2012

The severity of stages estimation during hemorrhage using error correcting output codes method

Yurong Luo
Virginia Commonwealth University

School of Engineering

Virginia Commonwealth University

This is to certify that the dissertation prepared by Yurong Luo entitled THE SEVERITY OF STAGES ESTIMATION DURING HEMORRHAGE USING ERROR CORRECTING OUTPUT CODES METHOD has been approved by her committee as satisfactory completion of the dissertation requirement for the degree of Doctor of Philosophy.

Kayvan Najarian, Ph.D., Committee Chair, Department of Computer Science

Meng Yu, Ph.D., Department of Computer Science, School of Engineering

David Primeaux, Ph.D., Department of Computer Science, School of Engineering

Rosalyn S. Hobson, Ph.D., Department of Electrical and Computer Engineering, School of Engineering

Kevin R. Ward, M.D., Department of Emergency Medicine, Univeristy of Michigan

Dr. Kathy Ryan, The U.S. Army Institute of Surgical Research

Dr. F. Douglas Boudinot, Dean of the School of Graduate Studies

---

Dr. J. Charles Jennett, Ph.D., Dean, School of Engineering

Data

THE SEVERITY OF STAGES ESTIMATION DURING HEMORRHAGE USING ERROR CORRECTING OUTPUT CODES METHOD

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

YURONG LUO
M.S., Department of Computer Science (Beijing, China), 2008
B.S., Department of Computer Science (Hebei, China), 2004

Director: KAYVAN NAJARIAN
ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE

Virginia Commonwealth University
Richmond, Virginia
August, 2012

Table of Contents

List of Tables

List of Figures

Abstract

THE SEVERITY OF STAGES ESTIMATION DURING HEMORRHAGE USING ERROR

CORRECTING OUTPUT CODES METHOD

By Yurong Luo, Ph.D. student

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2012

Major Director: Kayvan Najarian
Associate Professor, Department of Computer Science

As a beneficial component with critical impact, computer-aided decision making systems have infiltrated many fields, such as economics, medicine, architecture and agriculture. The latent capabilities for facilitating human work propel high-speed development of such systems. Effective decisions provided by such systems greatly reduce the expense of labor, energy, budget, etc. The computer-aided decision making system for traumatic injuries is one type of such systems that supplies suggestive opinions when dealing with the injuries resulted from accidents, battle, or illness. The functions may involve judging the type of illness, allocating the wounded according to battle injuries, deciding the

severity of symptoms for illness or injuries, managing the resources in the context of traumatic events, etc.

The proposed computer-aided decision making system aims at estimating the severity of blood volume loss. Specifically speaking, accompanying many traumatic injuries, severe hemorrhage, a potentially life-threatening condition that requires immediate treatment, is a significant loss of blood volume in process resulting in decreased blood and oxygen perfusion of vital organs. Hemorrhage and blood loss can occur in different levels such as mild, moderate, or severe. Our proposed system will assist physicians by estimating information such as the severity of blood volume loss and hemorrhage , so that timely measures can be taken to not only save lives but also reduce the long-term complications as well as the cost caused by unmatched operations and treatments.

The general framework of the proposed research contains three tasks and many novel and transformative concepts are integrated into the system. First is the preprocessing of the raw signals. In this stage, adaptive filtering is adopted and customized to filter noise, and two detection algorithms (QRS complex detection and Systolic/Diastolic wave detection) are designed. The second process is to extract features. The proposed system combines features from time domain, frequency domain, nonlinear analysis, and multi-model analysis to better represent the patterns when hemorrhage happens. Third, a machine learning algorithm is designed for classification of patterns. A novel machine learning algorithm, as a new version of error correcting output code (ECOC), is designed and investigated for high accuracy and real-time decision making. The features and

characteristics of this machine learning method are essential for the proposed computer-aided trauma decision making system. The proposed system is tested agasint Lower Body Negative Pressure (LBNP) dataset, and the results indicate the accuracy and reliability of the proposed system.

{CHAPTER 1 Novelty and Contribution}

The proposed computer-aided decision making system in traumatic injuries aims at estimating the severity of blood volume loss based on analysis of the physiological signals, and as such, can be used as an alarm system to be used by physicians and care givers to compensate for acute hemorrhage.

There are several systems for processing physiological signals to detect hemorrhage. Andriy [1] investigates nerve activity and heart rate variability during severe hemorrhagic shock in sheep. The explored features from R-to-R interval (RRI) of the Electrocardiogram (ECG) involve normalized high-frequency (HFnu) and low-frequency (LFnu) powers of RRI, and their ratio (LFnu/HFnu). Additionally, non-linear methods are also applied that includes RRI complexity, RRI fractal dimension etc. Another group of methodologies comes from the work of Ji and Bsoul [2,3], in which multi-signals, such as ECG, Arterial Blood Pressure (ABP), and impedance are analyzed at the same time. The features, including wavelet measures on different level of detailed coefficients and approximation coefficients, the energy of different signals, and statistical measures on spatial feature signals are extracted to represent the pattern of hemorrhage. These features are extracted from ECG, such as P-to-Q interval, S-to-T interval and so on, in both time domain and frequency domain mainly by wavelet decomposition.

The objective of our research is to design a computer-aided decision making system to estimate the severity of hemorrhage with higher prediction accuracy, near-real-time performance, and friendly human computer interface. Compared with the existing systems, the contributions of the proposed system will be demonstrated with the emphasis on the novelty of every sub-system of the proposed system as shown in Figure 1.1, which shows the general framework of the proposed computer-aided decision making system. The specific tasks in the proposed system and their novelty (contributions) are described below.

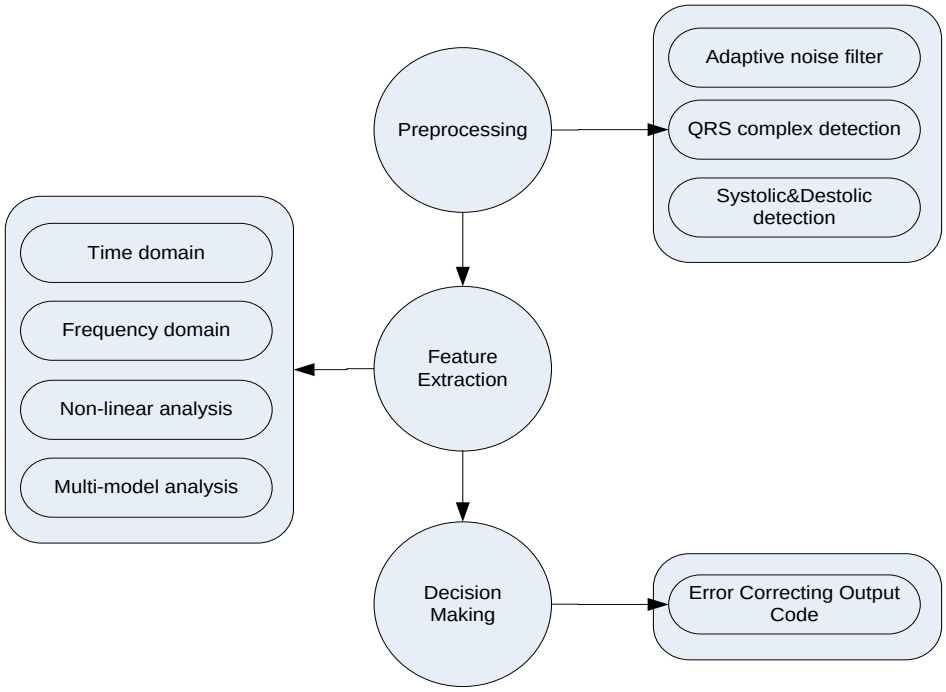


Figure 1. 1: Detailed tasks in the framework of computer-aided decision making system

1.1 Preprocessing

The proposed computer-aided decision making system is based on multiple signals. For initial investigation, two signals, i.e. ECG and ABP, which are known to be closely linked with hypovolemia, are considered. The objective of preprocessing is to eliminate signal corrupted noise and extract feature signals for further information acquisition. In our research, three novelties are incorporated into the system.

I. Design adaptive noise filter

Despite the significant noise on the biomedical signals, in particular in traumatic cases, the existing computer-aided decision making systems mainly make use of traditional filtering methods, filtering for both deterministic signal and stochastic signals, to eliminate noise. However, as to filter deterministic signal, the frequencies of the desired signals often overlap with the frequencies of the noise. Stochastic filters, on the other hand, often require that the desired signal to be a stationary stochastic process. In addition, such filters also require an approximate form of the desire signal to form the expected result. The proposed filter in this study can deal with the inputs that are deterministic or stochastic, stationary or non-stationary. Moreover, the proposed filter does not need any prior knowledge about the shape of the signals to be extracted, and as such, can be used for a variety of physiological signals. The details of designing adaptive noise filter are presented in Chapter 4.

II. Algorithm for Detection of QRS Complex and other waves

In this research, due to the central role of the QRS complex in determination of the state of the cardiovascular system, a main contribution of the proposed system is detection of the QRS complex with very high accuracy, in near-real-time and through a process that is robust against noise. Significantly different from the traditional QRS complex detection algorithms, the proposed QRS detection algorithm investigates the patterns with maximum energy of QRS complex. Hilbert transform and a novel threshold method are also applied in the proposed algorithm. Another novel step, called point insertion process, considers the variations of amplitude and period among different beats and uses these to improve the accuracy of the detection capability of the algorithm. Also, the detection of other waves in ECG signal is tried. The details of the algorithm are presented in Chapter 4.

III. Algorithm for Detection of Systolic and Diastolic Peaks

Arterial Blood Pressure (ABP) can be a suitable signal for diagnosis of blood volume loss. In the research, the proposed work develops a novel algorithm for detection of systolic and diastolic peaks. By determining the relationship among points intersected between positive and negative slopes in a feature signal extracted from the ABP, systolic and diastolic peaks are accurately detected. The advantages of the proposed method are two folds: a) It is not affected by the variations of the amplitude across the peaks, b) its fast and efficient processing performance makes it suitable for a number of real-time applications. The details of the algorithms are presented in Chapter 4.

1.2 Feature Extraction Methods

In this step, the proposed method extracts features and checks whether the extracted features can represent the pattern associated with blood loss. These features include a set of features extracted in time domain, such as statistical indicators on R-to-R interval, P-to-R interval, amplitude of R waves, the energy of the short-term signal (or the feature signal), as well as features in frequency domain, such as power spectral density, spectral analysis in the wide or selected frequency ranges, etc. Some of these features are novel, which will be discussed later. In addition, although some of other features used in this study are not novel, the bundling of these features with the novel / modified features form a pool of measures that prove to provide highest degrees of detection accuracy. One of novel / modified features introduced in this study is briefly described below.

Lempel-Ziv is a nonlinear measure that can potentially be used for complexity analysis of sequences. Biomedical signals, such as ECG and ABP during blood loss exhibits variations with nonlinear patterns. Capturing these highly nonlinear variations, in particular from ECG and heart rate variability (HRV) signals, however, calls for improved and enhanced versions of Lempel-Ziv. The proposed work evaluates the performance of the Lempel-Ziv measure for the blood loss application, identifies the disadvantages of Lempel-Ziv when addressing blood loss using ECG and HRV, and introduces an novel and improved version of the Lemple-Ziv to address these issues. The details of features extracted using nonlinear analysis are presented in Chapter 5.

In addition to specific features extracted from a single signal, the proposed method processes multiple signals and extracts features from them simultaneously. Multi-signal

processing is important vital as volume loss affect many biomedical signals at the same time and as such combined information from a variety of modalities as well as the relationship among these signals can better represent the severity of blood loss. The details of features extracted using multi-signal processing are presented in Chapter 5.

1.3 Classification Method for Decision making

In reality, many practical problems are multi-class applications. There are two prevalent models in designing classifiers for multi-class problems. One approach is one-versus-rest while the other method is all-pairs. Error correcting output codes (ECOC), a recently developed model for multi-class problems, supplies an option to deal with multi-class application effectively while increasing the flexibility in coping with multi-class problems. However, in order to make ECOC a more powerful multi-class tool, many components within the ECOC design needs further improvements. These components include: the code matrix, the encoding method, and the decoding process. Another contribution of the proposed work is the design of a new version of ECOC in which the performance, in terms of its time complexity and accuracy, is improved. The details of the novel version of ECOC are presented in Chapter 6.

{CHAPTER 2 Introduction}

2.1 Problem Statement

2.1.1 Definition

Trauma is a major cause of death and disability across the world. The events recognized as trauma include natural disasters (earthquakes, fires, floods, hurricanes, etc), physical assault (rape, incest, molestation, and domestic abuse), serious bodily harm, serious accidents (automobile, high-impact scenarios), falls or sports injuries, etc. The severity of the injury is directly related to the force of the impact, duration of impact, body part involved in the impact, injuring agent (blast, blunt or penetrating object), and any associated risk factors (age, preexisting medical conditions). Hemorrhage Shock (HS), a condition of depressed body functions as a reaction to trauma with loss of body fluids or lack of oxygen, is a potentially life-threatening condition that requires immediate treatment. Delay and lack of sufficient detailed information in dealing with hemorrhage will seriously threaten human life. While nowadays there are many measurement systems available to care givers to monitor patient's physiological signals, there is an urgent need for computational methods to process these signals and form informative recommendations based the details in the data not visible to the human eyes. The objective of the proposed research is to process physiological signals in order to provide assistance to physicians by detecting the existence and severity level (mild, moderate,

severe) of blood loss. This would allow physicians / medical staff to take appropriate measures in a timely manner. This will not only save lives but also help alleviate potential shot-term and long-term complications / disabilities. The system will also help cut down the cost of providing care and treatments to patients.

2.1.2 Motivation

Trauma is a global problem with high risk and high prevalence. Across the world, over 5 million people die every year (almost 16,000 persons a day) as a result of accidental or intentional traumatic injuries and around half a billion survivors of trauma suffer from a physical or psychological trauma as a result [4]. Motor vehicle accidents are one cause of trauma where the rate of fatalities is falling in many of the world's richest countries; however this rate is increasing rapidly in most developing countries where more than 80% of the world's population reside. Yet slightly less than 20% of the 5 million deaths from all types of injury that occur each year happen in the high-income 'developed' countries, which makes trauma a major issue for these nations. Age variations reveals that, in 2002, deaths from motor vehicle accidents were the second highest cause of death globally in the age-groups 5-14 and 15-29 years and the third highest cause of death among 30-44 year olds.

Trauma annually impacts hundreds of thousands of individuals and costs billions of dollars in direct expenditures and indirect losses. In 2000 alone, the 50 million injuries that required medical treatment ultimately cost $406 billion [5]. This includes estimates of $80.2 billion in medical care costs and $326 billion in productivity losses. Persons aged 25

to 44 years represented 30% of the U.S. population and 40% ($164 billion) of the total lifetime costs of injuries that occurred in 2000. Motor vehicle and fall injuries account for 22% ($89 billion) and 20% ($81 billion) of the total lifetime costs of injuries that occurred in 2000. Upper extremity and lower extremity injuries each account for 17% ($68 billion) of the total lifetime costs of injuries that occurred in 2000.

Considering the data given above, a computer-aided trauma decision making system that can help reduce mortality, long-term complications, and the cost associated with trauma will have a tremendous impact on the human life.

2.2 Triage for Trauma Systems

2.2.1 Introduction

This study focuses on triage and assisting caregivers to optimize their effort throughout this process. During a typical triage process, the level of severity of injuries for the patients is "scored" (or "sorted") on the basis of the actual or perceived degree of injury and they are assigned to the most appropriate care resources according to the score, in order to ensure optimal care and the best chance of survival. Generally speaking, the necessity of triage lies in the following aspects: 1) achieve a more optimal performance in allocating the resources used for trauma care, 2) cut down the time spent to take these measures, resulting in the increased survival rate.

2.2.2 Existing Trauma Triage Systems

An efficient "scoring" or "sorting" of potentially injured patients translates into customized care, better outcomes as well as improved resource utilization [6]. As

discussed above, a major complication in all traumatic injuries is hemorrhage and in particular management of hemorrhagic shock. The detection and scoring of hemorrhage which forms the foundations for any response to hemorrhagic shock is complex, under-investigated and influenced by the dynamics and magnitude of volume depletion. Next, the existing systems for detection of hemorrhage are briefly reviewed.

Heart rate variability (HRV) [7,8] is a physiological signal that is formed based on the variations of the time interval between consecutive heart beats. Other similar concepts used include: "cycle length variability", "heart period variability", "RR variability", where R is the time point corresponding to the peak of the R wave in the QRS complex in the ECG signal, and RR is the time interval between two successive R's. The most widely used methods of extracting characteristic features from HRV can be categorized into time-domain and frequency-domain methods. Other methods involve using different approached in non-linear analysis.

