throughout. Therefore, we propose to replace the unit step function with a sigmoid function as follows:

$$\begin{aligned} & AG(x,a,\mu,\sigma_1,\sigma_2) \\ & = a\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma_1^{\,2}}\right)\cdot\psi(x-\mu) \\ & +a\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma_2^{\,2}}\right)\cdot\left(1-\psi(x-\mu)\right) \end{aligned} \tag{4}$$

where $\psi(u)=1/(1+e^{-pu})$ is a sigmoid function with a parameter p which determines the function shape: as $p\to+\infty$, the sigmoid function evolves into a step function. Assigning a large enough value to p (e.g. p = 5), the AG defined in (4) is not only derivable everywhere now, but also has almost the same shape as that in (3), as shown in Fig. 7 (b).

2.2 Empirical Mode Decomposition

The Empirical mode decomposition (EMD) algorithm was firstly proposed in the study of fluid mechanics [19], and then immediately applied to biomedical engineering [23]. In contrast to previous methods like Fourier decomposition, EMD is intuitive, a posteriori and adaptive, with the basis functions not predefined, but derived from the given data.

The essential idea of EMD is to identify the intrinsic oscillatory modes by their characteristic time scales in a signal empirically, and accordingly decompose the signal into intrinsic mode functions (IMFs) by means of a *sifting* process. Therefore, the EMD

method is especially applicable for nonlinear and non-stationary signals, including ECG.

As a counterpart to the harmonic function in Fourier analysis, the IMF represents the oscillating mode embedded in the original data. By definition, an IMF should satisfy two conditions: (1) the total number of local extrema and that of zero crossings should be equal to each other or different by at most one, which means that within two consecutive zero crossings, there can be only one wave crest or trough, no complicated waves allowed; and (2) the mean of the upper and lower envelopes respectively defined by local maxima and local minima should be zero.

By virtue of the IMF definition, the *sifting* process can be simply summarized as follows:

Given a signal $X(t)$, the first step is to find out all the local maxima and minima. After that, all the local maxima are connected by a cubic spline line as the upper envelope, and the lower envelope then comes out using the local minima in a similar way. In expectation, the two envelopes can cover all the data between them, and the their mean is designated as $m_1(t)$. So far, the first *sifting* process has been done, and the prototype of the first IMF has come forth as

$$p_1(t) = X(t) - m_1(t) \tag{5}$$

Although $p_1(t)$ should be an IMF ideally, in reality, it may still contain more than one extrema between zero crossings due to the overshoots and undershoots involved in the envelope-generating step. Thus, the *sifting* process has to be repeated on the prototype of IMF until $p_k(t)$ satisfies the two conditions above, and the first IMF $c_1(t)$ is now obtained as the last $p_k(t)$. To terminate the *sifting* process, we have to establish a

criterion. A common method is accomplished by limiting the size of the standard difference (SD) calculated from the two consecutive sifting results as

$$\mathrm{SD} = \sum_{t=0}^{T} \frac{\left| p_{k-1}(t) - p_k(t) \right|^2}{p_{k-1}^2(t)} \tag{6}$$

A typical value for SD can be set between 0.2 and 0.3 [19]. Once $c_1(t)$ is obtained, it is then subtracted from the original data to get a residue $r_1(t)$:

$$r_1(t) = X(t) - c_1(t) \tag{7}$$

Obviously, $c_1(t)$ should represent the finest scale mode of oscillation, and the residue $r_1(t)$ still contains useful information about longer time scale components. Therefore, the residue is treated as a new signal, and repeated sifting processes are conducted to obtain:

$$\begin{aligned} r_2(t) &= r_1(t) - c_2(t) \\ &\vdots \\ r_n(t) &= r_{n-1}(t) - c_n(t) \end{aligned} \tag{8}$$

The whole process can be stopped when (1) $c_n(t)$ or $r_n(t)$ is less than a predetermined threshold, or (2) $r_n(t)$ becomes a constant or monotonic function. By combining equations (7) and (8), we finally get the decomposition result:

$$X(t) = \sum_{k=1}^{n} c_k(t) + r_n(t) \tag{9}$$

In this result, the original signal is decomposed into n IMFs and one residue. It's obvious that lower and higher order IMFs correspond to smaller and larger time scales respectively, and the residue reflects the trend of the signal. Moreover, the residue is often denoted as the last IMF $c_{n+1}(t)$ for convenience.

2.3 Baseline Wander Removal

2.3.1 Sources of Baseline Wander

Extraneous low-frequency components can severely influence the visual interpretation of an ECG as well as the results obtained from computer-based ECG analysis. Removal of baseline wander is therefore required in the analysis of ECG signals to minimize changes in beat morphology with no physiological counterpart. Respiration and electrode impedance changes due to perspiration are important sources of baseline wander in most types of ECG recordings.

The frequency content of the baseline wander is usually in a range well below 0.5 Hz. During the latter stages of a stress test, however, increased movements of the body will further add to the baseline wander activity. Patients unable to perform a traditional treadmill or ergometer stress test may still be able to perform a stress test either by sitting, running an ergometer by hand or by using a special rowing device. In these cases, baseline wander related to motion of the arms may distort the signal. The bandwidth of such baseline wander is then considerably larger than that caused by respiration and electrode impedance changes.

2.3.2 Baseline Wander Removal by EMD

Many methods have been previously presented to remove ECG baseline wander, using techniques such as robust averaging, polynomial interpolation, linear filtering and so on. Recently, Weng *et al.* [24] proposed a technique using empirical mode composition, and demonstrated superior performance to other methods. In this study, we adopt a

simplified version of their approach, which is less complicated but still has comparable performance as shown below.

As illustrated in Fig. 8, given a segment of ECG signal with baseline wander, we first decompose it into a set of intrinsic mode functions (IMFs). Since baseline wander is a low-frequency component, it is supposed to be located in the higher order IMFs. As shown in Fig. 8, the 9th order IMF can be regarded as the baseline wander, while the first 8 order IMFs can be combined to construct the ECG of interest. In practice, the decomposition of an ECG signal usually results in not only 9 IMFs. However, for the problem of baseline wander removal, we find that the first 8 order IMFs are enough to construct a satisfying ECG of interest.

Our simple approach to baseline wander removal can thus be summarized as follows:

(1) If the resulted number n of IMFs is not less than 9, then keep the first 8 IMFs to remove the baseline wander;

(2) If the resulted number n of IMFs is less than 9, then keep the first n-1 IMFs to remove the baseline wander.

2.3.3 Experiments on Baseline Wander Removal

To demonstrate the performance of the above method, an experiment is conducted. A linear filtering method is used as benchmark, which is a high-pass filter with a cutoff frequency of 0.1Hz. The results are shown in Fig. 9. The left column gives a whole view of the signal, and the right column shows the enlarged local view of part of the signal.

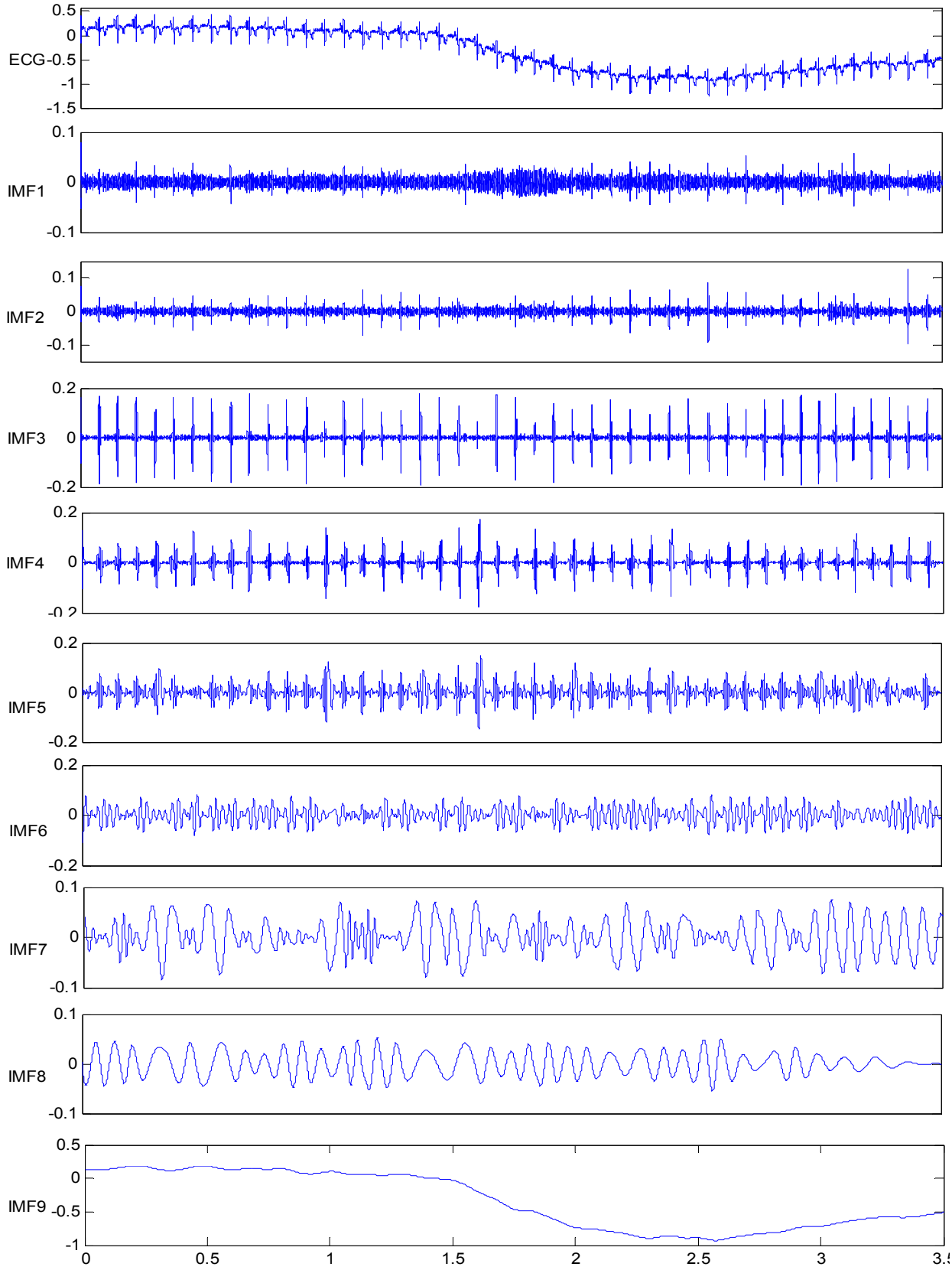


Fig. 8. ECG with baseline wander and its IMFs

Overall, it seems that the results of the two methods are comparable. However, a close examination of the local behavior reveals that the EMD method outperforms the high-pass filter in the aspect of avoiding morphological distortion of feature waves. Compared with the original ECG, linear filtering introduces an obvious convex wave between a T wave and the next P wave, and this artifact would definitely bring confusion into the following feature extraction and classification. On the other hand, the EMD method has not only removed the baseline wander, but also kept the original

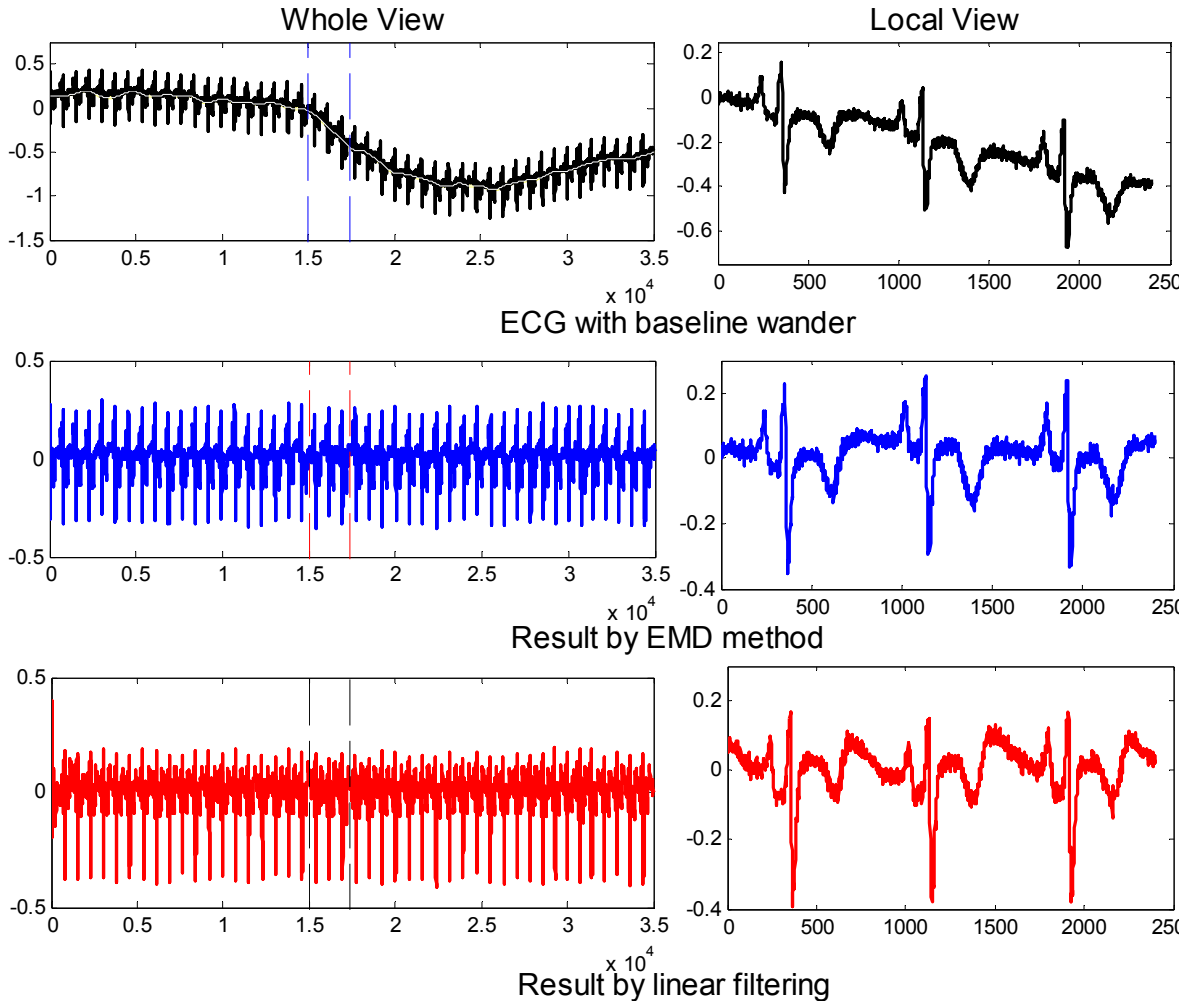


Fig. 9. Results of baseline wander removal using EMD and linear filtering.

morphology of ECG well.