As to time domain analysis, in [9] the standard deviation (SD) of average R-R intervals (RRISD), the square root of the mean squared differences of successive R-R intervals (RMSSD), heart rate, and the percentage of adjacent normal R-R intervals (pNN50) are calculated. For validation and comparison, two groups (patient and healthy) were formed. The results from time-domain analyses showed that heart rates (105 ± 6 versus 90 ± 6) and R-R intervals (605 ± 37 versus 712 ± 46), as well as global measures of HRV, including RRISD (58 ± 14.6 versus 47 ± 7), RMSSD (53 ± 12 versus 43 ± 7), and pNN50 (13 ± 5 versus 13 ± 4), are statistically indistinguishable across the two groups. In order to analyze HRV

in time domain the root-mean-square successive difference (r-MSSD), which calculates the square root of the mean of the squared differences between successive beat-to-beat intervals over twenty four hours, and the mean of all the five minute standard deviations of beat-to-beat (SDNN) were also proposed. In [10], geometric measures, HRV triangular index (or RR triangular index) and the triangular interpolation of RR interval histogram (TINN) are employed as the criteria to predict patients' outcome.

As to the frequency analysis, Andriy investigates nerve activity and HRV during severe hemorrhagic shock in animal models (sheep) [1] . The explored features based on the R-to-R interval (RRI) extracted from the ECG involve the normalized high-frequency (HFnu) and low-frequency (LFnu) powers of RRI as well as their ratio (LFnu/HFnu). In 2006, Cooke et al. demonstrated that an increase in the HF/LF ratio was associated with increased mortality in a series of 42 patients transported by helicopter to a trauma center [11]. Batchinsky et al. In [12], it showed that high frequency amplitude (HFA) distinguishes survivors from non-survivors with a data set as small as 100 beats. Ji and Bsoul build their model based on the wavelet transform. The features used in these studies include various statistical measures on different levels of both detailed and approximation coefficients formed by the wavelet transform.

For non-linear analysis, all types of entropy metrics of heart rate variability [13] are introduced, such as approximate entropy, sample entropy, multiscale entropy, similarity of distributions, etc. In Andriy's model, some features based on non-linear methods were used. The complexity of RRI was measured by the approximate entropy (ApEn) and

sample entropy (SampEn) methods, while the RRI fractal dimension was measured by curve length (FDCL) approach. To better represent the patterns of hemorrhage in trauma, some studies combine the features coming from multi-signals, such as arterial blood pressure (ABP), impedance, etc.

2.3 Proposed Computer-aided Decision Making System

In the proposed system:

1. new pre-processing systems are proposed to remove the noise.
2. more effective methods to detect QRS complex and Systole&Diastole are introduced.
3. better and more features are employed.
4. finally, a novel machine learning algorithm with superior classification accuracy and reliability is designed. The resulting system is expected to outperform the existing computer-aided trauma decision making systems for detection of the existence and severity of blood loss.

2.4 Summary

In this chapter, backgrounds to trauma care and trauma decision making systems for blood loss were presented. Also, the main advantages and shortcomings of some of the existing computer-aided decision making systems were briefly reviewed. Finally, the main objectives of the proposed systems in relation to the existing systems are outlined.

{Chapter 3: Proposed Method; Overview}

3.1 Proposed Computer-aided Decision Making System

The main objective of proposed computer-aided decision making system is to estimate the existence and severity of blood volume loss. The overall schematic diagram of the proposed computer-aided decision making system is presented in Figure 3.1. Each element of this diagram is briefly described below.

I. System Input:

The inputs of the system are physiological signals, namely ECG and ABP. In this study, the data for training and testing of the system come from U.S. Army's Lower Body Negative Pressure (LBNP) system that simulates blood loss [14]. The LBNP system has been consistently used by the community as an accurate model of the hemorrhage and has proved to a suitable system for such studies.

II. Computational Algorithms:

The computational algorithms include the steps shown in the schematic diagram (Figure 3.1). The first step is denoising and extraction of features from physiological signals. The features based on time domain, frequency domain, non-linear analysis, and multi-signal analysis are extracted to form informative patterns representing the existence and severity of volume loss.

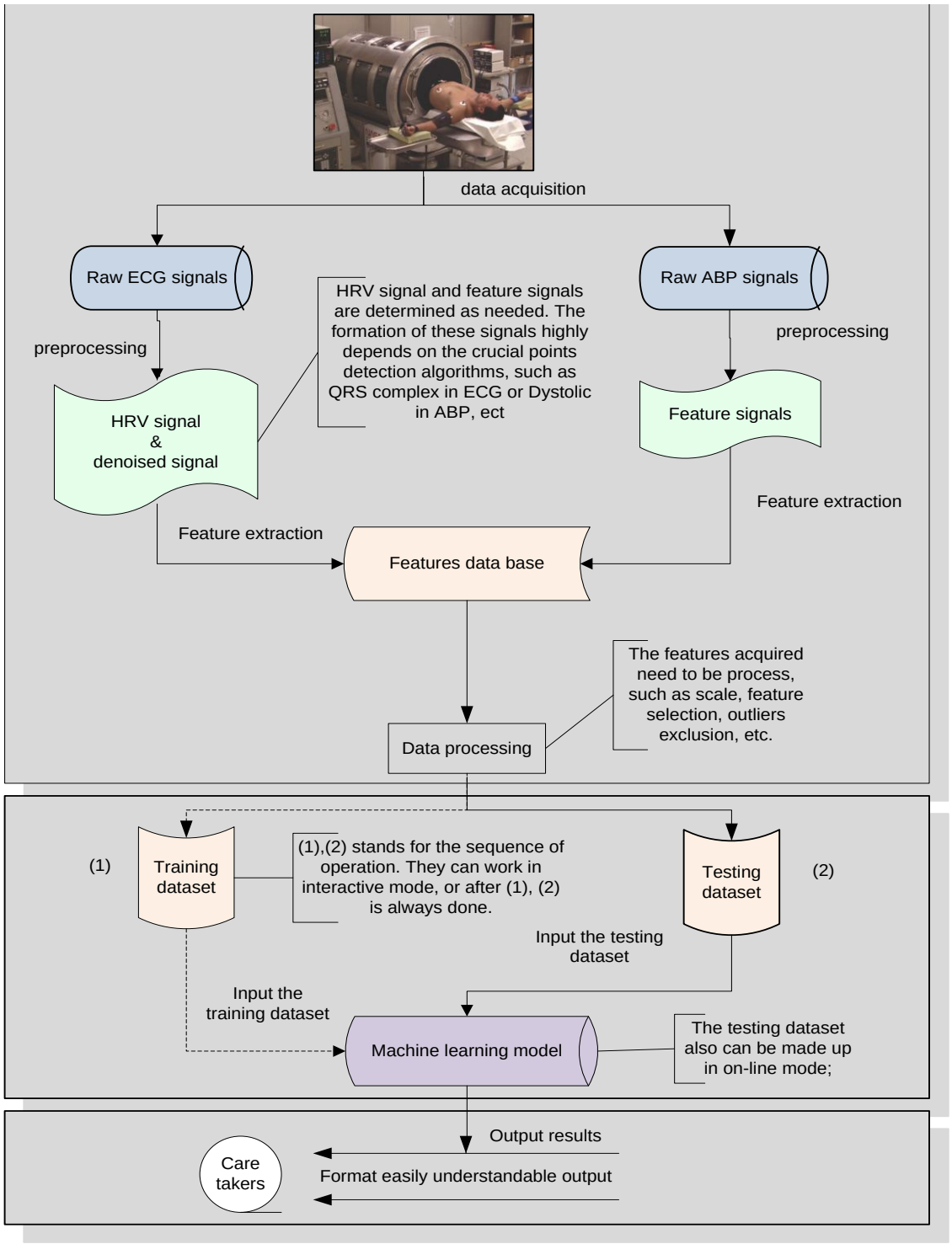


Figure 3. 1: Schematic diagram of the proposed computer-aided decision making system.

The formed features are further processed through operations such as scaling, feature selection, outlier exclusion, etc.

Once the optimal feature set is formed, the novel machine learning tool is trained and tested to detect volume loss. The machine learning method is an improved version of ECOC that provides superior accuracy and reliability for multiclass classification scenarios. The system can operate in two different modes. One is to train the classifier offline (and only once), while in the other mode the classifier can be interactively trained. In either case, the system is tested using a set of new data.

After building the model and all computational methods for the proposed system, an intuitive and user-friendly interface will be designed to allow users interact with the system and provide the computational methods the means to present the results to the care givers.

III. System Output:

The output of the system is the severity of blood loss. The proposed system classifies the degree of volume loss into three different levels: mild, moderate, and severe. The criterion for dividing into different levels for volume loss resorts based on the mapping between the stages of LBNP and the above mentioned three levels for each individual. This mapping, which will be discussed later, has been formed by collaborating physicians and physiologists based on the collapse stage of each individual during the LBNP experiment.

The details for each of the elements discussed above will be provided in the following chapters.

{CHAPTER 4 Preprocessing}

4.1 Overview

The objective of preprocessing is two folds: one is to eliminate noise that corrupts the signal, while the other is to extract characteristics features from the signal, which sometimes involves forming derivate feature signals from the original signal. Different sources of noise [15] corrupt the essence of the signal, resulting in inaccurate information. In ECG signal, for example, the typical sources of noise include power line interference, motion artifacts and baseline drift. As to feature signals, the typical signals extracted from ECG includes RR signal (signal composing of the timing between consecutive R points in ECG), heart rate variability (HRV), and so on. In the following sections, first the main above-mentioned preprocessing methods often used for ECG analysis are briefly reviewed and then the proposed pre-processing algorithms are presented.

4.2 Related Work

4.2.1 Adaptive filter

Biomedical signals are contaminated by several types of noise and interference. Filtering the noise is essentially important for acquiring accurate information. While some traditional filters are designed for pre-processing of deterministic signals, others deal with filtering of stochastic signals. The disadvantage of the former lies in the fact that the frequencies of the signals usually overlap with the frequencies of the noise. And the better result comes from the available knowledge about the deterministic signals. As such

stochastic filters are prevalently used in the field of biomedical signal processing. However, such filters often require some prior knowledge about the signal and make some assumptions on the nature of signal and noise. The most popular stochastic filters are Wiener filter and Kalman filter.

I. Wiener filter

Wiener filter [16] defines a class of optimal linear filters that involves linear estimation of a desired signal or sequence $d(n)$ from another related sequence $x(n)$ (Fig.4.1). $y(n)$ and $e(n)$ are filter output and the estimation error, respectively. $W(z)$ is the Z-transform of the filter coefficients $W = [w_0 w_1 \cdots w_{N-1}]^T$.

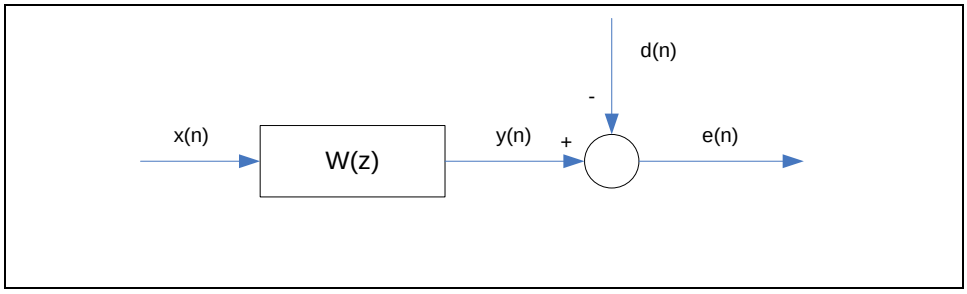


Figure 4. 1: The block schematic of Wiener filter

Wiener filters are characterized by the following:

--- Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross correlation.

--- Requirement: the filter must be physically realizable / causal.

--- Performance criterion: minimum mean-square error (MMSE).

In the following part, a briefly deduction about the Wiener filter is given. Let $W = [w_0 w_1 \cdots w_{N-1}]^T$ and $X = [x_n x_{n-1} \cdots x_{n-N+1}]^T$. Then, the output is $y(n) = X^T W$. The performance function, or cost function, is then given by

$$\begin{aligned} \xi &= E(e^2(n)) = E((d(n) - y(n))^2) \\ &= E[d^2(n)] - W^T E[X(n)d(n)] - E[X^T(n)d(n)]W + W^T E[X(n)X^T(n)]W \end{aligned} \quad (1)$$

For convenience, the cross-correlation vector, P, and autocorrelation matrix, R, are defined as:

$$P = E[X(n)d(n)]$$

And,

$$R = E[X(n)X^T(n)]$$

To obtain the set of weights which minimizes the performance function, the derivatives of the cost function are set to zero, i.e. $\frac{\partial \xi}{\partial w} = 0$. Finally, the optimal weight vector is obtained as:

$$W_{op} = R^{-1}P \quad (2)$$

This form of solution for the Wiener filter is also known as the Wiener-Hopf equation.

II. Kalman filter

The Kalman filter [17] is an efficient recursive filter that estimates the internal state of a linear dynamic system from a series of noisy measurements. It is used in a wide range of engineering and economic applications from radar and computer vision to estimation of structural macroeconomic models, to control theory and control systems engineering. As

in the Wiener filter that uses both auto correlation and the cross correlation of the received signal with the original data to optimize a cost function, the Kalman filter presents a prescription of the optimal MSE filter. However, Kalman describes the filter using state space techniques. Another difference is that the Kalman filter does not assume the signal to be stationary.

Kalman filters are characterized by the following:

---Assumption: the average value of the noise is zero; no correlation exists between the noise sources; the noise covariances are known.

--- Performance criterion: minimum mean-square error (MMSE).

Next, a brief description of the Kalman filter is provided. A linear state space system is a process that can be described by the following two equations:

$$\text{State equation: } x_{k+1} = \Phi x_k + w_k \tag{3}$$

$$\text{Output equation: } z_k = H x_k + v_k \tag{4}$$

where $x_k\,(n\times 1)$ is the state vector of the process at time k; Φ $(n\times n)$ is the state transition matrix of the process from state k to the state at k+1 which is assumed stationary over time; $w_k\,(n\times 1)$ is a white noise process with a known covariance; $z_k\,(m\times 1)$ is the actual noisy measurement of x at time k; H $(m\times n)$ is the noiseless connection between the state vector and the measurement vector, which is assumed stationary over time; and v_k (m$\times$1) is the measurement error.

The covariance of the two noise models are assumed stationary over time and are given by: $Q = E[w_k w_k^T]$ and $R = E[v_k v_k^T]$. The mean square error is then given by: $P_k = E[e_k e_k^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ ($\hat{x}_k$ is the estimate of x_k). Denoting the prior estimate of $\hat{x}_k$ as $\hat{x}_k^{"}$, it is possible to write an updated equation for the new estimate, combing the old estimate with measurement data as:

$$\hat{x}_k = \hat{x}_k^{"} + K_k(z_k - H\hat{x}_k^{"}) \tag{5}$$

where K_k is the Kalman gain. After substitution of the equation into the mean square error function:

$$P_k = P_k^{"} - K_k H P_k^{"} - P_k^{"} H^T K_k^T + K_k(H P_k^{"} H^T + R) K_k^T \tag{6}$$

Differentiating $T[P_k]$ with respect to K_k gives: $\frac{dT[P_k]}{dK_k} = -2(HP_k^{"})^T + 2K_k(HP_k^{"}H^T + R)$, where T means the trace of the matrix. After setting it to zero, finally:

$$K_k = P_k^{"} H^T (H P_k^{"} H^T + R)^{-1} \tag{7}$$

Again, substituting K_k into the new form of P_k , gives $P_k = (1 - K_k H) P_k^{"}$. Summarizing the process, the iterative Kalman filter follows the equations in Table 4.1.

Table 4. 1: Kalman Filter Recursive Algorithm

Description	Equation
The covariance of noise models	$Q = E[w_k w_k^T]$ and $R = E[v_k v_k^T]$
Update equation for the new estimate	$\hat{x}_k = \hat{x}_k^{"} + K_k(Hx_k + v_k - H\hat{x}_k^{"})$
State projection	$\hat{x}_{k+1}^{"} = \Phi \hat{x}_k$
The error covariance matrix	$P_k = (1 - K_k H)P_k^{"}$
The prior estimate P_k	$P_{k+1}^{"} = \Phi P_k \Phi^T + Q$
The Kalman gain	$K_k = P_k^{"} H^T (HP_k^{"} H^T + R)^{-1}$

III. Adaptive filter

As it can be seen from the above discussion, both the Wiener filter and the Kalman filter need to satisfy some assumptions and rely on some prior knowledge. Some of the knowledge can be estimated from the improved version of the Wiener filter and Kalman filter, which may not be easy to satisfy in many practical applications.

Widrow [18] put forward an alternative method for filtering noise, which is called adaptive filter (Fig.4.2). The method uses a "primary" input containing the corrupted signal and a "reference" input containing noise correlated with the primary noise. The reference input is adaptively filtered and subtracted from the primary input to obtain the signal

estimate. Before subtraction, adaptive filtering allows the treatment of inputs that are deterministic or stochastic, stationary or time variable.

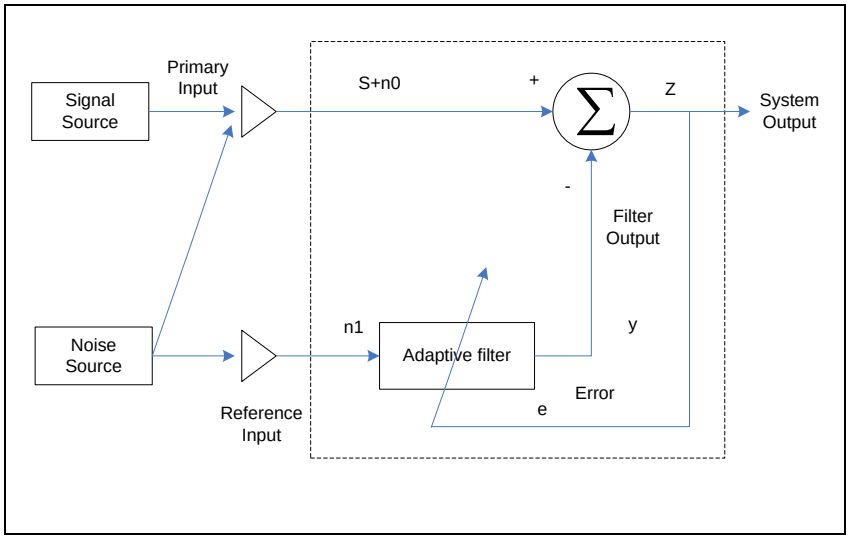


Figure 4. 2: The adaptive noise cancelling concept

As Fig 4.2 shows, a signal s is corrupted with a noise n_0 , which is uncorrelated with the signal. The combined signal and noise $s+n_0$ forms the primary input to the filter. A second input receives a noise n_1 uncorrelated with the signal but correlated, in some unknown way, with the noise n_0 . The noise n_1 is filtered to produce an output y that is as close to as to n_0 as possible. This output is subtracted from the primary input $s+n_0$ to produce the system output $z=s+n_0-y$.

Adaptive filters are characterized by the following:

---Requirement: little or no prior knowledge of s , n_0 , n_1 , or of their interrelationships, is required.

---Performance: Least Mean Squares (LMS).

The adaptation algorithm, LMS, is given by the following equations:

$$y_k = w_k^T x_k \tag{8}$$

$$\xi_k = d_k - y_k \tag{9}$$

$$w_{k+1} = w_k + \alpha e_k x_k \tag{10}$$

where w is the weight of the filter, k is the time index, α is the adaptation constant that determines the size of the step taken at each iteration along the estimated negative. LMS is an iterative gradient (e_k) descent algorithm that uses an estimate of the gradient on the mean square error surface to seek the optimal vector with respect to the minimum mean square error.

Next, a very description of the adaptive filter works is given. Squaring the output $z = s + n_0 - y$ gives $z^2 = s^2 + (n_0 - y)^2 + 2s(n_0 - y)$. Taking expectations of both sides of z^2 yields:

$$E[z^2] = E[s^2] + E[(n_0 - y)^2] + 2E[s(n_0 - y)] \tag{11}$$

Since s is uncorrelated with n_0 and y:

$$E[z^2] = E[s^2] + E[(n_0 - y)^2] \tag{12}$$

Accordingly, the minimum output power is $\min(E[z^2]) = E[s^2] + \min(E[(n_0 - y)^2])$. This means that when y is approaching to n_0, the output of the filter achieves desirable performance.

The applications of adaptive filter include adaptive noise cancellation, inverse plant modeling, adaptive inverse control, adaptive equalization, adaptive linear prediction, and

nonlinear filtering and prediction. For adaptive noise cancelling, if the noise is fixed in a fixed frequency, adaptive filter will act as notch filter [19,20]. In the later developments of adaptive filter, P. Laguna [21] designed a new version of adaptive filter (Fig 4.3), celled Adaptive Impulse Correlated Filter (AICF), in which in order to filter an event-related bioelectric signals d_k (such as ECG), an impulse correlated reference input x_k was used. Unlike the original adaptive filter (Fig 4.2), the reference signal was not correlated with the noise, but it was correlated with the deterministic component in the primary signal. For ECG, x_k is the actual signal, composed of a train of impulses positioned at the beginning of each pulse. This method shows superiority when referring to event-related bioelectric signals. However, the limitation of the method is that it assumes that the event-related bioelectric signals are periodic.

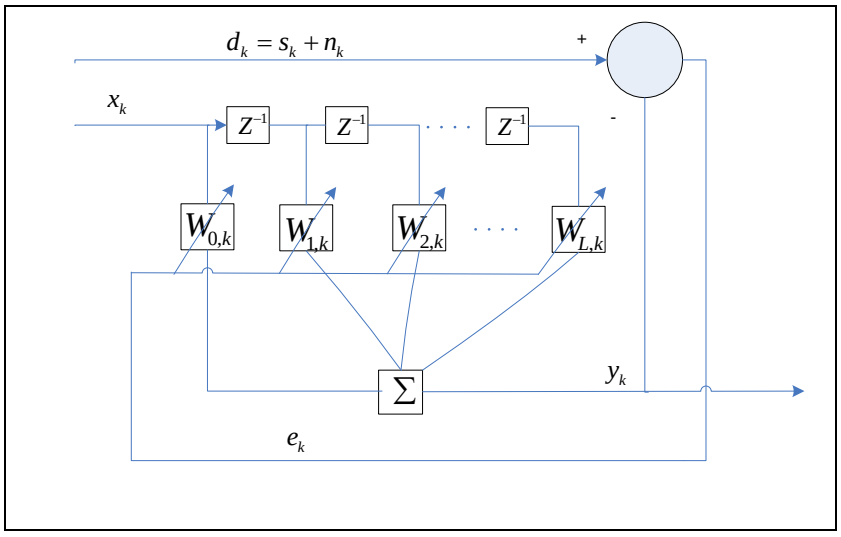


Figure 4. 3: Adaptive impulse correlated filter

Thus, despite the capabilities of the filters described above, there are some outstanding problems that need to further be studied and addressed, in particular along the promising ideas of adaptive filter (Fig.4.2). For example, usually in primary signal, there are uncorrelated noise with reference signal, and in reference signal, there are signal correlated component. These factors will influence the effect of adaptive filter.

4.2.2 QRS complex detection

One of the most widely used signals in clinical practice is Electrocardiogram (ECG). ECG reflects the electrical activity within the heart and the time of its occurrence. A typical ECG cycle consists of P, QRS, and T waves (Fig 4.4). Irregular shapes of these waves are associated with complications and illnesses. These wave irregularities include shortened or prolonged QT interval, flattened or inverted T waves, prolonged T waves, prolonged S waves, missing P or T waves, and many small or large ripples. As a result, in order to detect these irregularities and diseases associated with them, it is critical to develop accurate method for detection of P, QRS and T waves.

QRS detection algorithms [22,23] have been developed for several decades. The algorithms based on wavelet transform [24] are commonly used because they extract the most possible candidates as QRS wave from the original signal by applying different wavelet functions such as Daubechies 6 (db6), Haar, db2, db3, or even customized wavelets according to specific application . Also, wavelet methods remove noise by extracting fixed frequency signal from original signal. QRS detection algorithms based on the characteristic of signal shape are also effective methods for QRS detection [25,26,27,28].

These methods that acquire QRS complex according to the shape of ECG, usually use derivative approach. However, the disadvantage of these methods is that they require filtering noise so aggressively that they may affect the actual signal. Neural-Network-based methods for QRS complex analysis [29] train artificial neural networks (ANN) with digitized ECG signal samples, and use the trained ANN for testing. Also, algorithms based on more sophisticated digital filters [30], filter bank [31], fuzzy methods [32], Hilbert Transform [33], syntactic method [34], mathematical morphology [35,36,37], and some other hybrid algorithms [38,39,40,41] have been introduced.

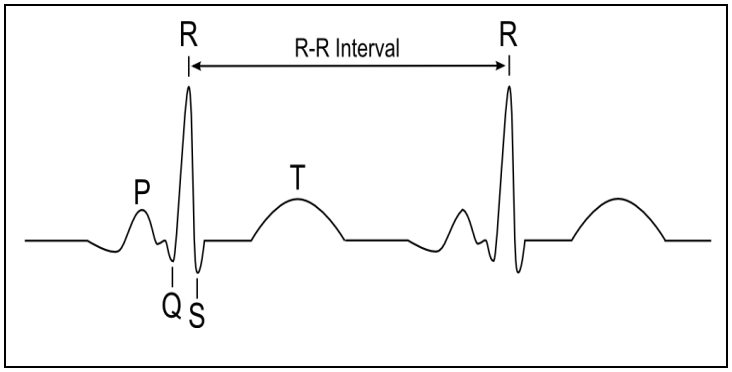


Figure 4. 4: The components in ECG signal

Generally, the common characteristic of these methods is using the following two steps: a) getting feature signal from the original ECG signal, and b) comparing the feature signal with a fixed or adaptive threshold. The efficiency and creativity of these methods lies mainly in the first step. Often, some post-processing methods are adopted to achieve better result.

4.2.3 Systole&Diastole Detection Algorithm

To the best of our knowledge there is no systematic signal processing method to detect the systolic and diastolic peaks from Arterial Blood Pressure (ABP). Therefore, in this study, by observing the characteristics of ABP signal, a new algorithm which we refer to as Systole&Diastole detection algorithm, is desired for this purpose. This algorithm will be described later.

Arterial Blood pressure (Fig 4.5) is produced when the heart pumps blood into the arteries of the body. As the heart beats, the pressure rises to a maximum level, called the systolic blood pressure, and as it relaxes, it falls to a minimum level called the diastolic pressure. Blood pressure is expressed as Systolic value over Diastolic value.

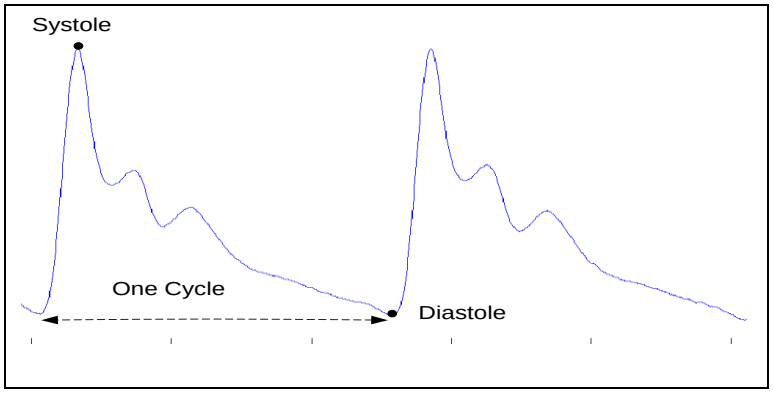


Figure 4. 5: The components in ABP signal

ABP signal has multiple diagnostic values in traumatic injuries, for instance: a) while the amplitude of Systolic and Diastolic values varies even in healthy subjects, and the degree of these variations is larger closer to onset to hemorrhagic shock; b) there may be some small “ripples” between Systolic and Diastolic peaks during hemorrhage.

4.3 Proposed Method

4.3.1 Noise removal

Through analysis and experiment towards adaptive filter, the feasibility dwells in notch filter using adaptive principle. The typical noise in ECG is baseline wander. In the proposed research, the realization of noise removal depends on adaptive notch filter and indepedent component analysis (ICA). Using ICA to remove noise is based on the assumption that the ECG baseline wander comes from an independent and unkonwn source. Experimental results in Chapter 7 prove the correctness of the assumption.

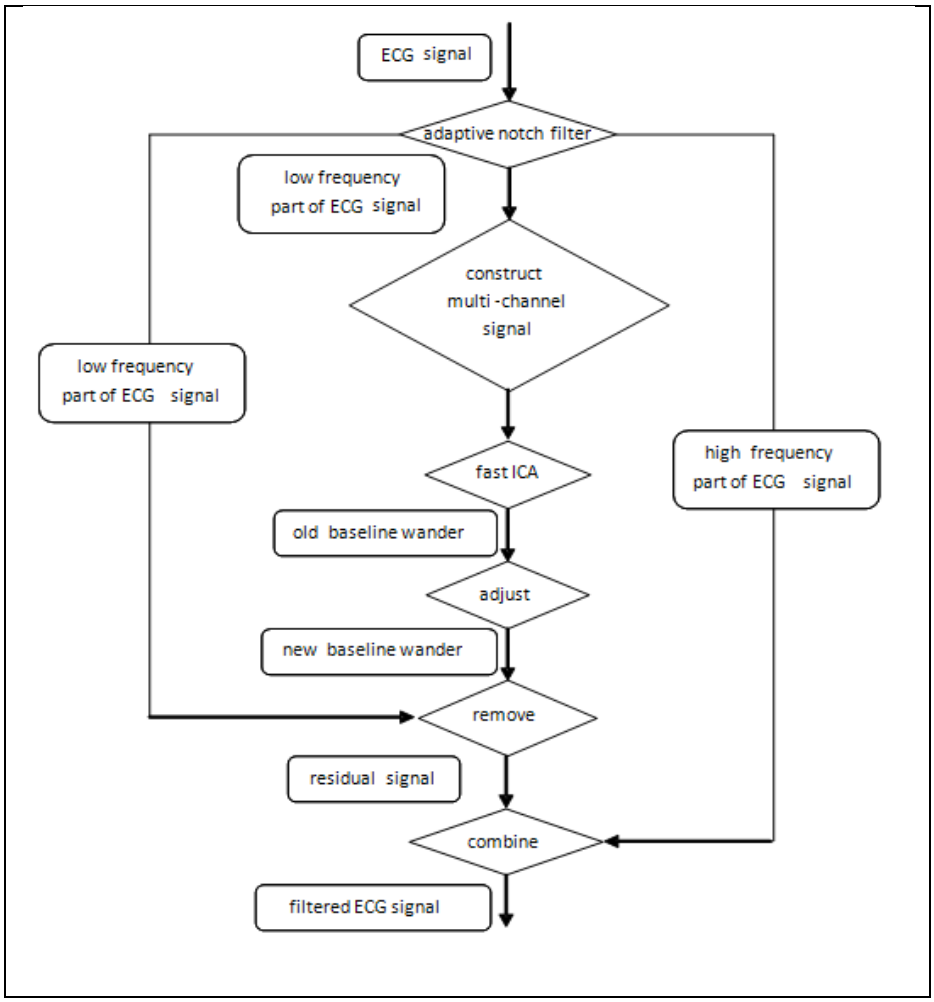


Figure 4. 6: Schematic diagram of noise removal

Figure 4.6 shows the framework of the proposed method. As it can be, the first step of the proposed method is an adapative notch filter, desinged to form sub-signals of the ECG, as described later. Next, the proposed method utilizes ICA to remove the baseline drift. Finally, the independent component formed by the ICA as the output, which is originally labeled as the baseline wander, needs to be further adjusted to form a better estimate of the baseline wander. The advantages of the proposed method include the following aspects: 1) it reduces the errors produced by the overlapping in frequency between baseline wander and low frequency part of ECG; 2) it also cut down the errors induced by estimating independent component in ICA through using just low frequency signal.

4.3.1.1 Adaptive notch filter

The adaptive notch filter is based on the same theoretical foundations as the adaptive noise cancellation. As to the adaptive notch filter, the reference signal is the signal with one or multi fixed frequencies, which are treated as the frequencies to be excluded.

The advantages of adaptive notch filter lie in the following aspects: 1) if the frequency of the interference is not precisely known, or the interference drifts in the frequency, the exact excluded frequency could be measured / adapted to during the filtering process; 2) the filter is tunable since the null point moves with the reference frequencies; 3) the adaptive notch filter can be made very sharp at the reference frequency; 4) through adjusting the parameters, the adaptive notch filter can be considered as a time-invariant filter by lessening the influence of the time varying components. The diagram of adaptive

noise cancelling is shown in Fig 4.7. The system is an N -stage tapped delay line (TDL). The weight of the filter is updated according to the following equations:

$$y_k = w_k^T x_k$$

$$\varepsilon_k = d_k - y_k$$

$$w_{k+1} = w_k + \partial \varepsilon_k x_k \tag{13}$$

where x is the reference input; d is the desired response; y is the output of the filter; w is the weight of the filter; ∂ is the adaptation constant; k is the time index. As described, the response from $E(z)$ to $Y(z)$ includes two parts. In practical applications, it is feasible to make the time-varying component to be insignificant (β/N ≈0) by changing the values of N and setting β as:

$$\beta = \frac{\sin(Nw_r T)}{\sin(w_r T)} \tag{14}$$

where w_r is the frequency of the interference. If the reference input is considered to be the following form:

$$x = C\cos(w_r T + \theta) \tag{15}$$

The transfer function of adaptive notch filter can be expressed as:

$$H(z) = \frac{z^2 - 2z\cos(w_r T) + 1}{z^2 - 2\left(1 - \frac{N\partial C^2}{4}\right)z\cos(w_r T) + \left(1 - \frac{N\partial C^2}{2}\right)} \tag{16}$$

Therefore, the parameter N can be set to the fixed value as described above. It can be seen that the above-mentioned filter is very flexible and can be adjusted using the

adaption constants ∂ and C to provide the desired bandwidth and depth of a suitable notch filter.