After removing the baseline wander from raw ECG signals, we turn our attention to ECG denoising aiming at reducing the high-frequency noises. In this study, a novel denoising method is proposed based on ECG model, EMD and instantaneous frequency. This is presented in the following section.

2.4 ECG Denoising

2.4.1 Introduction

While ECG has been extensively used for heart disease diagnosis in hospital as well as patient monitoring at home, reliable and efficient clinical applications highly depend on the accuracy of information extracted from the ECG recording, since ECG signals are usually corrupted with various artifacts.

Although a lot of studies on ECG applications have been reported during the past years, it still remains an important problem to extract high-resolution cardiac signals from noisy ECGs. The source of ECG artifacts can be cardiac-related, for example, reduction or disappearance of the isoelectric interval, prolonged repolarization, or atrial flutter; extracardiac noise sources include respiration, changes of electrode position, muscle contraction, and power line interference [25]. Therefore, the goal of ECG denoising is to separate the valid cardiac components from the background noises so as to obtain a signal that is qualified for reliable interpretation.

Due to the overlapping between cardiac components and noncardiac ECG contaminants in the frequency domain, especially from 0.01 Hz to 100 Hz, traditional linear filtering (e.g. low-pass or band-pass filter) is not adequate to eliminate such

noises while keeping valid components unchanged [26] [27]. Recently, numerous approaches have been proposed to denoise ECG signals, for example, principal component analysis (PCA) [28], independent component analysis (ICA) [29], neural networks (NN) [30], and wavelet transform (WT) based denoising techniques [31]. Although they demonstrated good performance, the model of the ECG in these methods is either fairly arbitrary or essentially based on the frequency content of the ECG and the location of the ECG peaks in time to some degree.

As described above, a dynamic model was firstly developed by McSharry *et al*. [20] for synthesizing artificial ECGs, based on three coupled ordinary differential equations. Later, Clifford *et al.* [32] presented a model-based filtering scheme which fitted the model to a noisy ECG in the minimum mean square error (MMSE) sense by performing a constrained nonlinear optimization. This method can capture much clinical information of the heartbeat, but its performance is still limited by the optimization process which easily falls into a local optimal solution.

Recently, a new signal processing method called Empirical mode decomposition (EMD) has been introduced by Huang *et al*. [19] for analyzing data from non-stationary and nonlinear processes. The major advantage of EMD is that the basis functions used to decompose a signal are not predefined like conventional methods such as wavelets, but derived from the signal itself via an iterative procedure *sifting*. Therefore, EMD has found vast applications in signal analysis, including biomedical engineering problems. However, directly applying EMD to the ECG denoising will not produce a desired result due to the physiological characteristics of ECG, as will be detailed later. In [33], Weng *et al.* proposed to avoid this question by preventing the QRS complex from being filtered, but this method only denoised ECG partly.

In the present study, we propose to apply the EMD algorithm to ECG denoising problem, together with the ECG dynamic model. The contribution of our method is that it not only combines the advantages ECG model of EMD algorithm for the ECG denoising problem, but also introduces instantaneous frequency into this problem, making the EMD-based denoising scheme applicable to different kinds of ECG noises.

Experiments have been conducted on real ECG records from the MIT-BIH Arrhythmia Database [34] by contaminating the original ECG with white noise and real noise respectively. Both quantitative and qualitative results show that our method offers a superior performance for ECG denoising.

2.4.2 Instantaneous Frequency

In the traditional Fourier analysis, frequency is defined for the sine or cosine function spanning the whole data with constant amplitude. Thus, we need at least one full oscillation of a sine or cosine wave to define the local frequency. However, such a definition makes no sense for non-stationary data for which the frequency changes values from time to time. Therefore instantaneous frequency (InF) was introduced based on Hilbert Transform.

For an arbitrary time series $X(t)$, we can always have its Hilbert Transform $Y(t)$, as

$$Y(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{X(\tau)}{t-\tau} d\tau = \frac{1}{\pi} \lim_{a \to \infty} \int_{-a}^{a} \frac{X(\tau)}{t-\tau} d\tau \tag{10}$$

where *p.v.* indicates the Cauchy principal value. This transform exists for all functions of class $L^p(\mathrm{R})$ with $1<p<\infty$. Based on this definition, $X(t)$ and $Y(t)$ form a complex conjugate pair, so we can have an analytic signal $Z(t)$, as

$$Z(t) = X(t) + \mathrm{i}Y(t) = a(t)e^{\mathrm{i}\theta(t)} \tag{11}$$

where $a(t) = \sqrt{X^2(t) + Y^2(t)}$, $\theta(t) = \arctan(Y(t) / X(t))$.

With $Z(t)$, the instantaneous frequency of $X(t)$ is then defined as

$$f(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}.$$

It should be noted that $f(t)$ is a single value function, so the instantaneous frequency makes no sense for signals which contain multiple frequencies at a given time. Therefore, in order to have a meaningful instantaneous frequency at any time, the function should be locally symmetric with respect to the zero mean level [19], and this is just satisfied by the properties of an IMF obtained from EMD (see Section 2.3).

For a discrete-time signal $x(n)$, the Discrete Hilbert Transform (DHT) has several forms and we adopt the basic DHT proposed by Kak [53]:

$$y(k) = \begin{cases} \frac{2}{\pi}\sum_{n\ odd}\frac{x(n)}{k-n} & k \quad even \\ \frac{2}{\pi}\sum_{n\ even}\frac{x(n)}{k-n} & k \quad odd \end{cases}$$

Then, $z(n) = x(n) + \mathrm{i}y(n) = a(n)e^{\mathrm{i}\theta(n)}$, and the instantaneous frequency of $x(n)$ can be estimated by the central finite difference:

$$f(n) = \frac{1}{4\pi}[\theta(n+1) - \theta(n-1)] \tag{12}$$

In the following, we will discuss the application of instantaneous frequency to ECG denoising on the basis of IMFs.

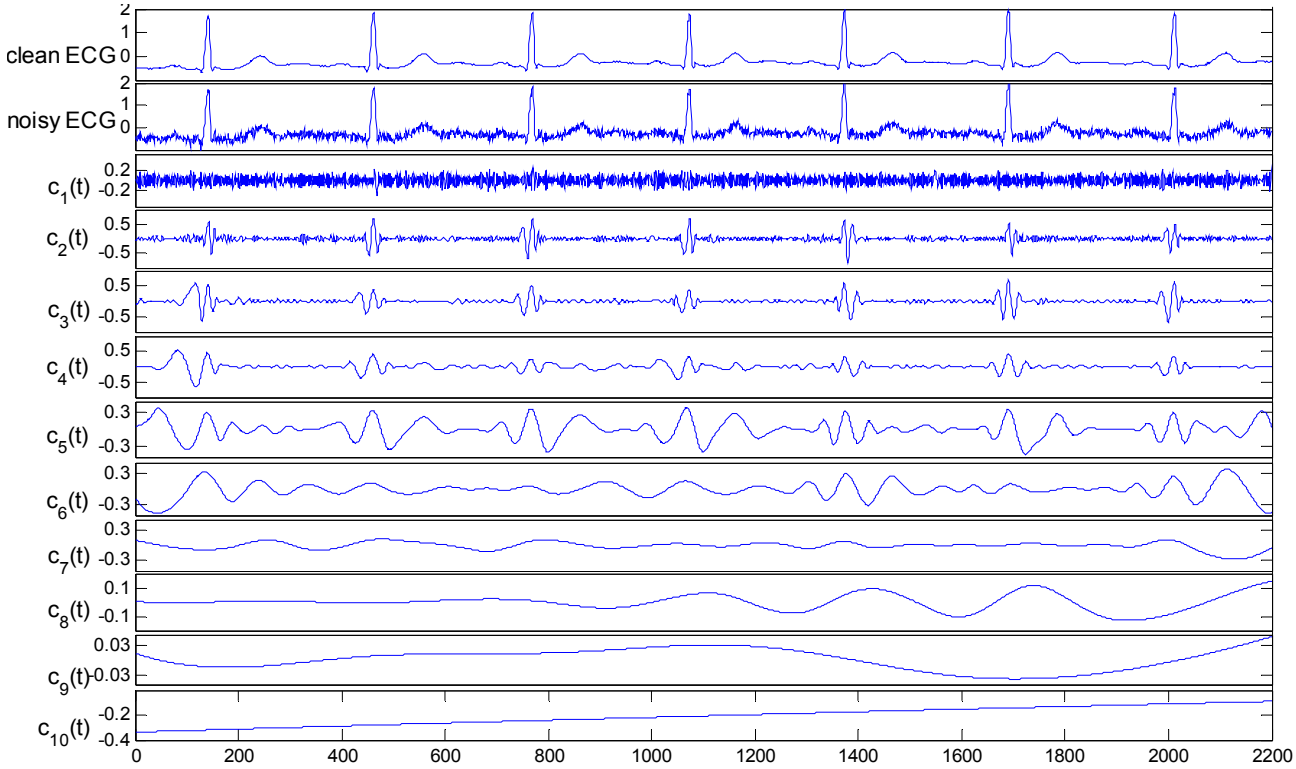


Fig. 10. Decomposition of noisy ECG. The top first plot is a clean ECG, and the top second is a noisy ECG generated by adding white noise to the clean one. Except the top two, all the others are the IMFs of the noisy ECG.

2.4.3 Problem of Direct ECG Denoising by EMD

Using the EMD algorithm, we now can decompose an ECG signal (with baseline wander removed) into a set of IMFs. Since our interest is ECG denoising problem, a noisy ECG signal is generated by adding white noise to a clean ECG, and one sample of decomposition result is shown in Fig. 10. The first and second top subfigures are respectively the clean and noisy ECGs, and below them are all the 10 IMFs, from the lowest order to the highest order. From this figure, we acquire two pieces of information: (1) the average time scale of an IMF increases as the IMF order increases; and (2) the time scale contained in one IMF is not uniform, but various with time.

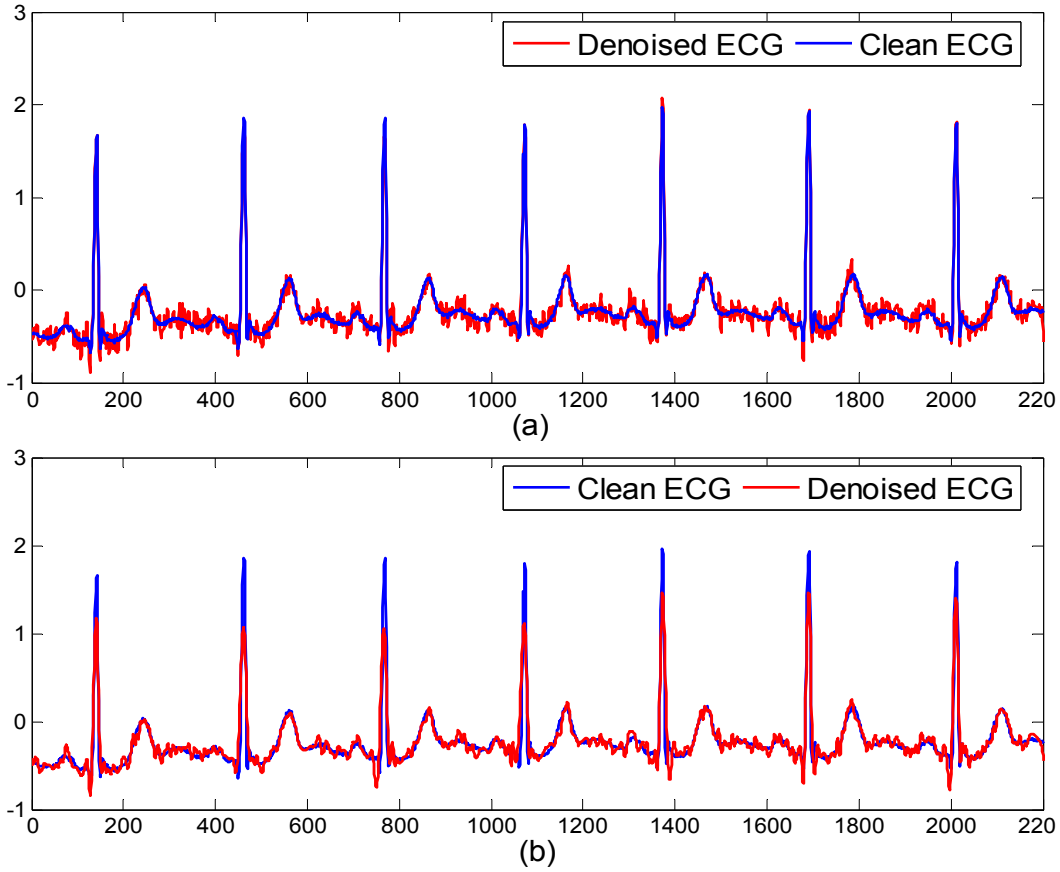


Fig. 11. Direct denoised results by EMD. (a) Removing the 1st order IMF, (b) Removing the top two IMFs. Note the residual noise in (a), and the R wave distortion in (b).