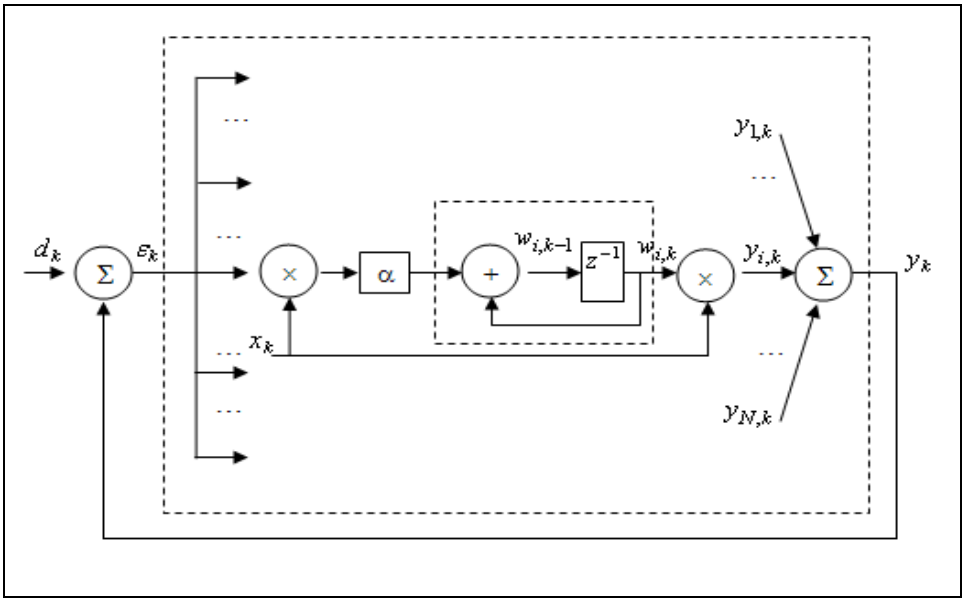


Figure 4. 7: The diagram of adaptive noise cancelling

4.3.1.2 Independent Component Analysis

After applying the notch filter, the main step used is ICA. First, the "standard" ICA is described. ICA can be briefly explained using a simple example of separating two source signals $s_1(t)$ and $s_2(t)$ that were mixed by an unknown linear process. Two different linear mixtures, $x_1(t)$ and $x_2(t)$, are given as:

$$x_1(t) = c_{11}s_1 + c_{12}s_2 \tag{17}$$

$$x_2(t) = c_{21}s_1 + c_{22}s_2 \tag{18}$$

where c_{11}, c_{12}, c_{21} and c_{22} are unknown coefficients. The objective of the problem is to recover the signal and $s_2(t)$ from mixture signals $x_1(t)$ and $x_2(t)$ without knowing any

prior information about the source signals $s_1(t)$ and $s_2(t)$ and the mixing process (i.e. c_{11}, c_{12}, c_{21} and c_{22}), except that $s_1(t)$ and $s_2(t)$ are statistically independent.

In the generalized case, where there are more latent sources and more mixture signals, the formal definition of ICA is as follows:

$$x_i(t) = c_{i1}s_1 + c_{i2}s_2 + \cdots + c_{in}s_n \qquad i \in [1, n] \tag{19}$$

where $s_i(t)$ is called latent source, $x_i(t)$ is the mixture signal, c_{ij} is the mixing coefficient between $x_i(t)$ and $s_j(t)$, and n is the number of latent sources and mixture signals. The above formulation can be expressed as the following matrix form:

$$X = C_{n\times n} \cdot S \tag{20}$$

where X is the matrix of mixture signals, in which each column is one mixture signal; S is the matrix of latent signals, in which each column is one latent signal; and $C_{n\times n}$ is the matrix for mixing coefficients.

The feasibility of solving the ICA problem lies in the condition that the latent sources are independent from each other. According to the Central Limit Theorem, the distribution of a sum of independent random variables approaches a Gaussian distribution. This implies that the solution of ICA can be achieved when distribution diverges from Gaussianity. The deviation from Guassianity can be measured using measures such as Negentropy.

Negentropy is one measure of nongaussianity defined based on the concept of Entropy, which is the fundamental concept of the information theory. Entropy, E, as a measure of information in random variables is defined for a discrete random variable y, as:

$$E(y) = -\sum_i P(y = c_i) \log P(y = c_i) \tag{21}$$

where c_i is the possible values of Y, and $P(Y = c_i)$ means the probability when the value of Y is c_i. For a continuous random variable y, entropy E is defined as the following equation:

$$E(y) = -\int f(y) \log(f(y)) dy \tag{22}$$

where f is the probability distribution function. Negentropy, J, is then defined as follows:

$$J(y) = E(y_{gauss}) - E(y) \tag{23}$$

where $y_{gaussian}$ is a Gaussian random variable with the same covariance matrix as y. A fundamental conclusion in information theory is that a Gaussian variable has the largest entropy among all random variables of equal variance. Hence, negentropy is always non-negative, and it is zero only if Y has a Gaussian distribution.

The exact calculation of negentropy requires an accurate estimation of the probability distribution function, which may be computationally costly or data intensive. Therefore, it is often preferred to find simple approximations of negentropy. Simple approximations of negentropy have been introduced [42], which are based on the maximum entropy principle. In general, the following family of approximations is the most commonly used group:

$$J(y) = \sum_{i=1}^{p} k_i [E(G_i(y)) - E(G(v))]^2 \tag{24}$$

where k_i are constants, v is a gaussian random variable with zero mean and unit variance. Often, the value of p and k_i can be set to one. Therefore, the above formulation becomes:

$$J(y) = [E(G(y)) - E(G(v))]^2 \tag{25}$$

The following formulations of G functions have proved very useful in practical applications:

$$G_1(y) = \frac{1}{a_1}\log\cosh(a_1 y) \qquad g_1(y) = \tanh(a_1 y) \tag{26}$$

$$G_2(y) = -\frac{1}{a_2}\exp(-a_2 y^2/2) \qquad g_2(y) = y\exp(-a_2 y^2/2) \tag{27}$$

$$G_3(y) = \frac{1}{4}y^4 \qquad g_3(y) = y^3 \tag{28}$$

where $1 \le a_1 \le 2$, $a_2 \approx 1$,and g is the first derivative of the function G.

Before applying the main processing operations of the ICA, it is often necessary to perform some preprocessing. Usually, the two different operations are conducted: centering and whitening. Centering requires that the random variable y is a zero-mean random variable and it is performed by subtracting its mean vector. Whitening will make the random variable uncorrelated and set their variances equal to unity by using the eigenvalue decomposition of their covariance matrix:

$$E\{yy^T\} = DVD^T \tag{29}$$

where D is the orthogonal matrix of eigenvectors, and V is the diagonal matrix of eigenvalues. Now, assuming that z is a new random variable after whitening:

$$z = DV^{-1/2}D^T y \tag{30}$$

Whitening makes the problem change from estimating mixing matrix to estimating a new one $\tilde{C}$:

$$z = DV^{1/2}D^T Cs = \tilde{C}s \tag{31}$$

Among several improvements of ICA, Fast and Fixed Point Independent Component Analysis [43], as a direct extension of the standard ICA, was developed for calculating latent sources with high speed. The basic rule of fast and fixed point independent component analysis is to find a direction, which can maximize nongaussianity of $w^T x$. Nongaussianity is decided according to the approximation of nongaussianity as mentioned above. The following is the basic description of the algorithm:

1. Initialize a weight vector w in one direction.
2. Change the weight vector according to the following criteria:

   $w^{'} = E\{xg(w^T x)\} - E\{g^{'}(w^T x)\}w$,

   and normalize the weight vector as: $w = w^{'}/\|w^{'}\|$
3. If the weights have not converged, go back to step 2.

where w is the weight vector to calculate latent source $s = w^T x$, and convergence means that the old weight vector and the new weight vector are in the same direction. In this study, the Fast and Fixed Point Independent Component Analysis is used as the implementation of ICA block.

4.3.2 QRS complex detection

As discussed in 4.2.2 about the common characteristic in all kinds of QRS detection algorithms, in the proposed research, the endeavor also embodies in three aspects: extracting feature signal, thresholding, and post processing. In the following section, the details about the algorithm are given.

4.3.2.1 The framework of QRS complex detection algorithm

Fig 4.8 depicts the schematic diagram of the proposed algorithm. In the proposed method, the amplitude of R wave is used only as point of entry. Hilbert transform is applied to the original ECG to form a new signal in which the inverted R waves are identified. A novel threshold method that produces the minimum errors between two different maximal points is then applied to obtain the most likely R waves. Finally, a validation process is used to exclude possible false R waves. In the point insertion process, amplitude and period controllers are applied. The period controller helps in deciding whether to insert a missing R wave between two successive detected R waves; while the amplitude controller checks whether there are possible points of the signal qualifying as the potential R waves satisfying the insertion condition. Finally, the validation process excludes possible false R waves.

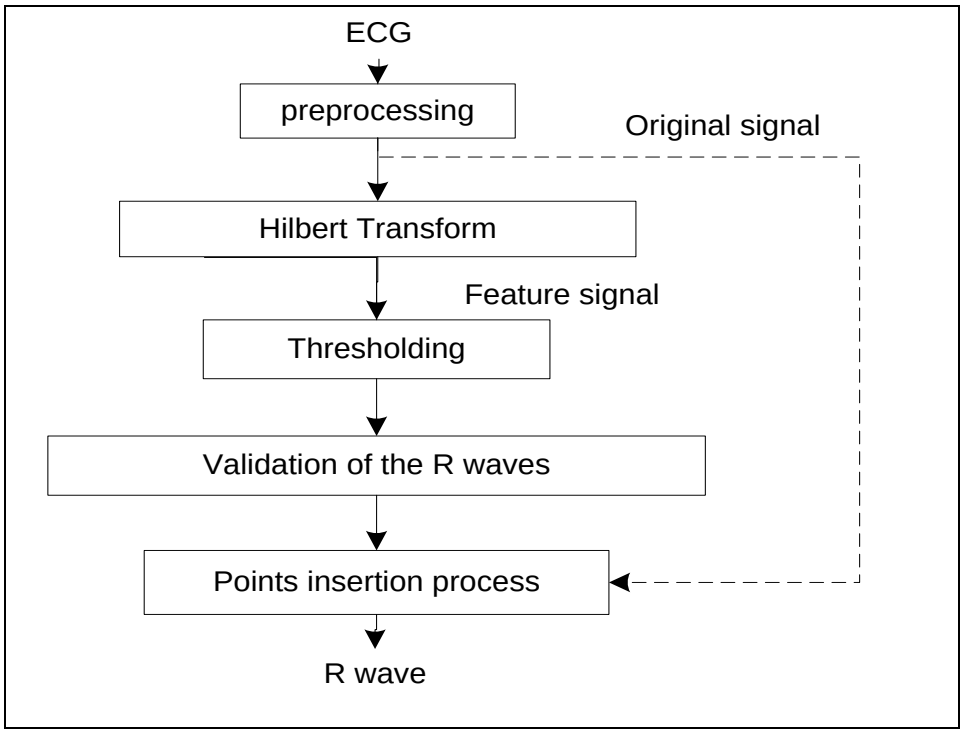


Figure 4. 8: Schematic diagram for R wave detection

In the following sections, preprocessing, Hilbert Transform, thresholding, validation of R waves, and point insertion are described in detail.

4.3.2.2 Preprocessing

Due to interferences from extraneous factor, raw ECG signals are often noisy. Gray describes several main sources of noises and interferences, including power line interference, electrode contact noise, motion artifacts, and baseline drift. In order to make the result as accurate as possible, it is necessary to filter the noise, as it can greatly influences the result of QRS complex detection. The proposed research adopts the conclusion that QRS energy resides approximately in 5-15Hz.

In the following section, it is illustrated the reason to choose the scope of frequencies, 5-15Hz. "the energy" refers to power spectral density (PSD) [44], describing the power of signal. The power spectrum $S(w) = \int_{-\infty}^{\infty} R(\tau)e^{-jw\tau}d\tau$ of a WSS process $x(t)$ is the Fourier transform of its autocorrelation $R(\tau)$,

$$R(\tau) = E\{x(t+\tau)\overset{*}{x}(t)\} \tag{32}$$

The spectral density captures the frequency content of a stochastic process and helps identify periodicities.

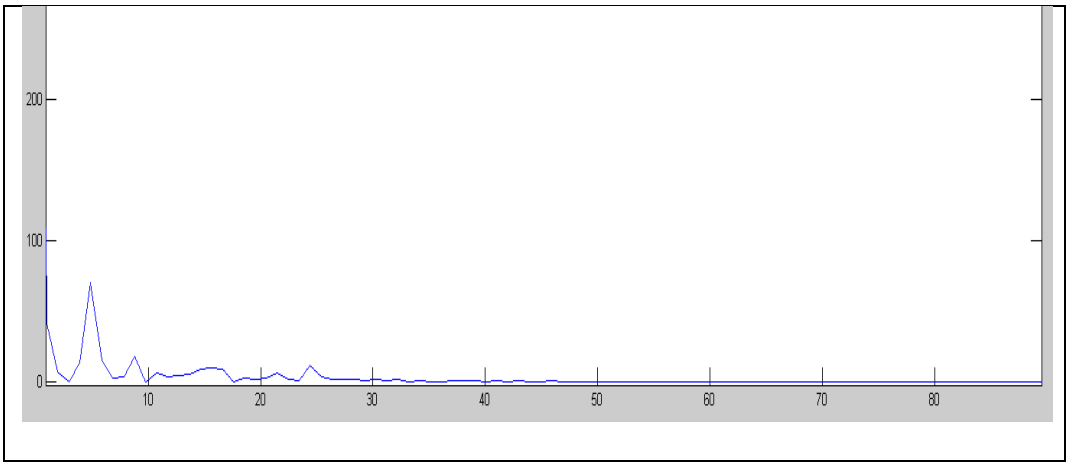


Figure 4. 9: Power spectral density of ECG signals

4.3.2.3 Hilbert Transform

In this study, Hilbert Transform is applied to form the feature signal. Hilbert transform is a linear operator, which takes a signal $x(t)$, and produces another signal $h(t)$ [45].

$$h(t) = H(x(t)) = \frac{1}{\Pi}\int_{-\infty}^{\infty}\frac{x(\alpha)}{t-\partial}d_{\alpha} \tag{33}$$

In the frequency domain, $H(w)$ is an all-pass filter with -90° phase shift.

$$H(w) = -j \operatorname{sgn} w = \begin{cases} -j & w > 0 \\ j & w < 0 \end{cases} \tag{34}$$

The following complex process, $z(t)$, is called analytic signal associated with $x(t)$:

$$z(t) = x(t) + jh(t) \tag{35}$$

The Hilbert transform $h(t)$ of $x(t)$ is used for the computation of the envelope $e(t)$ of the analytic signal $z(t)$.

$$e(t) = \sqrt{x(t)^2 + h(t)^2} \tag{36}$$

In this study, we adopt and customize the approach, introduced in Benitez's research. In this method, using the fact that the Hilbert transform is an odd function, the envelope of first differential waveform of a signal is calculated.

The advantage of adopting Hilbert transform to form the feature signal is that it can detect R wave correctly regardless of whether R wave is inverted or not. The claim is proved next. Showing the ECG signal as $x(t)$, $-x(t)$ is its inverted version, which has the inverted R waves. $f(t)$ is the feature signal and T is the operator forming the feature signal.

$$f_1(t) = T(x(t)) = \sqrt{(\frac{d(x(t))}{dt})^2 + H((\frac{d(x(t))}{dt}))^2} \tag{37}$$

$$f_2(t) = T(-x(t)) = \sqrt{(\frac{d(-x(t))}{dt})^2 + H((\frac{d(-x(t))}{dt}))^2}$$

$$= \sqrt{(\frac{d(x(t))}{dt})^2 + H((\frac{d(-x(t))}{dt}))^2} \quad (38)$$

Now,

$$H(-x(t)) = \frac{1}{\Pi}\int_{-\infty}^{\infty} \frac{-x(\alpha)}{t-\partial} d_\alpha$$

$$= \frac{1}{\Pi}\int_{-\infty}^{\infty} \frac{x(\alpha)}{-t+\partial} d_\alpha$$

$$= -\frac{1}{\Pi}\int_{-\infty}^{\infty} \frac{x(\alpha)}{t-\partial} d_\alpha$$

$$= -H(x(t)) \quad (39)$$

Therefore, for envelope of the signal, we final get the same value from the original signal and its inverted version.

4.3.2.4 Thresholding

After identifying the R wave candidates, one of crucial steps in accurate R wave detection is selection of a threshold. For the thresholding process, there are two main approaches for selecting the threshold: fixed threshold [46,47,48] and adaptive threshold mechanisms. In this study, inspired by a thresholding technique used in an image processing method called OTSU [49], we developed a new thresholding method based on the overall characteristics of the feature signal. OTSU automatically performs histogram shape-based image thresholding and reduction of a gray level image to a binary image. The algorithm assumes that the image to be thresholded contains two classes of pixels (e.g. foreground

and background) and calculates the optimal threshold value separating these two classes so that their combined spread (intra-class variance) is minimal.

Analyzing the feature signal reveals that indeed there exist two local maxima. One represents the baseline, while the other corresponds to the R peaks. Between RR intervals, irregular ECG signal may also have other local maxima because of ripples or other abnormalities. Fig 4.10 is the ECG signal. These local maxima can be taken as R waves by automated method for QRS detection, and as such, must be handled carefully. In order to reduce the errors caused by these irregular local maxima, in our proposed method, the thresholding technique is designed to consider these irregular maxima.

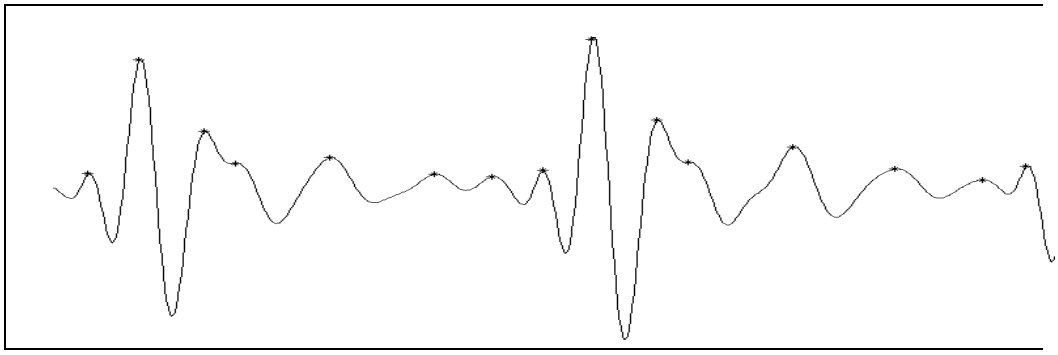


Figure 4. 10: ECG signal (dots stand for local maxima)

The proposed method of threshold selection is as follows. First, the values of all local maxima in feature signal are obtained. Then, maximum (Max) and minimum (Min) of the feature signal or segmentation of the feature signal are extracted. The algorithm compares every value a between Min and Max to minimize the error E. For each a, two sub-features signal $f_1(t)$ and $f_2(t)$ are split from $f(t)$, where $f(t)$ represents the feature signal that includes only the local maxima.

$$f_1(t):\{f(i)\mid f(i)>=a, i\in[1,n]\} \tag{40}$$

$$f_2(t):\{f(i)\mid f(i)<a, i\in[1,n]\} \tag{41}$$

Then, the average values of the sub-features $V_{\max}, V_{\min}$ are calculated by following equations.

$$V_{\max}=mean(f_1(t)) \tag{42}$$

$$V_{\min}=mean(f_2(t)) \tag{43}$$

Finally,

$$E=\min_a\{N_{\max}\times N_{\min}\times(V_{\max}-V_{\min})^2\}\quad a\in[Min,Max] \tag{44}$$

where $N_{\max}, N_{\min}$ are the amounts of local maximum points where values in feature signal are bigger or smaller than a, respectively. Therefore, the final threshold is a', which minimizes the error E. Fig 4.11 illustrates the optimum for thresholding.

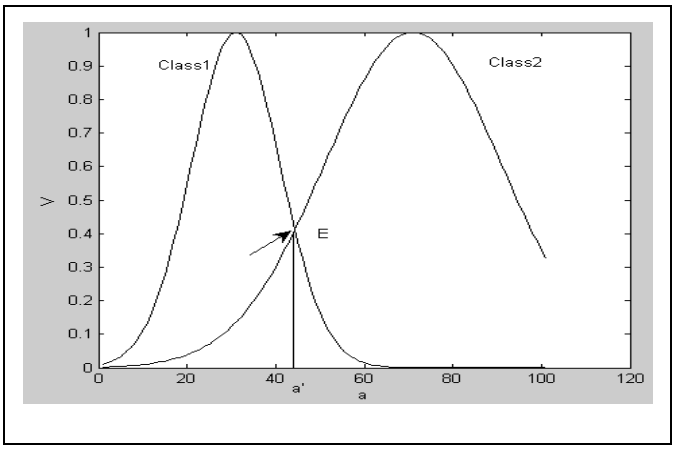


Figure 4. 11: Optimum for thresholding

Next, the proposed method is compared with three thresholding methods, which are considered as the most popular existing techniques:

1. $Threshold = \lambda \max[x(t)]$,
2. $Threshold = mean[f(t)]$ where $f(t)$ is the feature signal,
3. $Estimate(n) = (1-A).Estimate(n-1) + A.peak(n)$ where $Estimate$ is a function estimating threshold at point n, $peak$ finds the amplitude at point n, and A is a constant) ,
4. The method, presented in this study.

In the first method, if some R peaks exceed the amplitudes of other R peaks, many weak R peaks will be missed. The third method is highly susceptible to irregular peaks and changes in the RR interval. This method would fail to find true R peaks in patients with moderate or severe hemorrhage. The second method and the forth method (i.e. the technique presented here), are tested and further compared in this study. Fig 4.12 (a) is the result of the second method. The line in black is the threshold line, which is calculated based on the entire feature signal. Fig 4.12 (b) shows the result of our proposed method applied to the same ECG signal. In Fig 4.12, the dark points above the threshold line are regarded as R wave candidates. As it can be seen, our proposed method can better distinguish R peaks from other local maxima. Such a capability greatly improves the accuracy of QRS detection, and simplifies the following steps to a large extend.

The main advantages of our proposed thresholding can be summarized as follows:

1. Acquiring threshold is based on the overall feature signal or the segmentation of the feature signal. This simplifies the process and allows extraction of R waves in real-time.
2. It minimizes the error between two different types of local maxima, as described above, and therefore can improve the accuracy of R wave extraction.
3. Compared with the adaptive thresholding, it dramatically reduces the errors due to the variations in the RR interval and the amplitude of R waves.

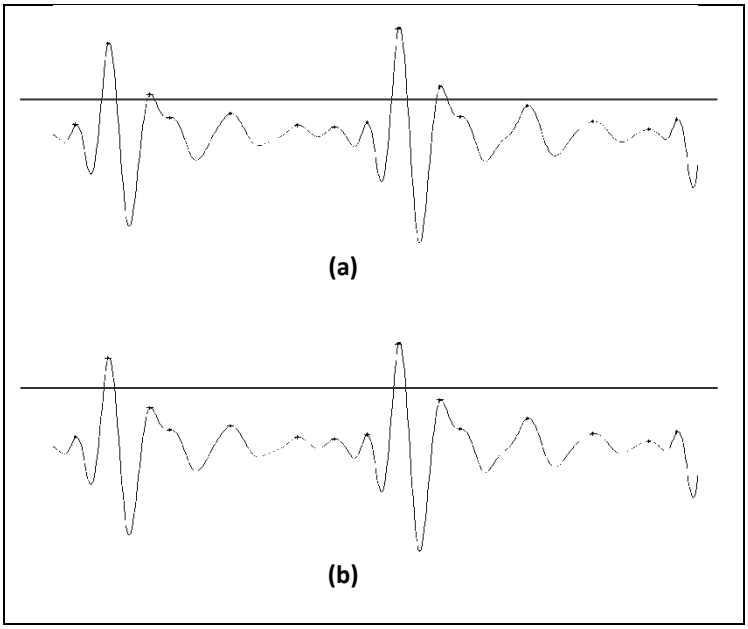


Figure 4. 12: Thresholding

4.3.2.5 Validation of the R waves

In any thresholding process, there is a likelihood of false R detection because of prolonged T waves or ripples between two successive R waves. In order to address this, the common trend is to find the point with the maximum amplitude in a local window as the R wave. However, it is difficult to find a proper window size for this maximization

step. In this study, we propose to exclude the false R waves by checking whether there is an overlap between two successive cycles based on the shape of ECG (Fig 4.13).

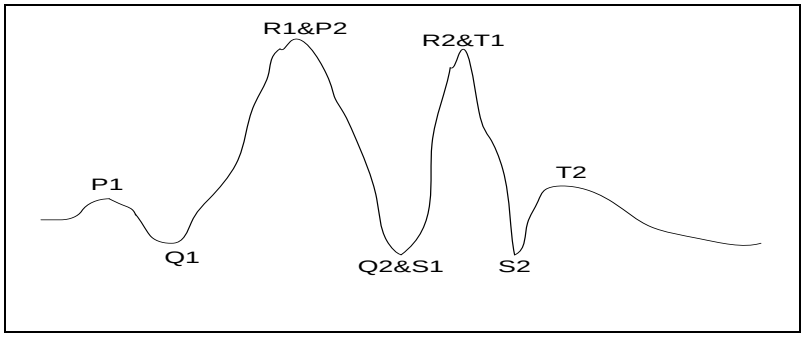


Figure 4. 13: Addressing interference effect due to cycle overlapping

For instance, in the example shown in Fig 4.13, R1, S1, T1 are overlapping with P2, Q2, R2, respectively; i.e. P1, Q1, R1, S1 and T1 belong to one cycle; while P2, Q2, R2, S2 and T2 belong to another. The proposed algorithm picks up the maximum R wave, that is, R1 as final R wave.

4.3.2.6 Point insertion process

Up to this stage, the algorithm may miss some R waves with small amplitudes. The purpose of this step is to insert missing R waves using two controllers: amplitude controller and period controller.

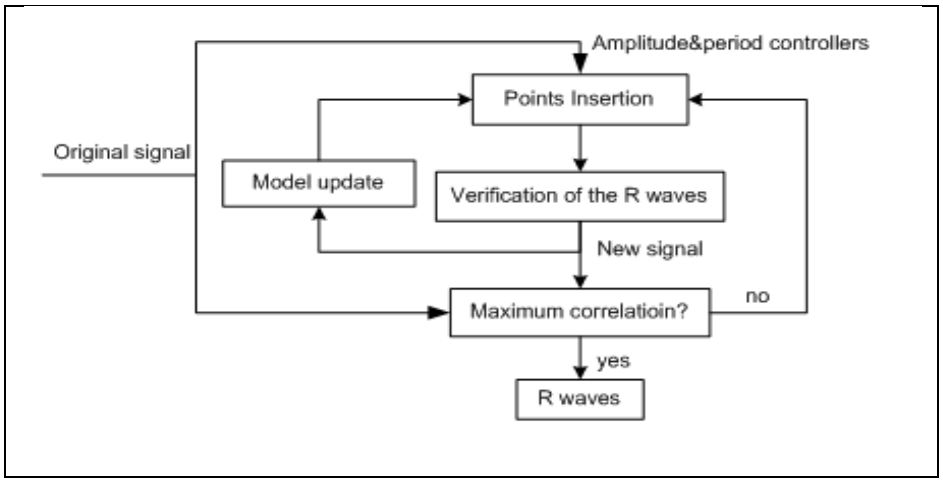


Figure 4. 14: Point Insertion Process

The period controller assists in deciding whether to search for and insert a missing R wave between two successively detected R waves. The amplitude controller checks whether there are peaks in the interval between the two detected R waves that could qualify as the possible R waves satisfying the insertion conditions. The algorithm automatically inserts an R wave if both conditions are satisfied. In this process, for different combinations of amplitude and period controllers, the correlation of the original signal and the new signal with the R wave insertion is calculated. When the value of the correlation reaches to the maximum, the algorithm achieves its final detection result. Fig 4.14 is the schematic diagram for this point insertion process. Next, the amplitude controller, period controller, and model update process are described in more detail.

1. Period controller

This parameter, adjusted by parameter α, controls the acceptable or anticipated time interval between two consecutive R waves. The scope of period controller parameter is between 0.2 and 0.4. The period controller also helps to calculate threshold,

c, which decides whether to insert a missing R wave between two consecutively detected R waves, i.e.

$$c = \alpha \times m_1^{\,2} \Big/ \sqrt{m_1^{\,2} + s_1^{\,2}} \tag{45}$$

where m_1 is mean of correctly detected RR interval in a given window, and s_1 is the standard deviation of correctly detected RR interval in the same window. When the gap between two successive R waves exceeds c, the algorithm decides that a R wave has to be inserted between the two existing R's.

2. Amplitude controller

This controller, adjusted by parameter β, decides how large a peak in the signal between two successively detected R waves has to be to qualify as a missing R wave. The scope of amplitude controller parameter is between 1.8 and 3.0. The amplitude controller helps to calculate threshold e, which marks a point that may be P, Q, R, S or T. The parameter e is defined as:

$$e = \beta \times m_2 \tag{46}$$

where m_2 is the mean of positive points in ECG signals in a given window. When the amplitude exceeds e, the algorithm marks the timing of the point and its amplitude.

3. Insertion rule

We adopt four points (P_a, P_b, P_c, P_d) rule to decide whether there is a possible R wave. The rule requires that the amplitude of P_b exceed the amplitude of P_a, P_c, and P_d.

The parameter P_d is also a threshold value for excluding noises (gap is used referred to P_d). Fig 4.15 shows the relation of the four points.

When it satisfies the four points rule, it is a possible location for R wave insertion. To further verify the need to insert an R wave, the steps described next, are performed.

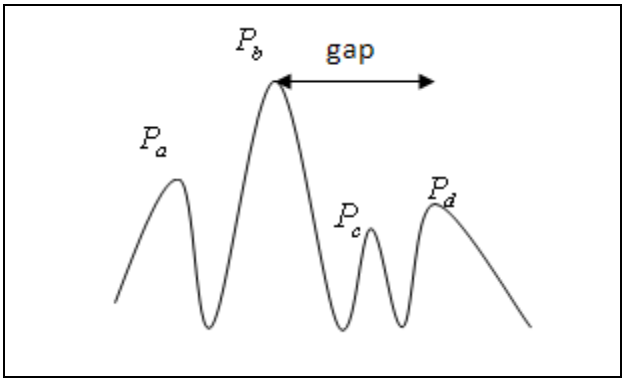


Figure 4. 15: The relation of the four points

4. Model update

First a model of ECG is formed. The initial model of ECG signal is formed randomly based on one cycle of ECG signal. As R wave detection proceeds, an updated model is created by calculating the average value of several most recent cycles of ECG.

$$Model = \sum_{1}^{n} f_i / n \tag{47}$$

where f_i is the signal for the ith detected cycle and n is the number of cycles included in the averaging process. These cycles are chosen to be the most recent ones.

5) Correlation

The calculation of correlation reflects the degree of matching between the original feature signal and new produced signal. First, the formation of new signal is described in detail.

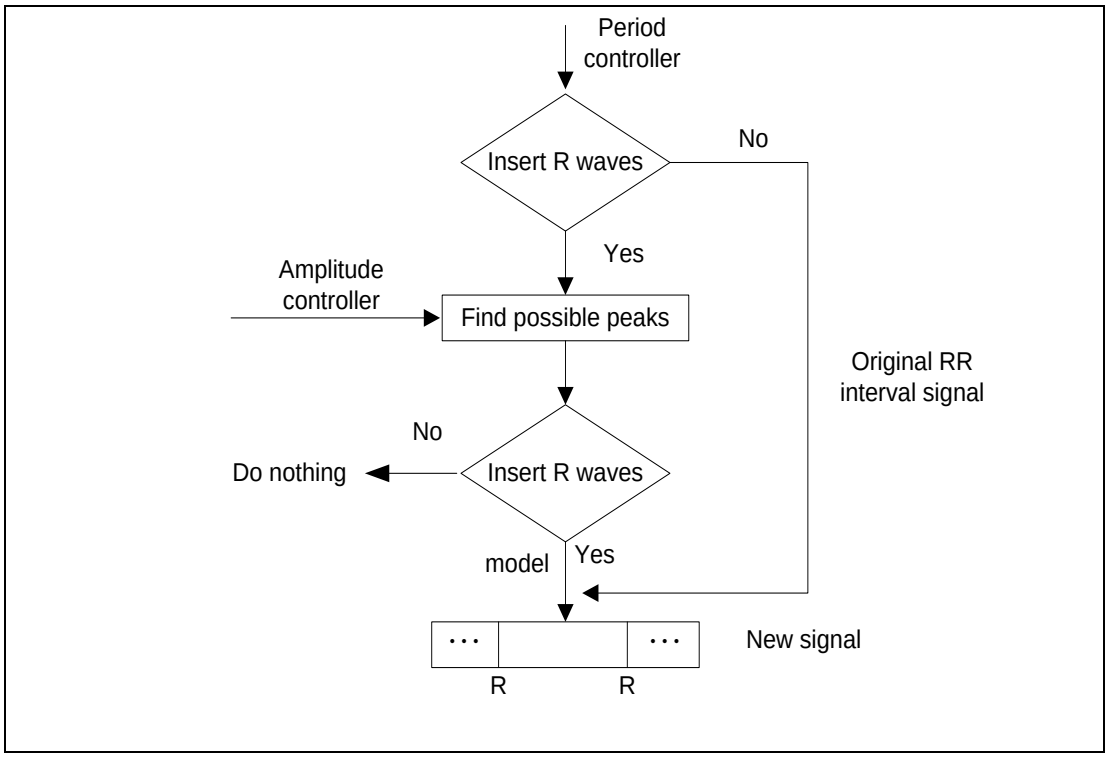


Figure 4. 16: Diagram of the formation of new signal

For each detected RR interval, after checking period controller, the algorithm decides whether to insert R waves. If it does not insert any R waves, it will keep the original RR interval signal. If it does, however, the algorithm uses the amplitude controller to decide possible peaks, and find possible R peaks using four points rule. Then, the algorithm inserts the template (model) in the position of possible R peak. Fig 4.16 shows the diagram of the formation of new signal.