The basic denoising principle using the EMD as a tool is to construct the desired signal with a partial sum of the IMFs. Although various approaches have been proposed to identify whether a specific IMF contains useful information or noise [35], their performances are not satisfactory when directly applied to the problem of ECG denoising, as explained below.

Referring to the IMFs of the noisy ECG shown in Fig. 10, our tendency would be to reconstruct a denoised signal using some of them. After examining the IMFs, it's easy to find that the first order IMF contains almost nothing but high frequency noise, and that IMFs ranging from the third to the last order, can be considered to mainly contain useful information about the ECG components. However, it is in the second IMF that

the problem lies, since it contains both high frequency noise and components of the QRS complex.

If we simply discard the first order IMF as noise, the output will still consist of considerable noise as illustrated in Fig. 11 (a). If we remove the second order IMF together, the resultant ECG will have the R waves heavily distorted as shown in Fig. 11 (b). Neither result is, unfortunately, desirable.

The cause of this problem is that the R wave of ECG has a sharp and high waveform, which easily falls in lower order IMFs together with noise components. To deal with this problem, [33] proposed to preserve the QRS complex with a window when discarding the noise-dominant IMFs. Although this method is able to remove much noise from the noisy signal, it suffers an obvious flaw that it has simply ignored the QRS complex in the denoising process, which is of high importance in ECG morphology. Therefore, a more considerate scheme is presented here, by virtue of the ECG dynamic model.

2.4.4 Model-based Pre-filtering

The motivation of the model-based method comes from our perspective that the problem caused by ECG characteristics should be resolved via the characteristics themselves. Fortunately, the ECG dynamic model is just an excellent tool to capture ECG morphological features. In fact, Clifford *et al.* [32] have demonstrated that this model can be used to filter a noisy ECG by fitting the equation (2) to a cycle of ECG segment, and the filtered signal can preserve much of the clinical information. In our study, we replace the model with the generalized model, and the fitting result is shown

in Fig. 12 (a). The advantage of this model-based filtering lies in the sound physiological background behind the dynamic model.

In practice, the ECG model $z(\theta)$ is fitted to a noisy signal $s(\theta)$ by minimizing the squared error between them. That is, we need to find

$$\min_{a_i, b_{1i}, b_{2i}, \theta_i} \left\| z(\theta) - s(\theta) \right\|^2 \tag{13}$$

over all five $i \in \{P, Q, R, S, T\}$, with $\theta \in [0, 2\pi]$, and equation (13) is then solved by

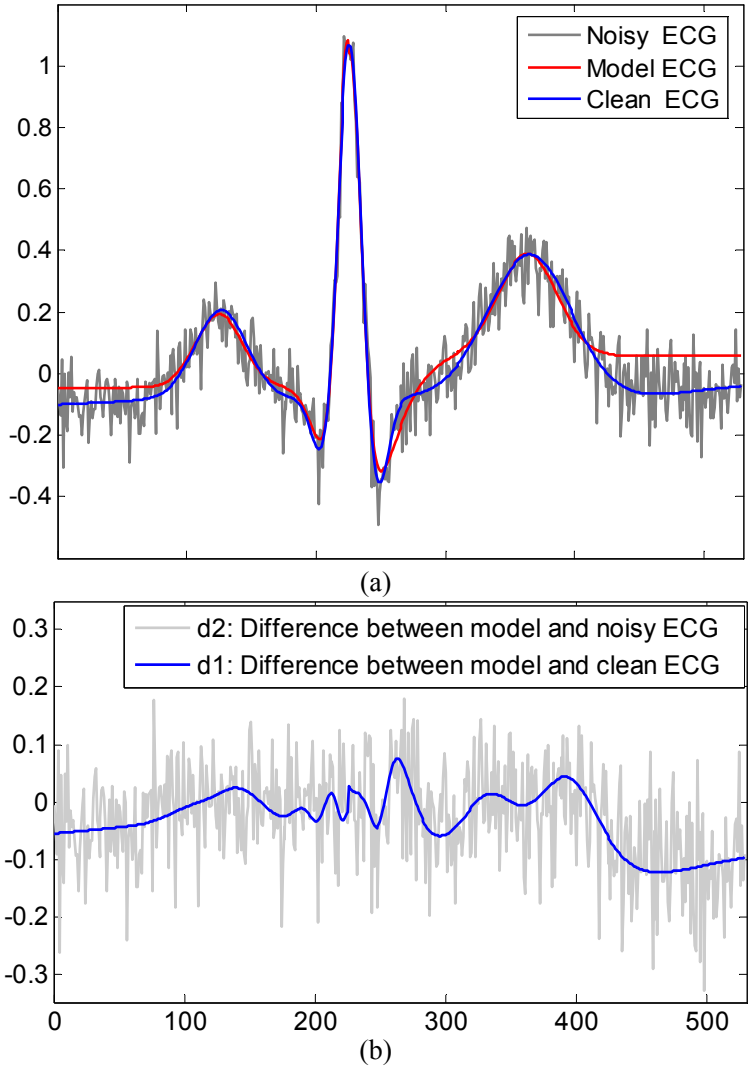


Fig. 12. (a) ECG model fitted to noisy ECG, (b) Differences between the model ECG and noisy and clean ECG.

a 20-D gradient descent method in the parameter space. The Matlab function *lsqnonlin.m* can be utilized to implement this nonlinear least-square optimization. However, this optimization problem can only find a local solution neighboring to the initial solution. Thus, the resultant model usually does not fit the noisy signal very well, and consequently differs from the clean ECG. As shown in Fig. 12 (b), d_1 represents the signal difference between the model and the clean ECG, while d_2 is the difference between the model and the noisy ECG.

Based on d_2, we wish to estimate the difference signal d_1, and combine it with the

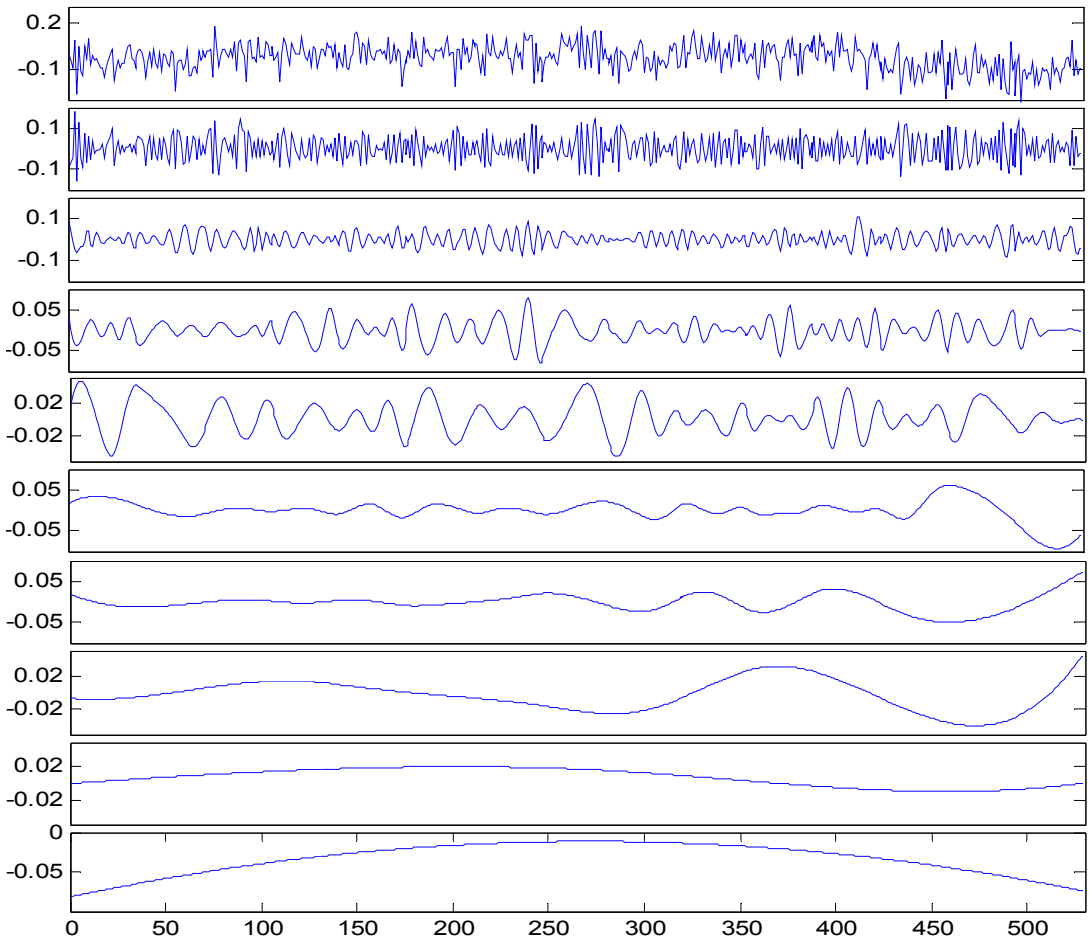


Fig. 13. EMD of d_2. The top subfigure is the signal d_2, and the others are all the IMFs from lower order to higher order.

model to obtain the finally denoised ECG, hopefully close to the clean one as much as possible.

2.4.5 EMD Denoising Using Significance Test

Our objective is to denoise the residue signal d_2 using EMD algorithm, in order to estimate the ideal signal d_1. After d_2 is decomposed into a set of IMFs, as shown in Fig. 13, we need to identify the noise components from them and use the rest to construct a filtered signal.

However, it is not a trivial problem to determine which IMF comprises useful information and which is primarily of noise, for different data will give various decomposition results. Through analyzing the EMD of fractional Gaussian noise, a significance IMF test procedure was proposed by Flandrin *et al.* [36]. They found that apart from the first IMF of the noise-only signal, the power spectra of the other IMFs

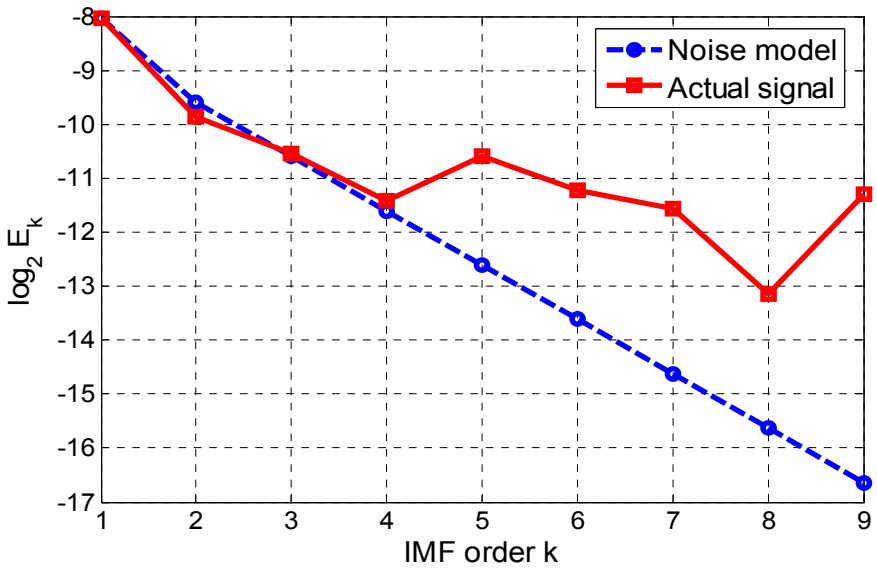


Fig. 14. Significance IMF test. Note that the IMF energy distribution of actual signal diverges from that of noise model at k=5.

exhibit self similar characteristics like those appearing in any dyadic filter structure. Therefore, the logarithm of the k-order IMF energy E_k, i.e. $\log_2 E_k$, should linearly decrease with respect to k.

In the case of white Gaussian noise (a special case of fractional Gaussian noise), the IMF energies of a white-noise-only signal can be described approximately by:

$$E_k = \frac{\sigma^2}{0.719} 2.01^{-k} \qquad k = 2,3,4\cdots \tag{14}$$

where σ^2 can be approximated by the variance of the first-order IMF of the noisy signal

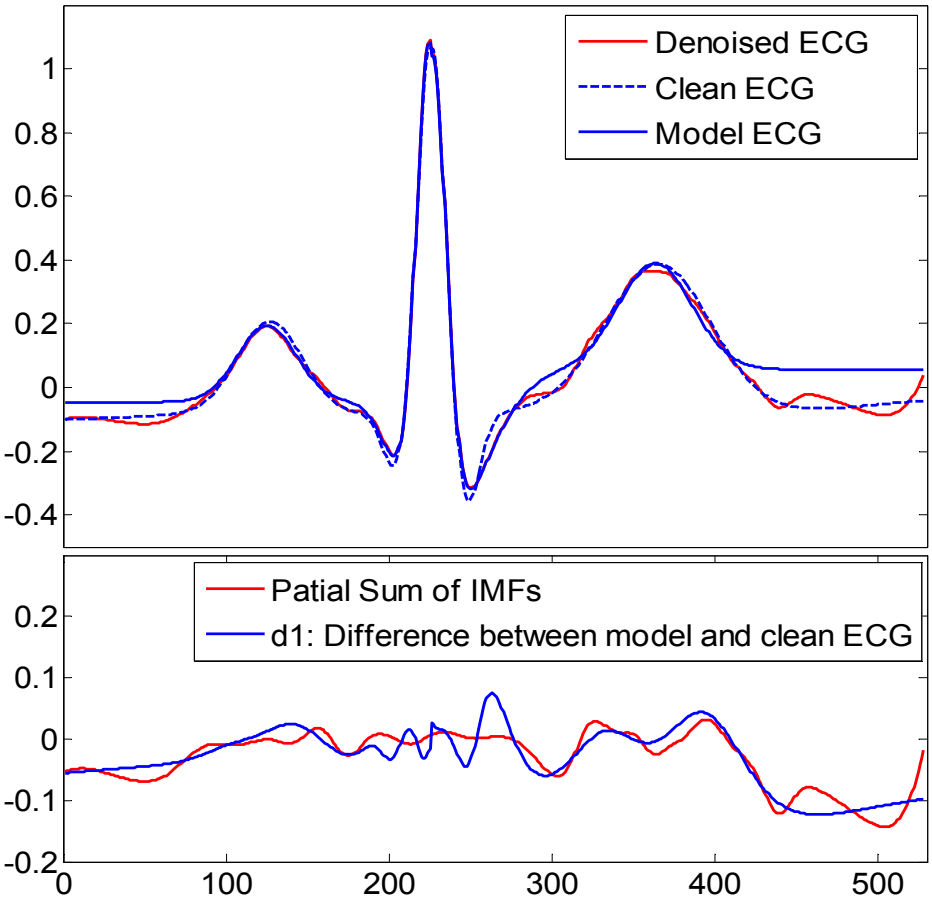


Fig. 15. The upper shows the finally denoised ECG, and the lower shows the partial sum of IMFs of d_2.

[37]. In Fig. 14, the dashed line shows the linear relationship between $\log_2 E_k$ and IMF order k for a white noise model, while the solid line denotes the actual relationship for the noisy signal d_2.

From Fig. 14, we observe that the first 4 IMFs of d_2 share very similar energy distributions with those of the noise-only signal, but from the 5th order, the two lines diverge significantly from each other. This phenomenon indicates that the top four IMFs of d_2 contain primarily noise components and therefore should be discarded, and useful signal information is expected to reside in the rest IMFs. To verify this, the partial sum of the last 5 IMFs is shown in the lower subfigure of Fig. 15.

Despite of some difference from the target signal d_1, the partial sum has captured its major information, especially the trend. Finally, we obtain the denoised ECG by adding the partial sum to the model (see section 2.5.4), and the result turns out much closer to the clean ECG as illustrated in the upper of Fig. 15.

In order to measure the similarity of IMF energy distribution between the noisy signal and the noise-only one, we can set a difference threshold for $\log_2 E_k$. That is, if the difference is larger than the threshold, we regard the corresponding IMF as a noise component and discard it. In this study, we empirically choose the threshold as $|0.05\log_2\sigma^2|$, where σ^2 is the variance of the first-order IMF.

2.4.6 EMD Denoising using Instantaneous Frequency

Although the above method has an excellent denoising performance, it works only when the noise can be regarded as white noise. In practice, real ECG noises, for example, muscle artifacts, are mostly far from white noise. Therefore, a more comprehensive approach is needed to deal with different kinds of noises.

The essence of the foregoing method is to find a critical order of IMF, above which all the IMFs are considered to be noise components and thus discarded. In this sense, it is like the low-pass filtering in Fourier analysis, where cut-off frequency is used. One drawback of this kind of method is the lack of local discrimination, which is an inherent drawback of Fourier methods. However, IMFs derived from EMD are different from harmonic functions, since they have time-variant frequencies, which can be measured by instantaneous frequency. Therefore, we present a novel method for EMD denoising based on instantaneous frequency.

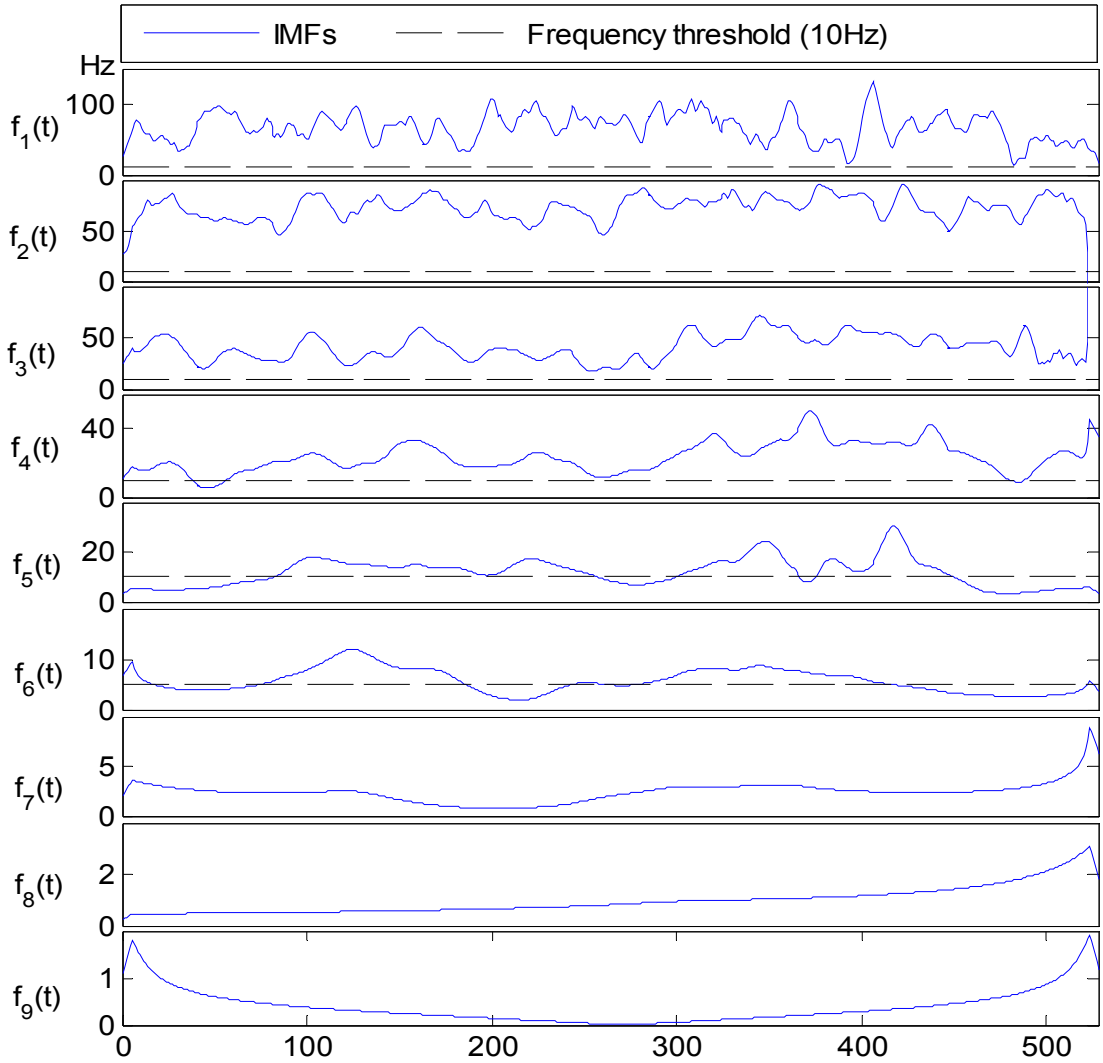


Fig. 16. Instantaneous frequency of the IMFs of d_2 and threshold.

For each IMF of d_2, we first calculate its instantaneous frequency using (12). Since the IMFs are not ideal due to quantization error, the raw instantaneous frequency would be too noisy to be used directly. Thus, we need to smooth the instantaneous frequency with an average filter, and the results are shown in Fig. 16. Next, we wish to remove high frequency noise components.

A natural way is to set a frequency threshold and discard the parts whose instantaneous frequencies are above threshold. However, this method suffers a flaw that the resultant signal will be discontinuous and rugged, because the selected IMF segments may have sharp edges at the ends, as illustrated in Fig. 17 (a). To avoid this problem, we propose to extend the two ends of the segment to the nearest zero-crossing points respectively, as shown in Fig. 17 (b). In this way, the final constructed signal becomes continuous.

In this study, we set the frequency threshold to be 10Hz, as the dashed line shown in

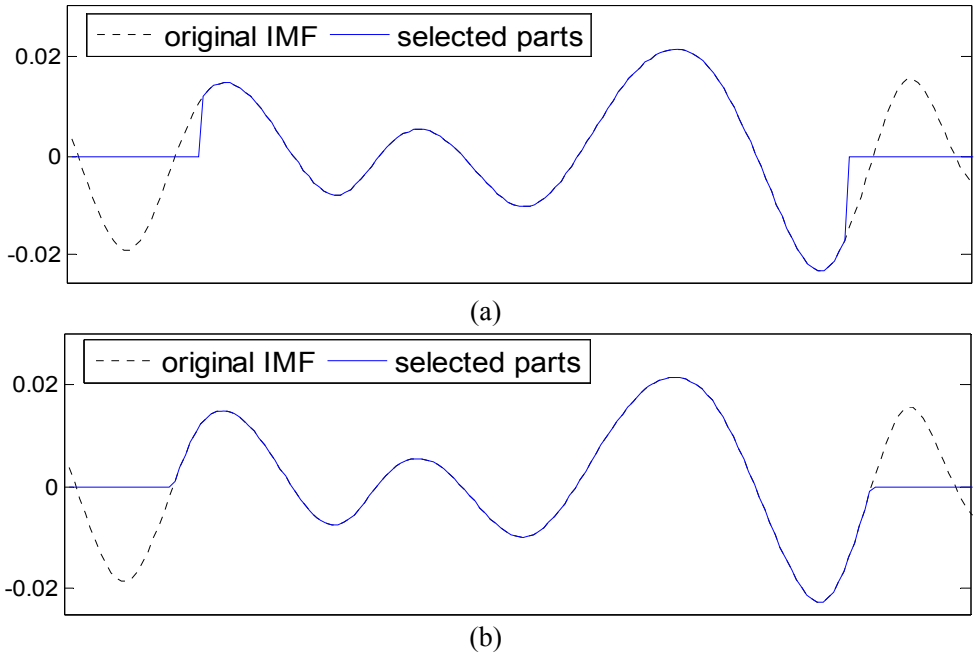


Fig. 17. Selected IMF segment by instantaneous frequency threshold: (a) with sharp edges, (b) with ends extended to zero-crossing points.

Fig. 16. From the relationships between the instantaneous frequency and threshold, we can see that the first four IMFs are almost regarded as noise components, and the last three IMFs are all interest signals, which is similar to the situations of the foregoing method.

However, the 5th and 6th IMFs are both partially divided into noise segments and valid segments, which differs from the foregoing method. Therefore, the instantaneous frequency based method has the advantage of local discrimination, and since it has no assumption on noise properties, it is applicable to both artificial noise and real noise. Using the method, the final denoising performance is basically as good as the foregoing one, as shown in Fig. 18.

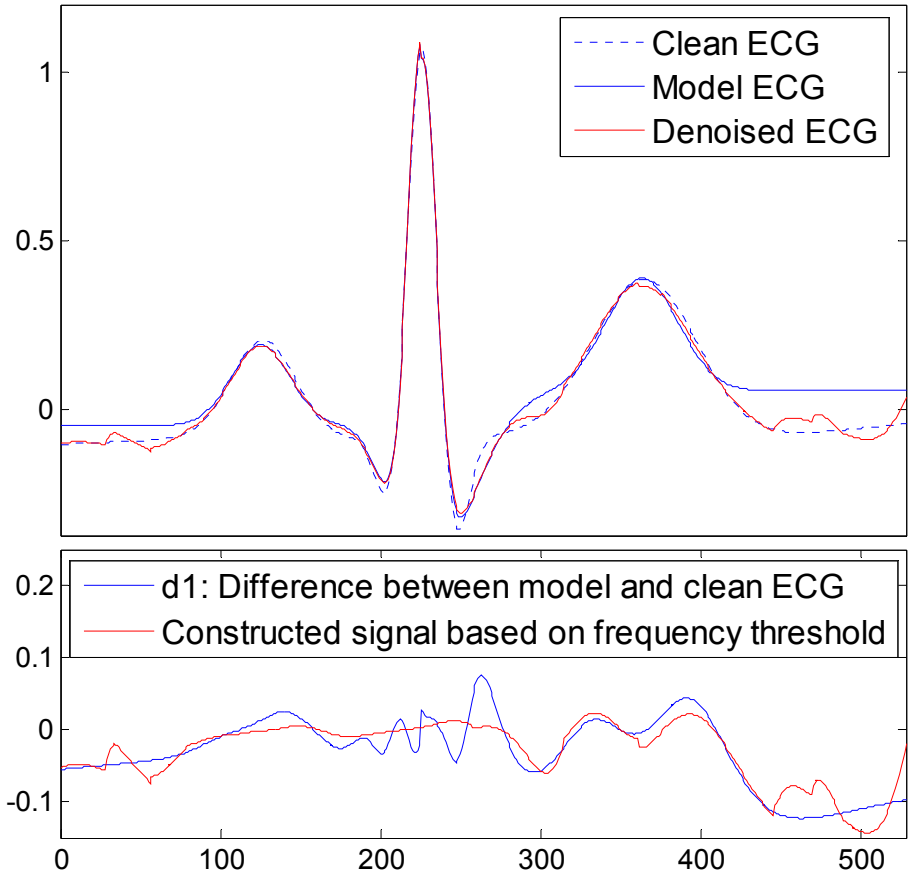


Fig. 18. Denoising results by instantaneous frequency threshold. Compare it with Fig. 15 to see the difference.