4.3.2.7 Improvement of QRS Detection Method

The proposed QRS complex detection algorithm, discussed above, can benefit from some further improvements such as the followings.

1. The general spectrum of the frequency of the signal analyzed in the proposed method is from 5 to 15Hz. However, for analysis of hemorrhage, restricting the frequency range to a smaller spectrum, as suggested in [5, 15], may improve the performance of methods such as ours. For different applications, this range may be needed to be adjusted. In general, in an application it is known beforehand that if there is too much noise in a certain range of frequency, then the frequency range of the method must be adjusted accordingly.
2. The final position of R wave needs to adjust, because the maxima correspond to the intersection between the positive and negative slopes of original ECG signal.
3. The parameters in point insertion process need to be adjusted for different ECG signals. Since a better algorithm is expected to have fewer parameter adjustments, a more systematic approach for the parameter adjustments can improve the algorithm.

4.3.2.8 The detection of other waves

Except for R wave, ECG signal includes other waves, which also are very important information suppliers. The following section illustrates the detection of P wave, Q wave, S wave and T wave. Often, the key factor in detection of other waves is that the other waves occur at relatively regular intervals with respect to the position of R wave. However, irregular ECG signals due to illnesses , such as the abnormal patterns shown in

Fig 4.17, can violate this perception. Therefore, according to the type of abnormal ECG, neccessary postprocessing is added to deal with these special cases.

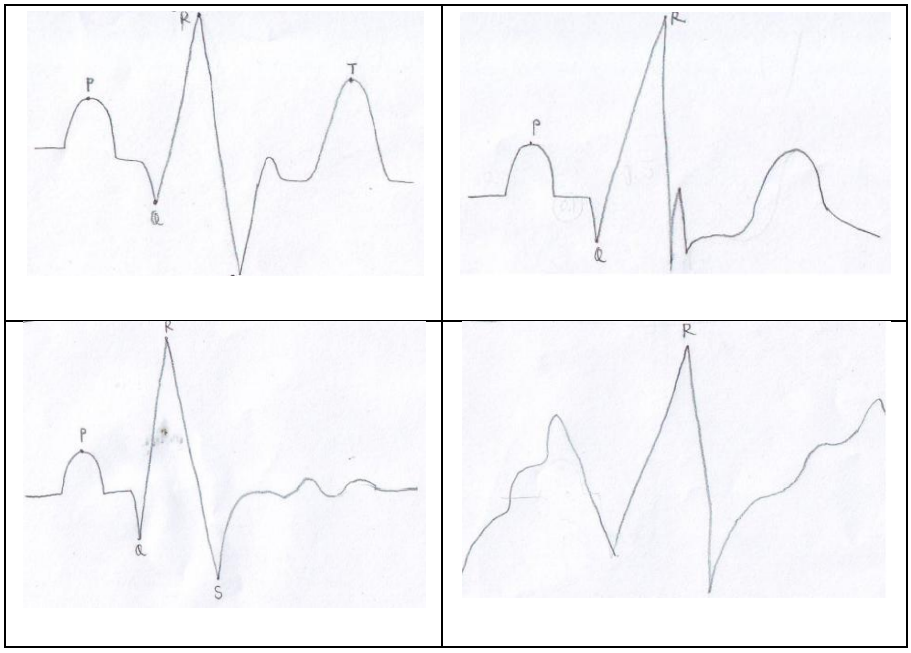


Figure 4. 17: Abnormal shapes of ECG signal

4.3.3 Systole&Diastole detection

The characteristic of Arterial Blood Pressure was shown above. In the first glance, it seems that simple thresholing might detect both systolic and diastolic peaks. However, in order to avoid the complications involved in setting and changing thresholds (due to the irregular period and amplitude of ABP signal), in the proposed method, the microscopic angle approach is taken. The design idea is to decompose the feature signal into several components, which make up all the possible situations. Then, detecting Systole&Diastole becomes a problem, which finds the target by considering the relations in the

components. In the following section, preprocessing, the formation of two signals, components of the model and operations for combination step are described in detail.

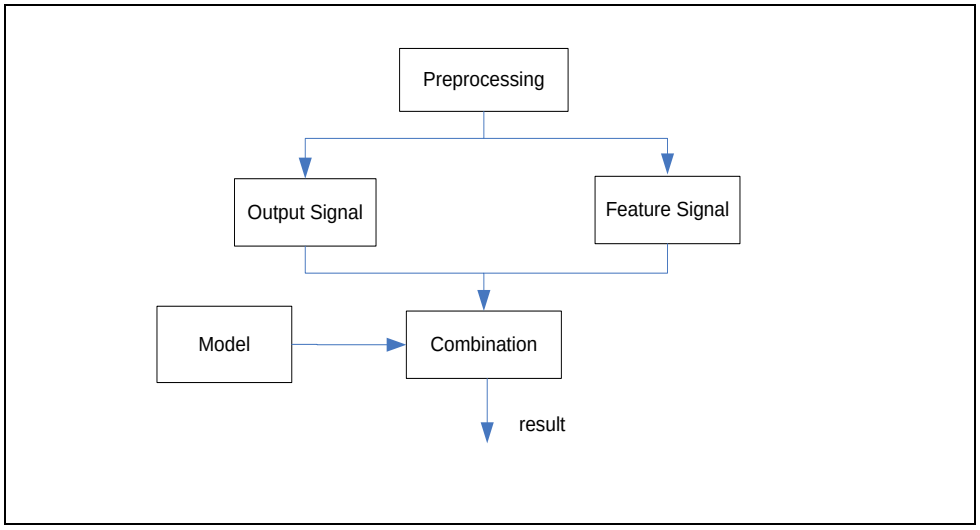


Figure 4. 18: Schematic diagram for Systole&Diastole detection

4.3.3.1 Preprocessing

In the proposed research, because of the typical lack of noise, the objective of preprocessing in Systole&Diastole detection in ABP signal is to obtain new signal that is more suitable for further processing. The operation used in this step is smoothing that reduces the ripples in the ABP signal. There are two types of new signal produced. One is the signal after smoothing, called the “source signal”. The other one is the signal for labeling position in order to facilitate the localization of the important turning points, called the “sign signal”.

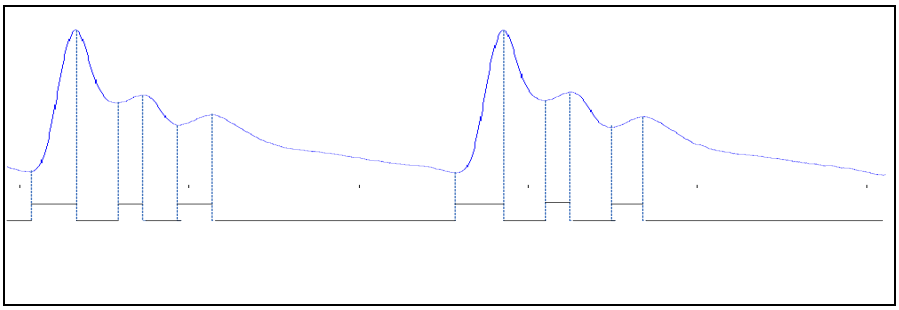


Figure 4. 19: The relation between source signal and sign signal

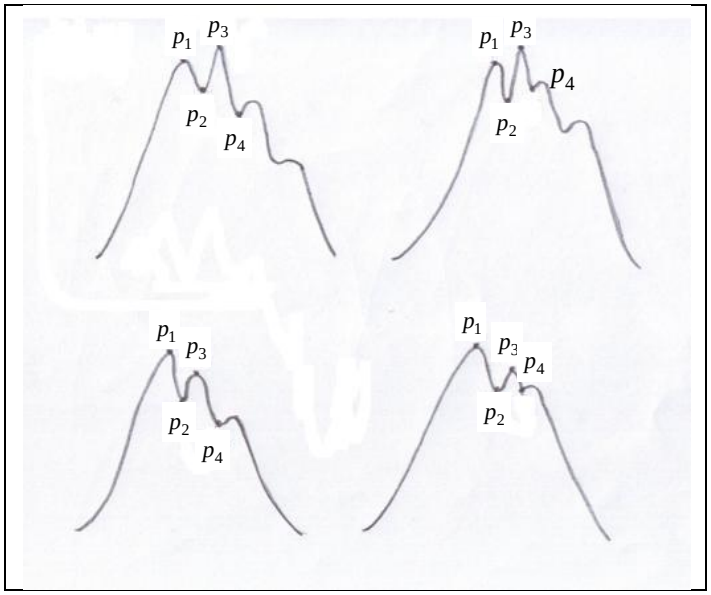
In the sign signal, for the points that correspond to the positive slope in the source signal, the value of the sign signal is “+1”; while the points corresponding to the negative slope in the source signal, get the value of “-1” in the sign signal. The relationship between the two signals is shown in Fig 4.19.

4.3.3.2 Model

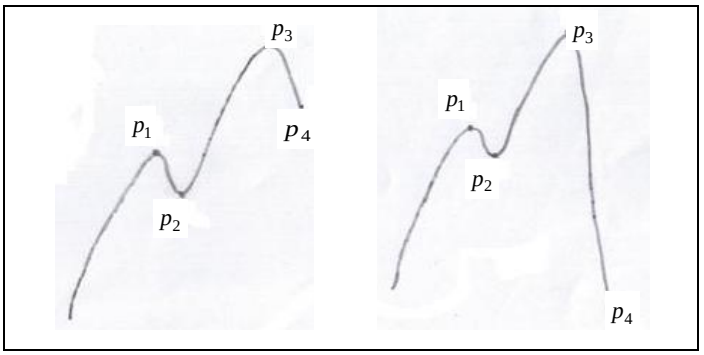
After smoothing, the source signal is decomposed into several components as shown in Fig 4.20. The model, designed to best capture the characteristics of the ABP signal, is built on four points, which are four successive intersections between negative and positive slopes. In the model, there is no limitation on the relation between amplitude of p_2 nad p_4; the parameter 'thresholdH' is used to decide the boundary of the components; the four points locate in one cycle, which is controled by another parameter 'thresholdP'. These components are defined and shown as follows:

1. Component#1: ‘|amplitude (p_3)-amplitude (p_1)|<thresholdH’.
2. Component#2: ‘|amplitude (p_3)-amplitude (p_1)|>thresholdH’ && ‘location (p_3)-location ()<thresholdP/4’.

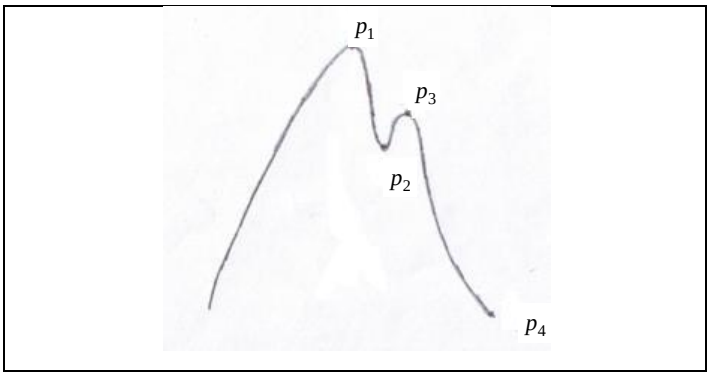
3. Component#3: '|amplitude (p_3)-amplitude (p_1)|>thresholdH' && 'location (p_3)-location (p_1)>thresholdP/4'.



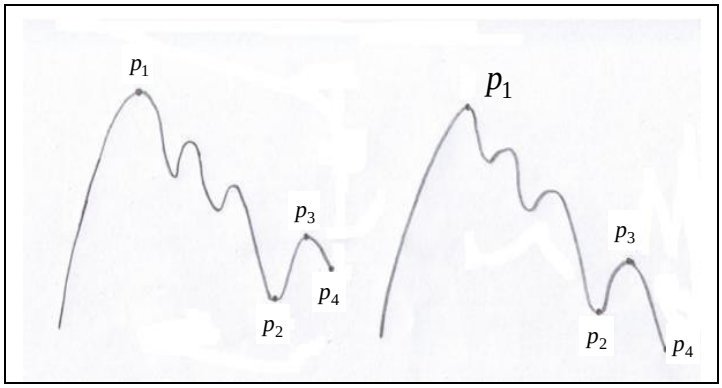
p_4 Component #1



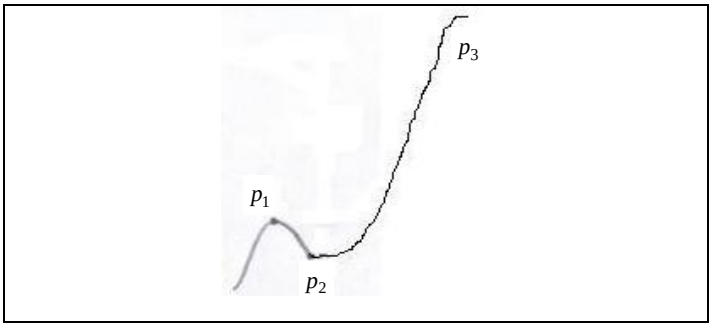
The first catogory in Component #2



The second catogory in Component #2



The first catogory in Component #3



The second catogory in Component #3

Figure 4. 20: The model of ABP signal

4.3.3.3 Combination

The objective of combination is to know how to use these components in the model. First, using the sign signal to find the first four points. Then, judge the relation among the four points by referring to the components in the model. The operations for detecting and shifting points to new places fall into three stages (Fig 4.21):

1. Stage1: the first component plays role. If p_1 and p_3 are in one cycle, p_1 is one systolic peak and shift p_2, p_3 and p_4. If not, p_1 is one systolic peak and shift p_1, p_2, p_3 and p_4.
2. Stage2: the second component plays role. If it belongs to the first sub-catogory, shift p_2, p_3 and p_4. If it belongs to the second sub-catogory, p_3 is one systolic peak and shift p_1, p_2, p_3 and p_4.
3. Stage3: the third component plays role. If p_1 and p_3 are not in one cycle, p_1 is one systolic peak and shift p_1, p_2, p_3 and p_4. If it belongs to the first sub-catogory, p_1 is one systolic peak and shift p_1, p_2, p_3 and p_4. If it belongs to the second sub-catogory, p_3 is one systolic peak and shift p_1, p_2, p_3 and p_4.

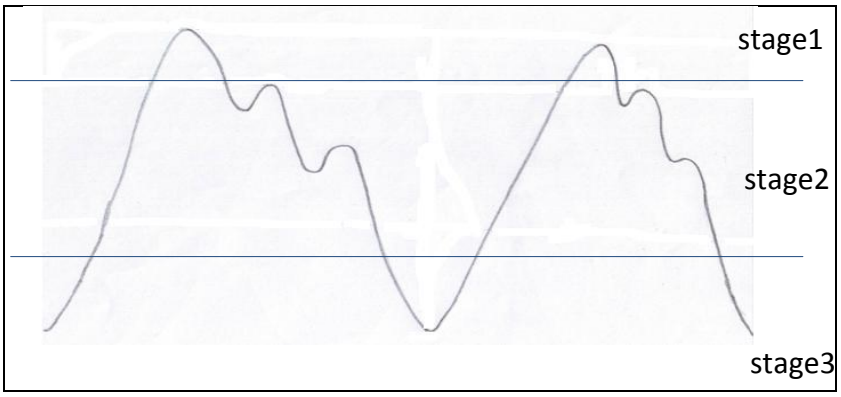


Figure 4. 21: The stage divide in ABP signal

4.3.3.4 Initial Assessment of Systole&Diastole Detection Method

For the proposed Systole&Diastole detection algorithm, the following observations and considerations that can be made.