2.4.7 Experiments

To verify the efficacy of the proposed method, we have carried out experiments on the MIT-BIH Arrhythmia Database, using both artificial noise and real noise.

As a measurement of the denoising performance, we adopt the notion of Signal-to-Noise Ratio Improvement (*SNRImp*), which is defined as

$$\begin{aligned} SNRImp[dB] &= SNR_{output} - SNR_{input} \\ &= 10\log\left(\sum\nolimits_i \left|x_n(i) - x_c(i)\right|^2 \Big/ \sum\nolimits_i \left|x_d(i) - x_c(i)\right|^2\right) \end{aligned} \tag{15}$$

where x_c denotes the clean signal, x_n represents the noisy signal, x_d is the denoised one.

To set a benchmark, we have also implemented the previous window-based EMD denoising scheme [33] which uses a window to preserve the QRS complex.

A. Test on White Noise

We arbitrarily choose ECG records from the database, adopt them as clean signals, and add white Gaussian noise to them to generate noisy ECGs with various Signal-to-Noise Ratios (SNRs). Then, we utilize the benchmark method (Win-EMD), the proposed method using significance test (ST-EMD), and the one based on instantaneous frequency (InF-EMD) to denoise the noisy inputs respectively. The quantitative denoising results *SNRImp* are listed in Table 1, where input SNRs are of values 15dB, 9dB and 3dB.

Table 1

Denoising Results on White Noise: *SNRImp* (dB)

Record No.	SNR (dB)	100	106	123	220	230	Mean
Win-EMD	15	4.13	4.27	4.06	4.52	5.20	**4.44**
	9	6.30	6.44	7.09	6.88	7.13	**6.77**
	3	7.79	7.65	8.15	7.53	7.82	**7.79**
ST-EMD	15	5.05	5.39	5.62	5.82	6.29	**5.63**
	9	8.09	7.88	8.61	8.17	8.33	**8.22**
	3	9.82	9.73	10.9	9.37	9.29	**9.82**
InF-EMD	15	5.11	5.23	5.79	5.64	6.21	**5.59**
	9	8.02	7.78	8.71	8.05	8.42	**8.20**
	3	9.86	9.75	10.1	9.25	9.41	**9.67**

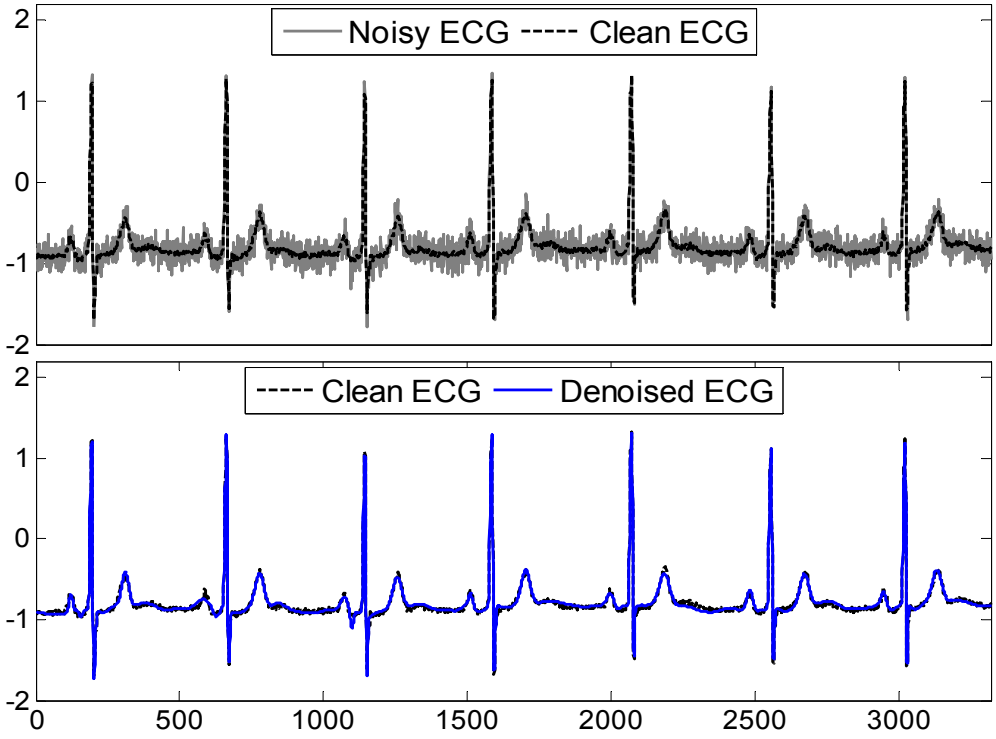


Fig. 19. Denoised result for Record 123 contaminated with white noise. Upper: Clean ECG and 9 dB noisy ECG, lower: Denoising result.

As we can see from Table 1, the *SNRImp* increases as the input ECG SNR decreases, for all three methods. Since white noise is used here, the proposed approach using significance test (ST-EMD) has demonstrated comparable performances to the one based on instantaneous frequency, and what's more, both methods have outperformed the benchmark one (Win-EMD).

In Fig. 19, a portion of Record 123 from the database is illustrated, together with the corresponding 9dB noisy ECG and the denoised result obtained by our method based on instantaneous frequency. As it's shown, the resultant ECG tracks the clean one closely and smoothly, producing a *SNRImp* 8.71dB, which means that the output SNR has been increased to 17.71dB.

B. Test on Real Noise

After discussing artificial noise, we next conduct experiments on real ECG noise to verify the performance of our method. From the MIT-BIH noise stress test database [38], we take two real ECG noise records: the muscle artifact "ma" record, and the electrode motion "em" record. It is worth noting that the baseline wandering in both records should be removed beforehand, since we are only concerned with ECG denoising issue (not baseline wander removal) now.

As shown in Fig. 20, real noise n_r is generated by combining both kinds of noises:

$$n_r = a \cdot n_m + b \cdot n_e \tag{16}$$

where n_m and n_e respectively denote the noise from record "ma" and "em", and weights a and b are chosen so that $a \cdot n_m$ and $b \cdot n_e$ contribute equally to the total noise n_r in power.

As in the foregoing experiment, we first contaminate the ECG records from the database with noises at different SNRs, and then denoise them by the benchmark method and the proposed method based on instantaneous frequency respectively. The

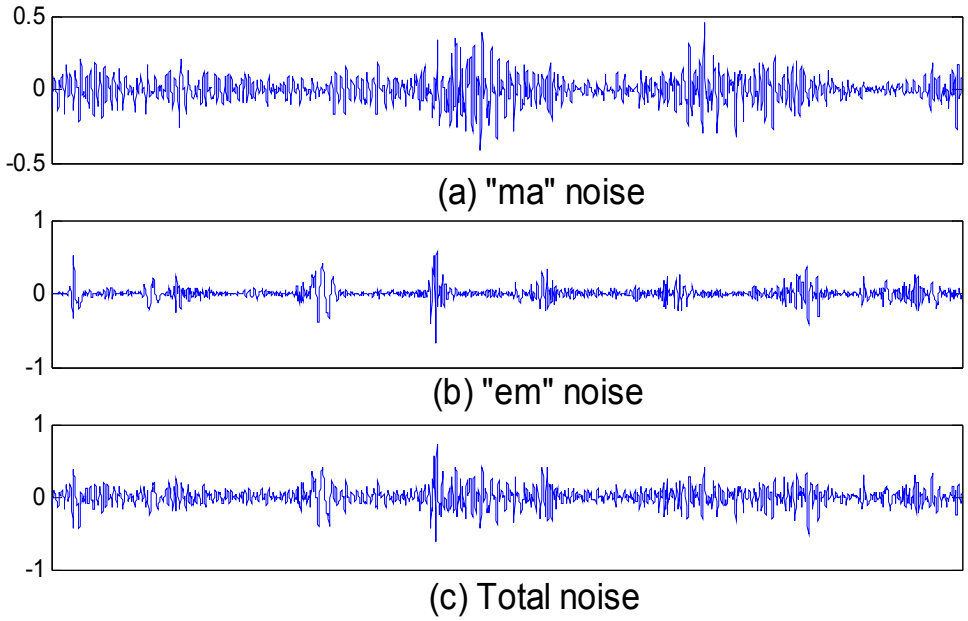


Fig. 20 Real noise and the components

Table 2

Denoising Results on Real Noise: *SNRImp* (dB)

Record No.	SNR (dB)	100	106	123	220	230	Mean
Win-EMD	15	2.75	2.16	2.68	1.90	2.21	**2.34**
	9	3.95	3.86	4.12	4.01	4.57	**4.10**
	3	5.85	4.92	5.96	5.03	5.89	**5.53**
InF-EMD	15	3.90	3.16	3.80	2.53	3.18	**3.31**
	9	5.65	5.36	5.83	5.62	6.32	**5.76**
	3	6.95	5.86	7.15	6.04	7.03	**6.61**

frequency threshold is 10Hz for all tests, and the denoising results are presented in Table 2.

From the quantitative denoising performances of the two schemes, we can see that the proposed method based on instantaneous frequency (InF-EMD) produces superior results to the benchmark one (Win-EMD) on all input ECG records at all SNRs. Moreover, compared with the *SNRImp* in Table 1, it's obvious that real noise is more difficult for each method to deal with than white noise, since the properties of real noise are more irregular and uncertain.

In Fig. 21, a segment of the denoising result of Record 123 is shown. The clean ECG is corrupted by real noise of SNR 9 dB, and then denoised by the proposed method

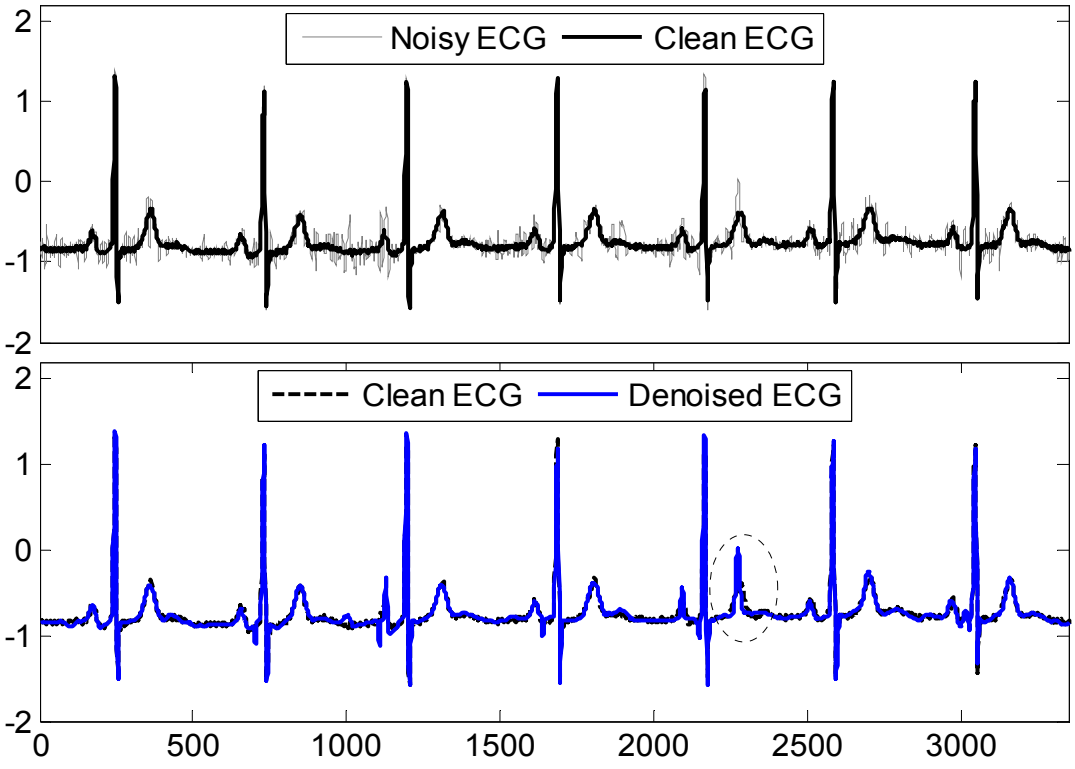


Fig. 21. Denoised result for Record 123 contaminated with real noise. Upper: Clean ECG and 9 dB noisy ECG, lower: Denoising result.

based on instantaneous frequency. From the figure, we can see that the denoised ECG can track the clean one at most time, but the performance is not so desirable in some region (for example, the one encircled by the ellipse). Therefore, the resultant *SNRImp* is only 5.83dB, lower than that of the white noise case.

2.5 Chapter Summary

In this chapter, we first introduced ECG model and its generalization, then discussed the ECG baseline wander removal and finally, presented the ECG denoising schemes.

ECG provides valuable information for CVDs diagnosis, but it's easily contaminated by various kinds of artifacts or noises. Therefore, ECG baseline wander removal and denoising are necessary steps before diagnosis is to be conducted. Here, we adopted an empirical mode decompostion based method to remove ECG baseline wander, and experiment results showed that this method could reduce baseline wander well while not distorting feature waves. For ECG denoising, a method based on EMD and instantaneous frequency is proposed in this work, and experiments on ECG database demonstrated that the performance of the technique is very good.

3. ECG Classification

In this chapter, we will discuss how to identify CVDs through computerized ECG classification. The chapter first introduces ECG database, then ECG feature extraction and classification, and finally, experiments conducted to verify the diagnosis performance.

3.1 Database

This study uses the ECG data from the Physikalisch-Technische Bundesanstalt database (PTB database) [39][40] to train and test the proposed classification system. Freely available at Physionet [41], the PTB database contains 549 records from 290 subjects and each subject is represented by one to five records. Each ECG record includes 15 simultaneously measured signals: the conventional 12 leads together with 3 Frank leads. Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of ±16.384 mV.

According to the "reason for admission" given in the annotation file of each record, the PTB database is categorized into 12 classes: Healthy control (80 records), Myocardial infarction (368 records), Valvular heart disease (6 records), Dysrhythmia (16 records), Heart failure (3 records), Cardiomyopathy (17 records), Palpitation (1 record), Angina (3 records), Hypertrophy (7 records), Bundle branch block (17 records), Myocarditis (4 records), N/A (27 records).

However, not all the data are employed in the present investigation, but only part of them that are selected according to the following two criteria. Firstly, the record amount

of a certain class should be large enough (no less than 5 in this study) to fully represent its characteristics. Secondly, since all heartbeats contained in one record will be regarded as one class in this study according to the record annotation, the actual beat varieties should be as few as possible for each record selected. As a result, the selected classes contain Healthy control (HC), Myocardial infarction (MI), Valvular heart disease (VHD), Cardiomyopathy (CM), Hypertrophy (HT) and Bundle branch block (BBB). Dysrhythmia, in which the beat morphology changes significantly, is discarded.

3.2 Feature Extraction

3.2.1 Feature Selection

Since ECG features in this study will be extracted based on one heart beat, beat segmentation is therefore necessary before feature extraction. The key step of beat segmentation is the detection of R wave. Fortunately, the R wave has been well studied in past years and accurate results have been reported in literature due to its distinguishing morphological character [42][43][44][45]. In this study, we adopt the simple but useful differentiator based algorithm in [42] to localize R waves.

Based on the detected location of R waves, a heartbeat is segmented starting from R_i–$0.45R_{i\text{-}1}R_i$ and ending at R_i+$0.6R_iR_{i+1}$, where R_i is the location of the ith R wave, and $R_iR_j = R_j - R_i$ is the RR interval between beat i and beat j, as shown in Fig. 22. The overlap between the ith T wave and the (i+1)th P wave is designed to preserve enough isopotential information and keep the complete configurations of T and P waves.

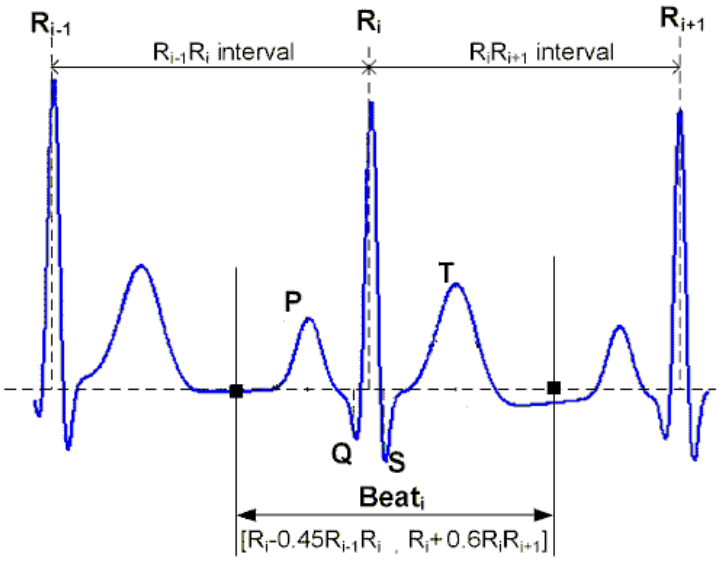


Fig. 22. Beat segmentation and beat features.

After segmenting an ECG signal into a sequence of heart beats, feature extraction is then conducted using the generalized ECG model (Section 2.1.2). According to (4), each beat of ECG can be described by 20 parameters, since each feature wave (P, Q, R, S, or T wave) is modeled by an asymmetric Gaussian with 4 parameters: amplitude a, location θ, and wave widths b_1 and b_2. Besides, the duration of a beat is also a key feature, which is thus included in the feature vector here. Specifically, we select R as the origin point of each beat, so P and Q will have negative locations, while S and T will have positive ones. Since the duration d of each beat is different, the location θ, widths b_1 and b_2 of each feature wave would be normalized by d, so that the elements of feature vector can be independent of each other.

Another point worth noting is how to use multi-lead ECG signals. Although ECG signals usually contain more than one lead (e.g., 12 leads), it is not necessary to use all of them for diagnosis in practice due to the information redundancy. In this study, we have 15 leads, and select Lead I and II (see Section 1.1.1) for ECG classification.

3.2.2 Feature Dimension Reduction by GDA

From Lead I and Lead II of each beat, 41 original features are extracted, to be categorized into 6 classes: HC, MI, VHD CM, HT, and BBB. For a classification problem, the optimal number of eigenvectors for the data transformation is generally equal to $N-1$, where N is the number of classes [46]. The current number of ECG features is so large that some insignificant features may cause redundant or even misleading information and thus affect the classification performance. Therefore, the feature dimension should be reduced before classification. Here we use generalized discriminant analysis (GDA) to explore the most distinguishable feature space. Proposed by Baudat et al. [47], GDA constitutes the kernelized version of linear discriminant analysis (LDA) and was previously applied to ECG analysis [48]. The key idea of GDA is to map the original feature space to a larger one by producing the kernel data projection, which increases class separability of the projected training data, and then reduce the dimension to the desired number by LDA.

Let x_{ij} be the jth feature vector in the ith class of the input data set $X \subseteq R^f$, n_i the number of feature vectors in class i, N the number of classes and $M = \sum_{i=1}^{N} n_i$ the total number of elements in X. The first step is to apply a nonlinear transfer function ϕ to map the feature vectors into a higher feature space R^F, i.e., $\phi : x_{ij} \in R^f \rightarrow \phi(x_{ij}) \in R^F$, with $F > f$. This nonlinear mapping is expected to dig out new features, with the objective to change the original linear unseparable problem to linear separable one.

In the higher-dimensional feature space R^F, the degree of aggregation is measured by covariance matrix. The inter-class inertia B is calculated by

$$B = \frac{1}{M}\sum_{i=1}^{N} n_i \mu_i \mu_i^T \qquad \mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \phi(x_{ij}) \tag{17}$$

and the total inertia of data set V is given by

$$V = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{n_i} \phi(x_{ij})\phi^t(x_{ij}) \tag{18}$$

To facilitate better illustration of the following contents, the class index i is omitted when there is no ambiguity in index of x_{ij}, keeping only the element index j=1,2,…,M.

The classical criteria for class separability are defined by the quotient between the inter-class inertia and the intra-class inertia [49]. Based on this principle, GDA aims at finding a set of eigenvectors which maximize the inter-class inertia and minimize intra-class inertia. Given the desired dimension of reduced feature space, for example D, this problem is equivalent to finding D eigenvectors v^i (i = 1, …, D) of the matrix $V^{-1}B$, corresponding to the largest D eigenvalues λ^i:

$$v^i = \arg\max_{v}(\lambda^i = \frac{v^T B v}{v^T V v}) \tag{19}$$

As eigenvectors are linear combination of F elements, there exist coefficients $a^i = (a_j^i)_{j=1,\ldots,M}$ such that

$$v^i = \sum_{j=1}^{M} a_j^i \phi(x_j) \tag{20}$$

Given the kernel function $k(x_p,x_q) = k_{pq} = \phi^T(x_p)\ \phi(x_q)$ and the corresponding kernel matrix $K = (k_{pq})_{p=1,\ldots,M;\ q=1,\ldots,M}$, the coefficients a^i for each eigenvector are normalized by

$$a^i = \frac{a^i}{(a^{iT} K a^i)^{\frac{1}{2}}}.$$

Finally, for a feature vector x from the original feature space R^f, the transformation to the reduced space R^D is expressed by the following equation:

$$y^i = v^{iT}\phi(x) = \sum_{j=1}^{M} a_j^i \phi(x_j)\phi(x) = \sum_{j=1}^{M} a_j^i k(x_j, x) \tag{21}$$

and the new feature vector $y = (y^i)$ is obtained in the D dimension space.

3.3 Classification by Support Vector Machine

Based on the reduced features above, we employ support vector machine (SVM) to conduct the multi-pattern classification. Basically, SVM solves binary classification problem in the following way. Given a set of training data $(x_1, y_1),\ldots,(x_l, y_l)$, where $x_i \in X \subseteq R^n$ is the feature vector, $y_i \in \{-1,1\}$ is the class label, and l is the size of data set, SVM is proposed to find the optimal separating hyperplane (OSH), which minimizes the following cost function with constraints:

$$\begin{aligned} &\min_{\omega,b,\xi} \quad J(\omega,\xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}\xi_i \\ &s.t. \quad \begin{cases} y_i(\omega\cdot x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \quad (i = 1,\cdots,l) \end{aligned} \tag{22}$$

The cost function $J(\omega, \xi)$ consists of two parts: the regularized term $\frac{1}{2}\|\omega\|^2$, of which the minimum value maximizes the distance between the hyperplane and the closest point of each class, and the penalty term $\sum_{i=1}^{l}\xi_i$ which represents the empirical risk for linear unseparable cases [50]. For a linear separable problem, $\xi_i = 0$. The weighting parameter *C* is the trade-off factor between regularized term and penalty term, of which a larger value emphasizes more on classification errors.

For the typical quadratic programming problem with linear constraints, Lagrange multiplier method is a conventional way to obtain the optimal solution by transforming the original minimization problem into a dual maximization problem, which is given by:

$$\begin{aligned} &\max \quad W(\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ &s.t. \quad \begin{cases} 0 \le \alpha_i \le C \\ \sum_{i=1}^{l}\alpha_i y_i = 0 \end{cases} \quad (i = 1, \cdots, l) \end{aligned} \tag{23}$$

where $\alpha = (\alpha_i)$ is the vector of Lagrange multipliers.

Given a function $\phi(x)$ which maps the input feature space into a higher-dimensional one [51], the objective function of (23) can be rewritten as:

$$\begin{aligned} W(\alpha) &= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \phi(x_i)\cdot\phi(x_j) \\ &= \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \end{aligned} \tag{24}$$

where $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ is the kernel function.

Denote the optimal vector of Lagrange multiplier obtained by solving the above optimization problem as α^*. The corresponding optimal linear weight vector ω^*, the optimal bias b^*, and the OSH are given as follows [52]:

$$\begin{aligned} \omega^* &= \sum_{i=1}^{l} \alpha_i^* y_i \phi(x_i) \\ b^* &= -\frac{\max_{y_i=-1}(\omega^* \cdot \phi(x_i)) + \min_{y_i=1}(\omega^* \cdot \phi(x_i))}{2} \\ \sum_{i=1}^{l} \alpha_i^* y_i K(x_i, x) &+ b^* = 0 \end{aligned} \tag{25}$$

Therefore, the decision function for identifying the class of the input x is given by

$$y(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i^* y_i K(x_i, x) + b^* \right) \tag{26}$$

The idea of using a hyperplane to separate the feature vectors into two groups works well when there are only two target categories, but how does SVM handle the case where the target variable has more than two categories? Several approaches have been suggested, but two are the most popular: (1) "one against many" where each category is split out and all of the other categories are merged; and, (2) "one against one" where $k(k–1)/2$ models are constructed where k is the number of categories. This study uses the more accurate (but more computationally expensive) technique of "one against one".

3.4 Experiments

In this part, experiments are conducted to verify the performance of the proposed classification system using the ECG data from the PTB database. After beat segmentation, the beat numbers of each class are listed in Table 3.

Table 3
Total beat number of each class

MI	HC	CM	HT	BBB	VHD	Sum
3541	3987	2226	985	2331	494	13564

To quantitatively evaluate the classification performance, the following indexes are employed.

$$\begin{aligned} Sensitivity &= \frac{TP}{TP+FN}\times 100\% \\ Specificity &= \frac{TN}{TN+FP}\times 100\% \\ +Predictivity &= \frac{TP}{TP+FP}\times 100\% \end{aligned} \tag{27}$$

where TP is true positive, FN is false negative, TN is true negative and FP is false positive. Sensitivity (Se) highlights the ability to categorize one kind of samples into its right class. Specificity (Sp) shows the ability to reject the samples from other classes. Positive predictivity (Pp) emphasizes the possibility that the samples classified into one class are really from this class. Finally, to evaluate the overall performance, we adopt the classification accuracy which is the ratio of the number of correctly classified beats to the number of total beats.