1. Observing the spectrum of the ABP signal, it becomes evident that there may be no fixed limits on the frequency range for Systole&Diastole events in the ABP signal. Therefore, any frequency adjustment of the Systole&Diastole detection may not be very helpful.
2. Compared with the method using a global or local threshold, the advantage is that the proposed method does not rely on such parameters. This is due to the fact that we adopt a microscopic parameters approach in which the settings for the parameters are flexible.
3. The proposed Systole&Diastole detection algorithm still needs to be improved. Again, in this approach, the use of fewer microscopic parameters and the logicality of the components in the model are preferred.

4.4 Summary

In this chapter, a brief review of the basic methods as well as the description of the proposed techniques for preprocessing is provided. These tasks involve denoising, QRS complex detection, and Systole & Diastole detection. For each task, the related research, the research direction/methodology, and the analysis of algorithm were presented.

{CHAPTER 5 Feature Extraction}

5.1 Overview

After preprocessing, feature extraction will operate on the processed signals to collect different types of informative features. In our particular application, these features are expected to better represent the patterns of hemorrhage / blood volume loss. The 'Existing Trauma Triage Systems' in Chapter 2 reviewed some trauma triage systems, in which different kinds of features are adopted to represent hemorrhage. Our proposed system will attempt to explore possibly superior features. The defined features will be in time domain, frequency domain, created using nonlinear analysis, and formed based on multi-model analysis. In the following part, the features will be listed, and necessary explanations will be added to some specific features.

5.2 Proposed Method

5.2.1 Features from ECG

During hemorrhage, the time variations, although not always visible to the human eyes, occur in the shape and structure of ECG signal. Therefore, the proposed work attempt to extract features based on these time patterns and indicators. Also, the features in frequency domain are explored to further increase and enhance the accuracy of the reasearch. Table 5.1 lists all the explored features in time domain, frequency domain, and nonlinear analysis from ECG signals.

Table 5. 1: Features in ECG

Feature Type	Name of the features
wave interval	PP interval; QQ interval; RR interval; SS interval; TT interval
relative wave interval	relative PQ interval; relative PR interval; relative PS interval; relative PT interval; relative QS interval; relative QT interval; relative RT interval; relative ST interval;
wave amplitude	amplitude of P wave, S wave and T wave
relative wave amplitude	relative PQ amplitude; relative PS amplitude; relative PT amplitude; relative QT amplitude
frequency in one window	frequency with maximum amplitude
frequency in one cycle	maximim amplitude in frequency spectrum; frequency with maximum amplitude
signal averaging	PQ interval averaging; QS interval averaging; ST interval averaging; PQRST interval averaging
nonlinear analysis	improved Lempel-ziv

5.2.1.1 Details for some specific features

I. wave interval.

The wavel interval feature represents the same information as the heart rate, because the calculation of heart rate obeys the following equation (fs is sample rate; RR is the average sample points for one RR interval):

$$Heartrate = \frac{60 \times fs}{RR\ interval} \tag{48}$$

II. relative wave interval.

This feature reflects the relative time intervals in one cycle and is calculated as:

$$PQ\ relative\ wave\ \text{int}\ erval = \frac{PQ\ \text{int}\ erval}{RR\ \text{int}\ erval} \tag{49}$$

III. relative wave amplitude

The calculation of relative wave amplitude is based on the following equation:

$$relativePQ\ amplitude = P\ amplitude - Q\ amplitud \tag{50}$$

IV. frequency in one window and in one cycle

The frequency spectrum is calculated based on signal in one window and in one cycle. The features are maximim amplitude in frequency spectrum and frequency with maximum amplitude.

V. signal averaging

For signal in every window, there are several subsignals (such are RR interval). The features are obtained based on the sum of subsignals. Although the length of each subsignal in every window is different, the interpolation is employed to resize all subsignals into fixed lengths. Then, similar frequency information are extracted from the summation of the subsignals.

VI. improved Lempel-Ziv

As mentioned above, the conventional Lempel-Ziv can not extract enough information from biomedical signals to better represent hemorrhage. An improved version of Lempel-ziv has been tried to capture more information. The methodology and the effectiveness of improved Lempel-Ziv will be illustrated in the following sections.

5.2.2 Features from ABP

ABP is another important biomedical signal whose variations are also known to be closely associated with blood volume loss. The intention of this study is to find features that better capture these variations/patterns/featrues.

Table 5. 2: Features in ABP

Feature type	Name of the features
features based on cycle	number of bumps; duration of each bump; energy; the frequency in one cycle; phase; concave&convex of the signal; relative value toward the left and the up;
features based on one window	frequency information; time duration between Systole and Diastole; amplitude difference between Systole and Diastole; mean arterial pressure; number of Systoles in one window;
others	variation of the wander; slope of lines; proposed novel

	/ improved Lempel-ziv; respiration rate extracted from ABP;

Table 5.2 lists the explored features based on ABP signal. This list is expected to grow and the ones mentioned in Table 5.2 will be further enhanced and improved in the future.

5.2.2.1 Details for some specific features

In the proposed method, the informative featrues are extracted from individual cycles of a subsignal. The following figures provide better a demonstratation of the extracted features.

I. number of bumps

For ABP signal, in each cycle, there are several "bumps" as shown in Fig 5.1. Depending on the complications and/or illness, the ABP shows differences in the number of these bumps in each cycle. In the proposed method, the number of these bumps is captured as one feature to represent the possible existence and/or severity of blood loss.

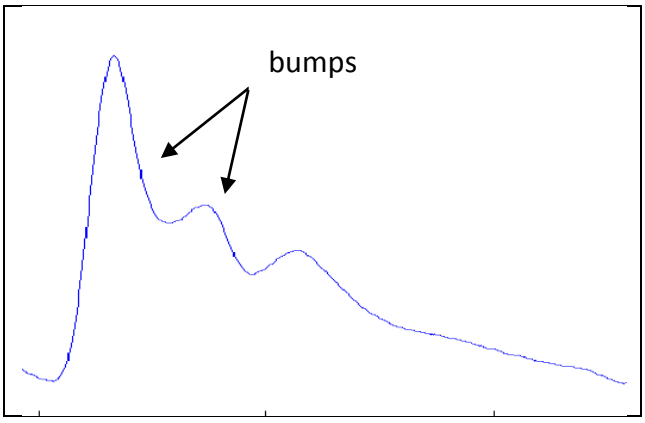


Figure 5. 1. Bumps in one cycle

II. duration of each bump

The duration of each bump, e.g. the time interval between A point and B point as shown in Fig 5.2, forms another family of features.

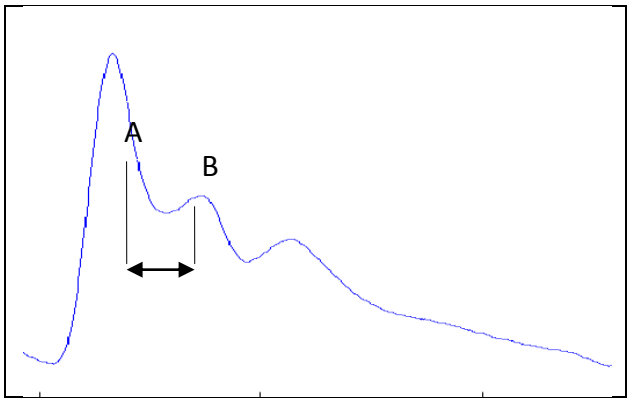


Figure 5. 2. The duration of each bump

III. concave&convex of the signal

The purpose of the feature is to show the area difference under the red line and the blue line. The difference between these two cureves show how convex (or concave) the ABP signal is.

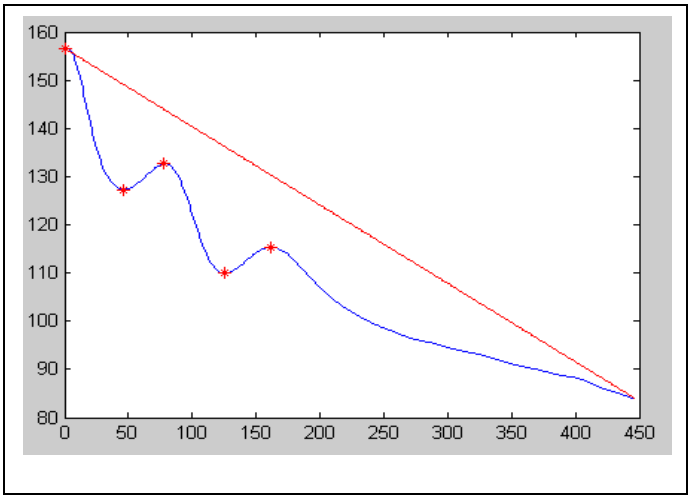


Figure 5. 3. The concave&convex of the signal

IV. relative value toward the left and the up

The relative value toward the left is obtained by dividing w using W, and the relative value toward the up is gotten by dividing h using H, where W is the width of the entire cycle measure from the first peak, w is the time interval between the frist and the second peaks, H is the height of the first peak, and finally h is the height of the second peak.

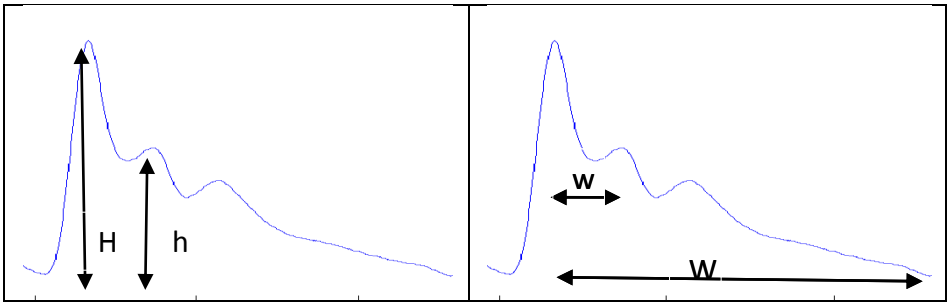


Figure 5. 4. The relative value toward the left and the up

V. mean arterial pressure

The calculation of the mean arterial pressure obeys the following equation:

$$\text{MeanArterialPressure} = \text{Diastole} + 1/3 \times (\text{Systole} - \text{Diastole}) \quad (51)$$

VI. variation of the wander

The variation of the wander reflects the level of deviations from the amplitude of Systolic peaks and is evaluated as

$$\textit{The variationof the wander} = \textit{Standard Deviationof}\ (\ \textit{systole} - \textit{min}\ (\ \textit{Systole}\ )\ ) \quad (52)$$

where systole is the amplitude in one systolic peak and Systole is a vector containning all amplitudes of systolic peaks.

VII. slope of the lines

For ABP, the Systolic peaks present rise and fall as the proceeding of hemorrhage. In the research, the lines built on Systolic peaks are interpolated, and the weighted slopes are used as features.

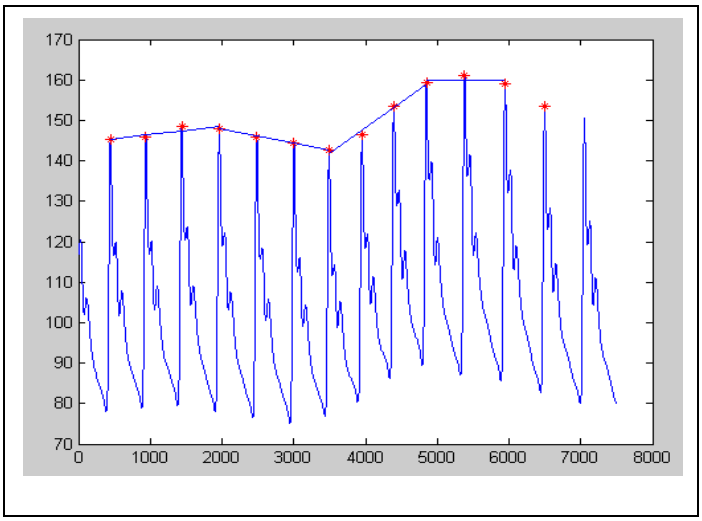


Figure 5. 5. The slope of the lines

The formulation of weighted slopes is based on the following equation:

$$\text{Weight slope} = \text{slope} * \text{Num} \tag{53}$$

where "slope" is the slope of the interpolated line and Num is the number of points in the interpolated line.

VIII. improved Lempel-Ziv

For dealing with ABP signal, the improved Lempel-ziv is also adopted to capture more information to better represent blood volume loss. The methodology and the effectiveness of the improved Lempel-Ziv will be explained later.

IX. respiration rate

In this study, respiration rate extracted from ABP is employed as another feature. The curve of Systolic peaks is formed using interpolation, and the maximum frequency is acquired by using Fourier Transform. Our results showed that the respiration rate extracted from ABP using our method can achieve more than eighty percentage of accuracy in predicting the real (actual) respiration rate +/- one breath/minute. As such, the proposed method allows avoiding the inconvenience of using conventional respiration rate measurements systems, and at the same time, provides reasonable accuracy in predicting the real respiratory rate. In the following part, the simulated results are given. As a standard to verify the accuracy for estimation, PETCO2 from LBNP is used. The real respiration rate is acquired by locating the maximum frequency of PETCO2 signal beween 0.1 Hz and 0.5Hz. Table 5.3 is the estimation result. The first column shows the window size and the frequency scope to locate the maximum frequency. The second and the third column present the accuracy difference by predicting the real respiration rate with or without +/- one breath/minute.

Table 5. 3 Estimation result

Seconds/frequency scope	Without "+/- one breath/minute"	With "+/- one breath/minute"
15 sec / [0.1, 0.5]	58.24%	80.53%
30 sec / [0.15, 0.5]	59.91%	79.03%
60 sec / [0.15, 0.5]	58.7%	72.50%

5.2.3 Details for features from nonlinear analysis

Lempel-Ziv [50] is a measure that analyzes symbolic sequence complexity. Lempel-ziv algorithm is an explicitly computable measure for finite sequences measuring the complexity of a sequence by quantifying the minimal number of steps required for its synthesis throughout a specific process. At each step of the synthesis process, two operations are allowed: generation of a new symbol, and copying a fragment from the part of the sequence that has already been synthesized.

As a powerful tool to estimate complexity, Lempel-Ziv has recently become popular in many fields [51,52,53,54] and is used to measure the degree of randomness in complex sequences. The Lempel-Ziv's applications in processing of physiological signals have been numerous (see for instance [55]) in which this measure has been used to analyze the complexity of biomedical signals in different conditions.

Hemorrhage causes changes in the shape of biomedical signals such as ECG and ABP. Capturing potential hemorrhage-caused changes in patterns of randomness in these signals may not only require the use of general Lempel-Ziv but also improved versions of Lempel-ziv. Therefore, the proposed work will include expansions and extensions of the general idea of Lempel-Ziv. Specifically, the project will target enhancing the Lempel-Ziv's performance in detecting the changes in the instinct randomness when hemorrhage happens.

5.2.3.1 General Lempel-Ziv

First, some basic concepts necessary to describe Lempel-ziv measure are introduced and explained.

Assume that A^* is the set of all finite sequences over a finite alphabet A; and S is a sequence and $l(S)$ is the length of S. Next define A^n as the formulation of the above mentioned object:

$$A^n = \{S \in A^* \mid l(S) = n\}, n \geq 0 \tag{54}$$

Further define Λ as the null sequence (the length of the sequence is zero). Note that Λ is an element of A^*. In this formulation, a substring of S will be expressed as $S(i, j)$, which starts from position i and ends at position j. If $i > j$, $S(i, j)$ will be Λ.

Possible relationships may exist between two sequences $Q \in A^m$ and $R \in A^n$. For instance, if a sequence is defined as: $S = q_1 \cdots q_m r_1 \cdots r_n$; Q is called a prefix of S, and S is called an extension of Q. If the length of Q is less than the length of S, Q is called a proper prefix of S.

The vocabulary of a sequence S, $v(S)$, is defined as the subset of all substrings or words $S(i, j)$ of S. Also, a word is called an eigenword if it does not belong to the vocabulary of any proper prefix of S. The eigenvocabulary of S is defined as composed of all eigenwords of S. There is an operator Π, which can identify all prefixes of S, i.e. $S\Pi^i = S(1, l(S) - i)$. It means that the prefix of S is the sub-string from the beginning of the string S to the position, which is i points far from the end of the string S.

The concept of Lempel-ziv measure treats the measure of complexity as a self delimiting learning machine, i.e. it scans a sequence $S = s_1 \cdots s_n$ from left to right, and if a new word (eigenword) is met, the measure of Lempel-Ziv complexity will increase. Therefore, the measure of Lempel-ziv is built based on the structure of the sequence S .

The building of the sequence S can be viewed as extending a sequence S by adding one of its words Q to generate an extension $R = SQ$. The only provision is that Q is an element of $R = v(R\Pi)$.