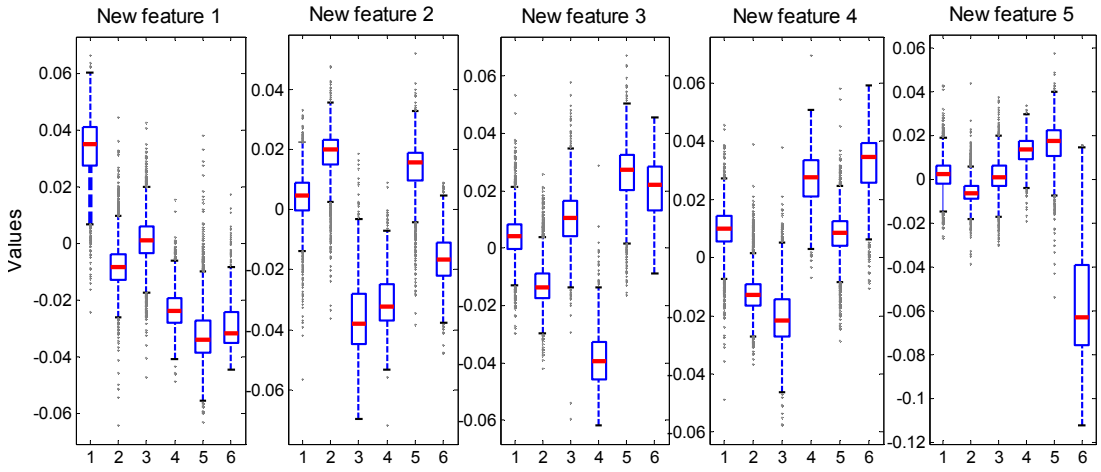


Fig. 23. Box-plot of new features. X axis of each plot stands for the disease class (1=MI, 2=HC, 3=CM, 4=HT, 5=BBB, 6=VHD).

3.4.1 Performance of Feature Reduction

Although a large number of features will preserve more information of the original signal, it also runs the risk of obtaining futile features or misleading features caused by noise, and requires long computation time to complete classification. In this study, we apply GDA to reduce the original 41 features to 5 features.

In this experiment, we only use 1/4 randomly selected beats from each disease class to train the eigenvectors based on the Radial Basis Function (RBF) kernel, i.e., $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$. The small GDA training set is helpful to test the robustness of the reduced features.

Figure 23 shows the box-plot of each new feature for all beats. The lower and upper lines of the box are the 25th and 75th percentiles of the samples. The red line in the middle represents the feature median. The whiskers extend above and below the box, showing the extent of the rest of the features, except for outliers. In this work, we apply

the default definition of outlier to any value that is more than 1.5 times the interquartile range away from the top or bottom of the box. The outliers are shown as gray dots in the figure. It is clear that the boxes of different classes are centralized at different positions at least along one feature axis. For example, only using the new feature 1, MI, HC and CM can be differentiated roughly. However, a close examination of the outliers shows the difficulty of ECG classification that individual physiological differences cause the feature spreading in a relatively large range. Additionally, the large number of beats and the comparatively small training set may also reduce the degree of centralization. This illustrates the non-ideal experiment environment, which in turn demonstrates the robustness and practicability of the proposed system.

Figures 24 and 25 show the distribution of 200 randomly selected beats from each disease in the spaces spanned by the new features 1, 2 and 5 and by 3 and 4 respectively. These two figures intuitively exhibit how the samples are distributed in the new feature space.

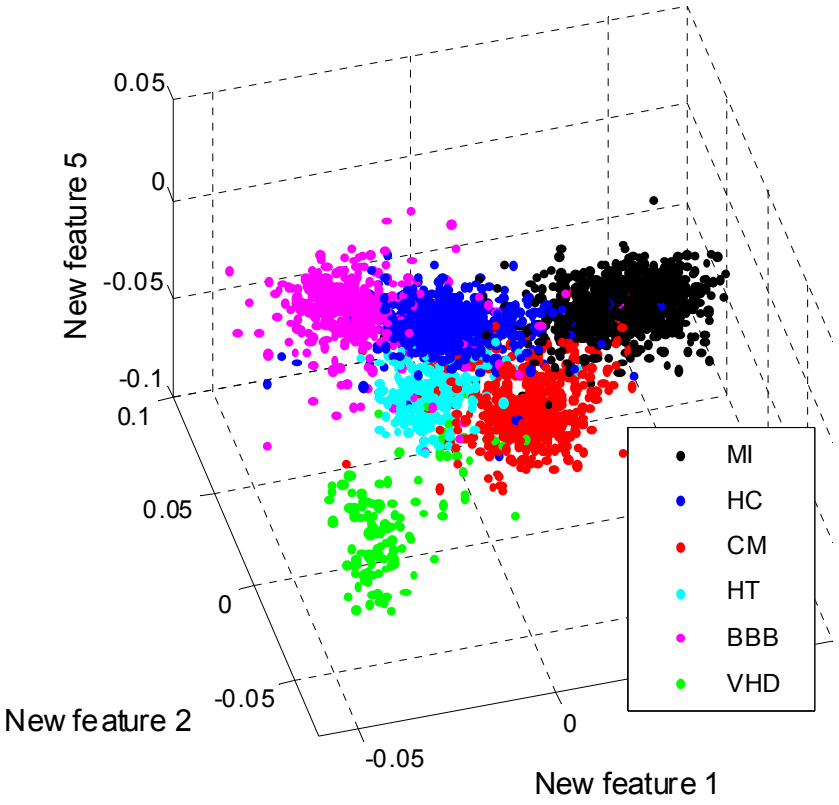


Fig. 24. Locations of 200 beats randomly selected from each class in the subspace spanned by the new features 1, 2 and 5.

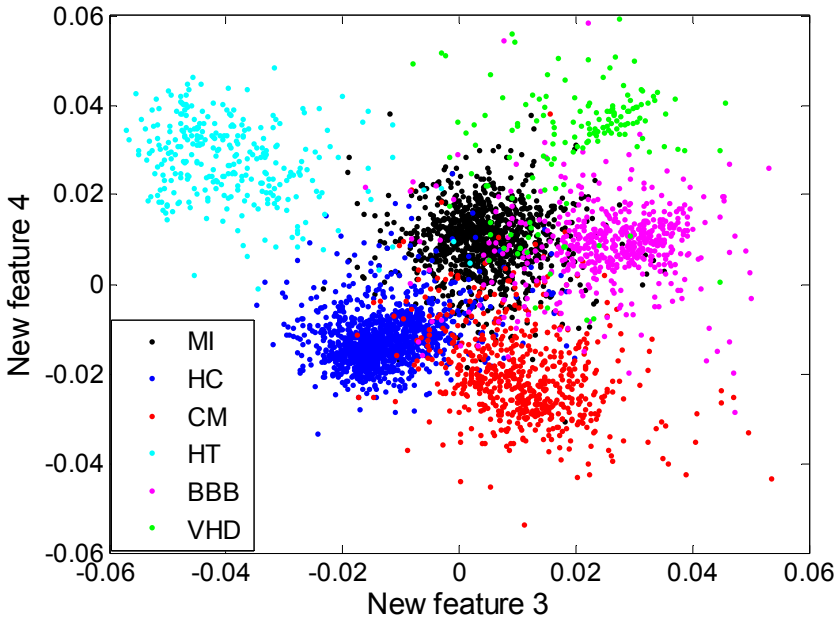


Fig. 25. Locations of 200 beats randomly selected from each class in the subspace spanned by the new features 3 and 4.

3.4.2 Performance of Classification

Based on the reduced features illustrated above, we conduct the classification experiments using SVM. The first investigated issue is the performance under different sizes of training set. A certain percent of beats are randomly selected from each class to make up of the training set of an appointed size, while the rest serve as the test set. In Fig. 26, the classification accuracies under different training set sizes are compared between two approaches: "ORG" (using 41 original features) and "GDA" (using features obtained by GDA). Both curves in this plot are the average results of 100 repeated tests.

Generally speaking, the classification accuracy has close relation with the size of training set. On the one hand, larger training sets could contain more various individual samples, obtain more comprehensive information from the feature space, and therefore

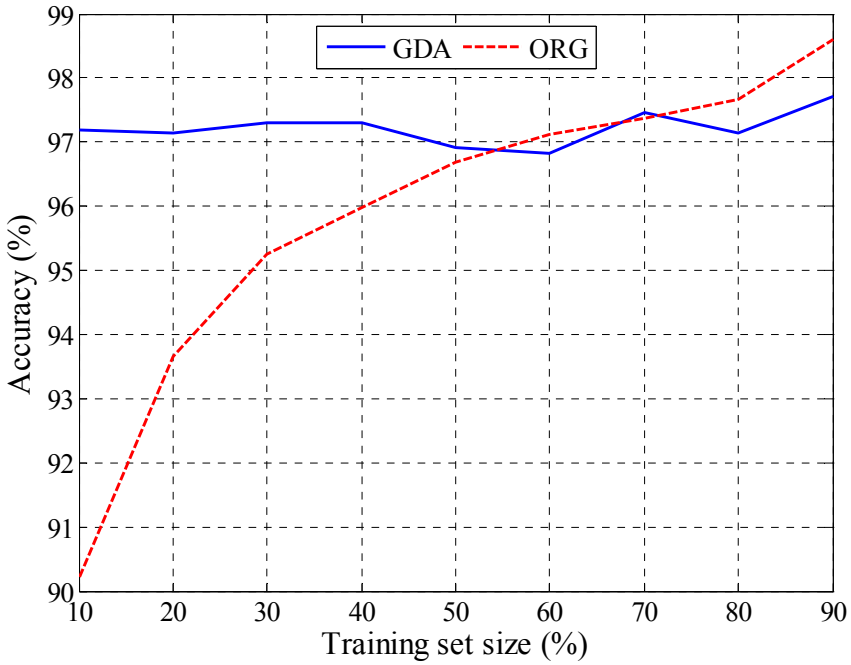


Fig. 26. The total classification accuracy of test set versus the size of training set.

produce a better classification system. On the other hand, if the size of training set is too small to cover the variety of samples, the resultant classifier would have difficulty in working well. As illustrated by the red dashed curve in Fig. 26, the classification performance varies just in a way that is described above, when the 41 original features are used.

However, large size of training set cannot be always guaranteed in reality, and can even be not available in some situations. Therefore, it is desirable to obtain a well-performed classifier while using as few training samples as possible. In practice, using fewer but more significant features turns out to be an efficient way to reduce the training set size requirement. In other words, if we can extract a few features which can capture the underlying characteristics distinguishing one category from another, then the classifier can be trained by few samples while obtaining desirable performance. The blue solid curve in Fig. 26 has just verified the above discussion, since the classification accuracies of small training set sizes are almost as high as those of large sizes. Comparing the two curves in this figure, it is clear that reducing feature dimension by GDA can give a more robust classification performance under different training set sizes.

Another issue worth noting is that the accuracy of classification using original features is a little greater than that of classification using reduced features, when the size of training set is very large, e.g., 90%. This is because larger number of features can preserve more information of the original signal, and given so many training samples, the individuals details are also counted into the classifier. However, the superior performance cannot always be achieved. To some extent, it increases the risk of over-training the classifier, e.g., excessively focusing on individual characteristics or

falsely considering some fake properties caused by noises. In order to avoid over-training, we show some experiment results in detail using small training set size in the following.

Table 4 shows detailed classification performances including Se, Sp and Pp for 25% training set size. The table compares the classification performances using original features with those using reduced features. It can be observed that when the size of training set is not large, more accurate results can be obtained by the classifier using reduced features.

Table 5 presents the confusion matrix also for 25% training set size, and the entries have been normalized by the actual beat number of each class. Each column of the confusion matrix represents the instances in a predicted class, while each row

Table 4
Classification performances using original and reduced features
(training set size: 25%)

Class	Original Features			Reduced Features		
	Se(%)	Sp(%)	Pp(%)	Se(%)	Sp(%)	Pp(%)
MI	95.32	98.63	95.64	98.49	98.66	96.52
HC	96.57	98.13	96.38	98.22	98.37	95.57
CM	92.55	98.30	92.12	95.10	99.08	96.73
HT	94.08	99.26	96.27	98.47	99.46	97.08
BBB	90.48	98.81	92.02	93.81	99.29	97.80
VHD	94.67	99.57	95.31	91.64	99.42	98.59
Average	93.95	98.78	94.62	95.96	99.05	97.05

Table 5

Confusion matrix for classification using reduced features

(training set size: 25%)

		Predicted Class (%)					
		MI	HC	CM	HT	BBB	VHD
Actual Class	MI	98.49	0.53	0.64	0.00	0.34	0.00
	HC	1.03	98.22	0.20	0.08	0.42	0.05
	CM	3.03	0.49	95.10	0.66	0.63	0.09
	HT	0.08	0.39	0.72	98.47	0.25	0.07
	BBB	0.45	5.20	0.54	0.00	93.81	0.00
	VHD	1.53	2.85	2.50	0.82	0.66	91.64

represents the instances in an actual class. It is easy to see from the confusion matrix how the system can be tuned and improved if the system has difficulty in differentiating two classes (i.e. commonly mislabeling one as another).

3.4.3 Performance Comparison with Other Works

It is meaningful but difficult to compare our experimental results with those of other works for several reasons. Firstly, most works on ECG classification problem are only focused on arrhythmias, e.g., PVC, VF, PAC, AF etc, which are not covered in this study. Secondly, some works on detection of CVDs didn't use ECG but other cardiac signals like heart sounds (HSs) [17] or body surface potential maps (BSPM) [16]. Thirdly, many studies on ECG-based CVD detection obtained their results from experiments conducted on very limited amount of samples (e.g., [18]), while we have conducted a large scale experiment based on over 13000 samples. Despite all that, we would still make some rough comparisons between our results and others in the following.

In [17], detection of VHD are studied using wavelet packet decomposition and SVM, and the performance is Se=100% and Sp=96.67%. It should be noted that their

experiment used very few samples: 38 for training and 82 for test. What's more, their classification problem only involved two patterns: VHD and normal ECGs. Considering the larger experiment scale and problem complexity in this study, our results are very acceptable: Se=91.64% and Sp=99.42% for 25% training set size.

In [18], detection of several CVDs is studied using wavelet aided SVM. Table 6 presents the detailed comparison. The column "Test Set Num" indicates the number of samples for each class contained in the test set. It is clear that the experiment scale of our work is much larger than that of [18], and moreover, our classification performance is superior to [18] from the comparison of Se for each class.

Through the above two comparisons, we can conclude that our classification system outperforms other works, because our classification accuracy is no inferior to others while our experiment scale is much larger than theirs.

Table 6

Performance comparison between our work and another ([18])

Class	Results of [18]		Results of this study	
	Se (%)	Test Set Num	Se (%)	Test Set Num
HC	90.00	20	98.22	2990
CM	91.60	12	95.10	1670
MI	93.33	30	98.49	2656
HT	100.00	2	98.47	739
VHD	100.00	1	91.64	371

3.5 Chapter Summary

In this chapter, we discussed how to do ECG classification using SVM. The PTB database provides a rich source of ECG signals corresponding to various CVDs. From these data, six classes of ECG signals are selected for study. ECG features are first

extracted using generalized ECG model, and then GDA is employed to reduce the feature dimension. Finally, large scale experiments are conducted and compared with other works as well. The experiment results show that (1) feature dimension reduction by GDA can improve the performance robustness for classification under different training set sizes; and (2) the whole classification framework works well, achieving high classification performances.

4. Conclusions

The present study is aimed at an approach to diagnose cardiac conditions automatically from ECG signals via computerized signal processing. The goal is to help physicians and patients detect CVDs and take corrective treatments as early as possible. It is well known that CVDs have been the No. 1 killer of the world for over ten years, killing millions of people every year. While ECG can provide important cardiac information for CVD diagnosis, there are still difficulties in using ECG more efficiently. One problem is that ECG obtained by portable monitors cannot be well interpreted in time due to the lack of expertise for common people. Another problem is that it is very tedious and time consuming for physicians to read long time ECG records in hospital. Therefore, this study aims to explore an automatic approach to analyze ECG signals and detect CVDs by computer.

The first step of our approach is ECG signal preprocessing, including baseline wander removal and ECG denoising. For baseline wander removal, EMD is used decompose the ECG and discard the low frequency components. Experiment result shows that this method can preserve the feature waves from distortion better than the low-pass filtering method. For ECG denoising, we have proposed a novel method based on ECG dynamic model, EMD, and instantaneous frequency. The method calls for pre-filtering the noisy ECG and generating a dynamic model that preserves most of the ECG morphological features, especially the QRS complex. The model is then subtracted from the original ECG and the residual signal is decomposed by EMD. Denoising of the decomposed residual signal is aided by the use of instantaneous

frequency to separate the noisy segments from the information segments. The denoised ECG is finally obtained by combining the model and the denoised residue. Experiments are conducted on ECG data contaminated by both Gaussian and real noise, and the results have demonstrated the advantage of the proposed method over other ones for ECG denoising.

After preprocessing, ECG classification is then conducted for detection of CVDs. This study proposes to extract features from each heartbeat using the generalized ECG dynamic model due to the model's accurate capturing of the ECG morphological characteristics, which form the basic evidence for clinical diagnosis. Additionally, GDA is employed to reduce the feature dimension from 41 to 5, which helps to extract the most significant features mathematically. SVM then serves as the classifier for the multi-category ECG classification problem. The PTB database provides rich ECG records corresponding to various CVDs, from which six classes of ECG signals are selected for study, namely, HC, MI, VHD, CM, HT and BBB. Finally, experiments are carried out on the PTB database and also compared with other works. The experimental results show that our proposed method produces not only accurate but also robust classification performance.

To conclude, the presented ECG-based approach for diagnosis of cardiac conditions demonstrated desirable performance and would seem to help both physicians and patients to detect CVDs if put to use. Our future work includes improving the proposed approach for realtime use in order to detect abnormities earlier. Also, more kinds of CVDs and ECG data should be investigated for a thorough evaluation of our approach in the future. The final goal of this work is to develop a set of equipment for practical clinic use.

Reference

[1] G. D. Clifford, F. Azuaje, P. E. McSharry, "Advanced methods and tools for ECG data analysis," Boston: Artech House, 2006

[2] P. Davey, "A new physiological method for heart rate correction of the QT interval," *Heart*, vol. 82, pp. 183-186, 1999

[3] P. J. Schwartz and Wolf S., "Q-T interval as predictor of sudden death in patients with myocardial infarction," *Circulation*, vol. 57, pp. 1074-1077, 1978

[4] http://en.wikipedia.org/wiki/Cardiovascular_disease#cite_note-0

[5] World health organization., Cardiovascular disease. Available at http://www.who.int/mediacentre/factsheets/fs317/en/index.html

[6] M. Risk, J. Sobh, J. Saul, "Beat detection and classification of ECG using self organizing maps," in *Proc. Int. Conf. IEEE EMBS*, vol. 19, 1997, pp. 89-91

[7] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, L. Sornmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, 2000, pp. 838-848

[8] L. Khadra, A. Al-Fahoum, S. Binajjaj, "A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques," *IEEE Trans. Biomed. Eng.*, vol. 52, 2005, pp. 1840-1845

[9] X. Alfonso, T. Nguyen, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, 1999, pp. 192-202

[10] W. Cheng, K. Chan, "Classification of electrocardiogram using hidden markov models," in *Proc. Annu. Int. Conf. IEEE EMBS*, vol. 20, 1998, pp. 143-146.

[11] M. Tsipouras, Y. Goletsis, D. Fotiadis, "A method for arrhythmic episode classification in ECGs using fuzzy logic and Markov models," *Computers in Cardiology*, 2004, pp. 361-364.

[12] S. Osowsaki, T. Linb, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng*., vol. 48, 2001, pp. 1265-1271

[13] L. Shyu, Y. Wu, W. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, 2004, pp. 1269-1273

[14] O. Inan, L. Giovangrandi, GTA. Kovacs, "Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Trans. Biomed. Eng*., vol. 53, no. 12, 2006, pp. 2507-2515

[15] S. Kara, M. Okandan, "Atrial fibrillation classification with artificial neural networks," *Pattern Recogn.*, vol. 40, no. 11, 2007, pp. 2967-2973

[16] H. Zheng, H. Wang, C. Nugent, D. Finlay, "Supervised classification models to detect the presence of old myocardial infarction in body surface potential maps," *Computers in*

Cardiology, vol. 33, 2006, pp. 265-268

[17] S. Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Systems with Applications*, vol. 35, 2008, pp. 1679-1687

[18] D. Ghosh, B. Midya, C. Koley, P. Purkait, "Wavelet aided SVM analysis of ECG signals for cardiac abnormality detection," *IEEE Indicon 2005 Conference*, 2005, pp. 9-13.

[19] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," in *Proc. Roy. Soc. Lond.*, vol. A 454, pp. 903-995, 1998

[20] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamic model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289-294

[21] R. Sameni, M. B. Shamsollahi, C. Jutten, and M. Babaie-Zadeh, "Filtering noisy ECG signals using the extended Kalman filter based on a modified dynamic ECG model," in *Proc. 32nd Annu. Int. Conf. Comput. Cardiol.*, Lyon, France, Sep. 25–28, 2005, pp. 1017–1020.

[22] T. Kato, S. Omachi, and H. Aso, "Asymmetric gaussian and its application to pattern recognition," in *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, Springer-Verlag, 2002

[23] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung, "Engineering analysis of biological variables: An example of blood pressure over 1 day," in *Proc. Nat. Acad. Sci. USA.*, vol. 95, pp. 4816–4821, 1998

[24] B. Weng, M. Blanco-Velasco, K. E. Barner, "Baseline wander correction in ECG by the empirical mode decomposition, " in *Proc. 32nd Annu. Northeast Bioeng. Conf.*, 2006, pp. 135-136

[25] D. Zhang, "Wavelet Approach for ECG baseline wander correction and noise reduction," in *Proc. 27th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Shanghai, China, 2005, pp. 1212-1215

[26] I. I. Christov, I. K. Daskalov, "Filtering of electromyogram artifacts from the electrocardiogram," *Med. Eng. Phys.*, vol. 21, pp. 731–736, 1999

[27] A. Gotchev, N. Nikolaev, and K. Egiazarian, "Improving the transform domain ECG denoising performance by applying interbeat and intrabeat decorrelating transforms," in *Proc. 2001 IEEE Int. Symp. Circuits Syst. (ISCAS)*, Sydney, Australia, 2001, pp. 17–20.

[28] G. Moody and R. Mark, "QRS morphology representation and noise estimation using the Karhunen-Loève transform," *Comput. Cardiology*, vol. 16, pp. 269–272, 1989.

[29] A. Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ECG signals using independent components analysis," *Neurocomputing*, vol. 22, pp. 173–186, 1998.

[30] G. D. Clifford, L. Tarassenko, "One-pass training of optimal architecture auto-associative neural network for detecting ectopic beats," *Electron. Lett.*, vol. 37, no. 18, pp. 1126–1127, Aug.

2001.

[31] P. M. Agante, J. P. M. de Sá, "ECG noise filtering using wavelets with soft-thresholding methods," in *Proc. Comput. Cardiology 1999*, 1999, pp. 535–542.

[32] G. D. Clifford, A. Shoeb, P. E. McSharry, and B. A. Janz, "Model-based filtering, compression and classification of the ECG," *Int. J. Bioelectromagnetism*, vol. 7, no. 1, pp. 158–161, 2005

[33] B. Weng, M. Blanco-Velasco, and K. E. Barner, "ECG denoising based on the empirical mode decomposition," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, 2006, pp. 1-4

[34] The MIT-BIH Arrhythmia Database [Online]. Available: http:// www.physionet.org/physiobank/database/mitdb

[35] H. Liang, Q.-H. Lin, and J. D. Z. Chen, "Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 10, pp. 1692–1701, October 2005

[36] P. Flandrin and P. Gonçalves, "Empirical mode decompositions as data-driven wavelet-like expansions," *Int. J. Wavelets, Multires., Inf. Process.*, vol. 2, no. 4, pp. 477–496, 2004

[37] Y. Kopsinis, S. McLaughlin, "Empirical mode decomposition based denoising techniques," in *Proc. 1st IAPR Workshop on Cog. Inf. Process. (CIP)*, Santorini, Greece, 2008, pp. 42-47

[38] The MIT-BIH Noise Stress Test Database [Online]. Available: http://www.physionet.org/physiobank/database/nstdb/

[39] R. Bousseljot, D. Kreiseler, A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB uber das Internet," *Biomedizinische Technik*, Band 40, Erganzungsband 1 (1995) S 317

[40] D. Kreiseler, R. Bousseljot, "Automatisierte EKG-Auswertung mit Hilfe der EKG-Signaldatenbank CARDIODAT der PTB," *Biomedizinische Technik*, Band 40, Erganzungsband 1 (1995) S 319

[41] The PTB Diagnostic ECG Database [Online]. Available: http://www.physionet.org/physiobank/database/ptbdb/

[42] J. Pan, W. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. on Biomed. Eng.*, vol. 32, 1985, pp. 230-236

[43] C. Li, CX. Zheng, CF. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. on Biomed. Eng.*, vol. 42, 1995, pp. 21-28

[44] JP. Martinez, R. Almeida, S. Olmos, AP. Rocha, P. Laguna, "A wavelet-based ECG delineator: evaluation on standard databases," *IEEE Trans. on Biomed. Eng.*, vol. 51, 2004, pp. 570-581

[45] S. Tabakov, I. Iliev, V. Krasteva, "Online digital filter and QRS detector applicable in low resource ECG monitoring systems," *Annals of Biomedical Engineering*, 2008, 36(11):1805-1815

[46] Y. Liu, H. Huang, C. Weng, "Recognition of electromyographic signals using cascaded kernel learning machine," *IEEE/ASME Trans. on Mechatronics*, vol. 12, 2007, pp. 253-264

[47] G. Baudat, F. Anouar, “Generalized discriminant analysis using a kernel approach,” *Neural Computation*, vol. 12, 2000, pp. 2385–2404

[48] B. Asl, S. Setarehdan, M. Mohebbi, “Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal,” *Artificial Intelligence in Medicine*, vol. 44, 2008, pp. 51-64

[49] K. Fukunaga, “Introduction to statistical pattern recognition (2nd edition),” Orlando, FL: Academic Press, 1990

[50] K. Muller, A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, V. Vapnik, “Using support vector machines for time series prediction,” *Advances in Kernel Methods: Support Vector Learning*, Cambridge, MA: MIT Press, 1999, pp. 243-253

[51] L. Wang, B. Liu, C. Wan, “Classification using support vector machines with graded resolution,” in *Proc. Int. Conf. on Granular Computing*, vol. 2, 2005, pp. 666-670

[52] N. Cristianini, J. Shawe-Taylor, “An introduction to support vector machines and other kernel based learning methods,” Cambridge University Press, England, 2000

[53] S. Kak, “The discrete Hilbert transform,” in *Proc. IEEE*, vol. 58, 1970, pp. 585-586