Next, let's distinguish two items: producibility and reproducibility. An extension $R = SQ$ of S is reproducible from S (S -> R) if $Q \in v(R\Pi)$. A nonnull sequence S is producible from its prefix, i.e. $S(1, j)$ ($S(1, j) \Rightarrow S$), if $S(1, j) \to S\Pi$ and $j < l(S)$. From the definition of these two items, the distinction is clarified, i.e. the last symbol of $S(1, j)$ may be different from the last symbol of S if it is producibility.

Finally, let's focus on the production of complexity of a sequence. The nonnull sequence S is regarded as a production ($Q \Rightarrow S$) from some proper prefix of S . A production process of S is a mechanism to generate S step by step. The result $S(1, h_i)$ of step i is called the ith state of the process. Assuming that the entire process has m steps, the parsing of S will become:

$$H(S) = H_1(S)H_2(S)\cdots H_m(S) = S(1, h_1)S(h_1 + 1, h_2)\cdots S(h_{m-1}, h_m) \qquad (55)$$

where $H_i(S)$ is called the components of S , and $h_0 = 0$, $h_m = l(S)$. A component $H_i(S)$ and the corresponding production step $S(1, h_{i-1}) \Rightarrow S(1, h_i)$ are called exhaustive if

$S(1,h_{i-1}) \not\rightarrow S(1,h_i)$. The proposed Lempel-ziv complex $c(S)$ is the number of components in $H(S)$, whose components are all exhaustive.

The following is the description of Lempel-ziv complexity.

1. Let S and Q denote two subsequences of P and SQ be the concatenation of S and Q, while sequence $SQ\Pi$ is derived from SQ after its last character is deleted. $v(SQ)$ denotes the vocabulary of all different subsequences of $SQ\Pi$. Initially, $c(n)=1$, $S=s(1)$, $Q=s(2)$, and therefore, $SQ\Pi=s(1)$.

2. In general, $S=s(1),s(2),\cdots,s(r), Q=s(r+1)$, then $SQ\Pi=s(1),s(2),\cdots,s(r)$; if Q belongs to $v(SQ\Pi)$, then Q is a subsequence of $SQ\Pi$, and not a new sequence.

3. Renew Q to be $s(r+1),s(r+2)$, and judge if Q belongs to $v(SQ\Pi)$ or not.

4. Repeat the previous steps until Q does not belong to $v(SQ\Pi)$. Now, $Q=s(r+1),s(r+2),\cdots,s(r+i)$ is not a subsequence of $SQ\Pi=s(1),s(2),\cdots,s(r+i-1)$, so increase $c(n)$ by one.

5. Thereafter, S is renewed to be $S=s(1),s(2),\cdots,s(r+i)$, and $Q=s(r+i+1)$. Repeat until Q is the last character.

5.2.3.2 Improved Lempel-Ziv

As it can be seen from the above description, the measure of Lempel-ziv is the number of eigenwords in building the sequence. As shown in Chapter 7, the results of using this measure on biomedical signals do not exhibit a clear pattern for hemorrhage. In order to better represent the patterns of hemorrhage, an improved version of Lempel-Ziv will be

need. By looking through the related research, the possible improvements can come from the improvements made to: 1) the thresholds used in forming the symbol signals [56]; 2) the number of symbols used in the symbol signals .

The conventional Lempel-Ziv loses many information due to the transformation to the symbolic signal. At the same time, if the conventional Lempel-Ziv is operated on raw signal, it will be extremely time consumption, which is not desirable for a real time system. Considering the two factors in this study, extending the element in the conventional Lempel-Ziv becomes the entry point for the improved Lempel-Ziv.

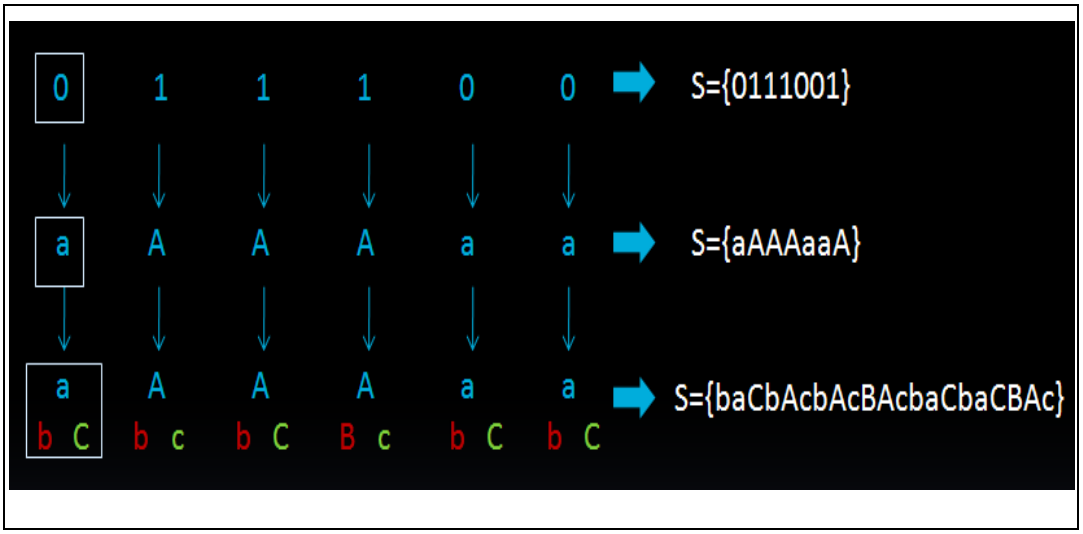


Figure 5. 6. The improved Lempel-ziv

As it can be seen from Fig 5.6, the first line is the sysmbolic signal, which is transfromed from the raw signal using a particular thresholding method. In the improved Lempel-Ziv, the '0'/'1' sysmbol is represented by another symbol. For example, '0' can be replaced with 'a', and '1' can be replaced with 'A'. As a matter of fact, the symbolic signal in the first line and the symbolic signal in the second line are the same. The only difference is

that they use different symbols to represent the symbolic signal. The novelty in the improved Lempel-Ziv lies in extending the basic element. Seeing from the third line in Fig 5.6, there are two different symbolic signals represented by two different colors. These two symbolic signals are acquired using the same method as the conventional Lempel-Ziv, but they represent different information in biomedical signals compared with the symbolic signal in the first line. One is the red one, which is represented by 'b' and 'B'; while the other one is the green one, which is represented by 'c' and 'C'. Therefore, the original element 'a' is extended to 'baC', and the symbolic signal '0111001' is extended to the symbolic signal 'baCbAcbAcBAcbaCbaCBAc'. Then the same implementation mechamism in the conventional Lempel-Ziv is applied to obtain the complexity measure based on the new symbolic signal.

By analyzing the improved Lempel-Ziv, if one more complex symbolic signal is added to the original symbolic signal, in an ideal context, the complexity will increase approximately to the multiplication of the two complexities. Ideally, the improved Lempel-Ziv can achieve to the maximum ratio (1: the number of extended element in the symbolic signal) between two stages in symbolic signal. Certainly, the imporved measure needs to consider a lot of factors to achieve the aim, such as the added symbolic signals, etc. Through the experiments, it is concluded that the improved Lempel-Ziv can not only capture more information from the biomedical signal but also satisfy the real time requirements. The results in Chapter 7 verify the effectiveness of the improved Lempel-Ziv. Howerver, in order to make Lempel-Ziv more effective, the followings can further

explore the results and the future directions of our analysis of biomedical signals using Lempel-Ziv features:

1) The concept of the Lempel-Ziv measure is based on the number of eigenwords (created based on building the sequence by coping its prefix to the sequence) in the symbol sequence. One possible further expansion of this measure can explore quantifying complexity based on factors other than eigenwords.

2) The relationships between successive eigenwords can also be used as a source of information to extract the degree of randomness.

3) In our proposed method, 'Window' is used to reduce the time consumption. However, such a division and its size affect the value of the Lempel-Ziv to some degree, and therefore needs further attention in the future.

5.2.4 Multi-model analysis

Multimodal signal processing [57] is an important research field that process signals and combines information from a variety of modalities - speech, vision, language, text- which significantly enhance the understanding, modeling, and performance of human-computer interaction devices or systems enhancing human-human communication.

Hemorrhage affects multiple physiological systems at the same time. Moreover, many of such systems, e.g. cardiovascular and respiratory systems are highly correlated. As such, it is expected to find the highly informative relationship among the physiological signals captured from these systems. As a result, the proposed system will incorporate the

relationship among these physiological signals and their features to better represent the patterns of hemorrhage. More research in this field will be in the plan for the future work.

5.3 Summary

In this chapter, the proposed features in time domain and frequency domain were described. In addition, the chapter provided a brief description of the basic ideas of the Lempel-ziv measure, which is a nonlinear based method. Multi-model analysis was also briefly described.

{CHAPTER 6 Error Correcting Output Code}

6.1 Overview

In the real world applications, learning problems contain both categories of binary learning problems and multiclass learning problems. A typical learning problem involves estimating an unknown function $f(x)$ whose domain input x is the feature vector which can be treated as a discrete set. If the learning problem is a binary one, k will be equal to 2; while in multiclass learning problems, k will be larger than 2. As examples of multiclass learning problems, in digit recognition, the classifier is expected to map each hand-written digit to one of $k = 10$ classes, while in weather forecast, the classifier predicts the weather regime as sunny, windy, rainy, or cloudy.

For multiclass learning problems, there are two general methodologies: 1) direct application of multiclass algorithms, such as artificial neural network (ANN) and decision tree algorithm; 2) using binary learning problem for multiclass learning problem. For the latter, the general trend includes one-against-one method and one-against-all method. An alternative approach is to employ distributed output code [58], pioneered by Sejnowski and Rosenberg. In this approach, a k classes learning problem with n typical attributes is considered as a distributed output code formulated by a $k \times n$ code matrix. The code matrix has k different vectors, called codewords, each having n different elements ('-1' or '1'). For each binary learning problem, according to the code matrix, the example is assigned two classes after going through the binary classifier. Finally, the outputs of these

n binary classifiers make up an n bits string. By attributing the n bits string to the nearest codeword, the example is classified to one of the k possible classes. Consider a digit recognition task as an example, as shown in Table 6.1 (for codewords) and Table 6.2 (for attributes). The digit recognition task has nine classes; each class having six attributes. Therefore, the code matrix has nine rows (corresponding to nine classes) and six columns (corresponding to six attributes). The learning problem with nine classes becomes the combination of six different binary learning problems.

Table 6. 1: Distributed code for the digit recognition task

Class	Codeword					
	vl	hl	dl	cc	ol	or
0	0	0	0	1	0	0
1	1	0	0	0	0	0
2	0	1	1	0	1	0
3	0	0	0	0	1	0
4	1	1	0	0	0	0
5	1	1	0	0	1	0
6	0	0	1	1	0	1
7	0	0	1	0	0	0
8	0	0	0	1	0	0
9	0	0	1	1	0	0

Table 6. 2: Meanings of columns

Column Position	Abbreviation	Meaning
1	vl	contains vertical line
2	hl	contains horizontal line
3	dl	contains diagonal line
4	cc	contains closed curve
5	ol	contains curve open to left
6	or	contains curve open to right

Inspired by distributed code, Dietterich and Bakiri [59] introduced Error Correcting Output Codes (ECOC) as a multiclass learning problem. The approach views machine learning as a communication problem in which the identity of the correct output class for a new example is being "transmitted" over a channel. In the process, due to errors introduced by the finite training sample, poor choice of input features, and flaws in the learning process, the class information is corrupted. By encoding the class in an error correcting code and then "transmitting" the bits, the system may be able to better detect and correct the errors and therefore identify the output class.

From the definition of distributed code and ECOC, it can be seen that the number of columns in distributed codes is identified by the attributes, while for ECOC the number of columns is not fixed.

6.1.1 Conventional Error Correcting Output Codes

ECOC solves multiclass learning problems by combing the output of several binary classifiers and implements multiclass learning problems by an error correcting output code matrix. Fig 6.1 is the framework of Error Correcting Output Codes.

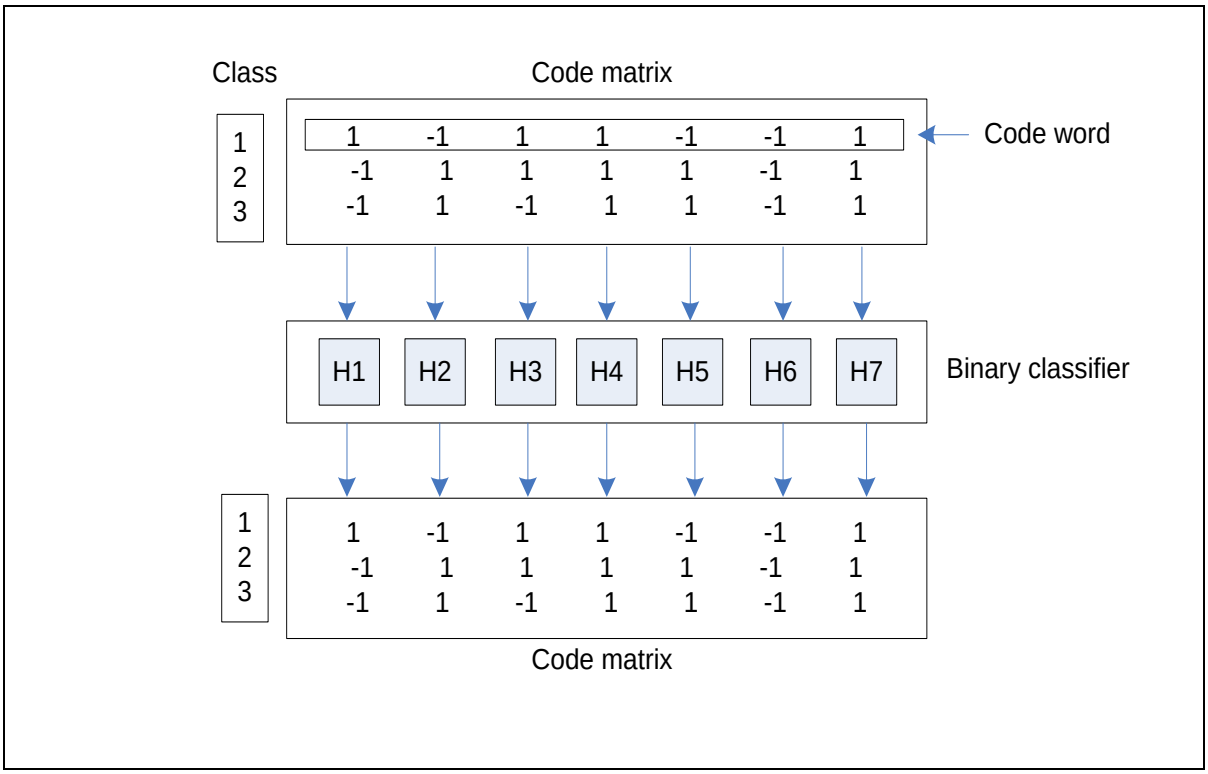


Figure 6. 1: Framework of Error Correcting Output Codes

Assuming a k class problem, what is needed is a matrix with k rows, but the number of columns, n, is not limited as long as it satisfies the basic properties [56]. Each row represents a different class and the number in the matrix can take only two values, -1 or 1.

In the training stage, training samples are redefined according to the error correcting output code matrix. Resorting to each column, each training sample finds the matched row. The row number has the same value as class label of the training sample. Then, new

class label is defined using the value in the corresponding row. As such, training samples are fed into n binary classifiers to train them.

In the testing stage, the testing samples are presented to the trained classifier. The output of the classifier is a vector with n elements, '-1' or '1'. By calculating the Hamming distance between the vector and different rows of error correcting output code matrix, the class label of the testing sample with minimum Hamming distance is selected as the output class.

6.1.2 Analysis

For a code matrix M, if the Hamming distance of code matrix (i.e. the minimum Hamming distance among all different codewords) is d, the code can correct at least $\lfloor (d-1)/2 \rfloor$ single bit errors. In other words, if there are only $\lfloor (d-1)/2 \rfloor$ errors, the nearest codeword will still be assigned as the correct codeword, or equivalently the correct class. That is why this method is called error correcting output codes.

The main idea of ECOC is to form the prediction through "voting" by multiple binary classifiers. The boundaries between pair of classes can be learned and then improved many times according to the code matrix.

Through the verification process, the use of the ECOC for multiclass learning problems can improve bias and variance of prediction [60]. Also, ECOC can be used to estimate probability [61]. At the same time, ECOC provide reasonable robustness against the changes in the size of the training sample and/or the distribution representation of

classes. These characteristics have made ECOC a very popular classification tool in many applications [62,63].

In this study and base on our recent work [64], we have focused on the following improvement of ECOC:

1) Since ECOC could correct $\lfloor (d-1)/2 \rfloor$ errors, in order to make the result as accurate as possible one would like to make d/n as large as possible. Now, for a classification with k classes, there will be at most $2^{k-1}-1$ non-trivial columns after removing complement columns as well as the column with all-ones or all-zeros (all-minus ones). Considering this exponential relationship, if k is too large, the time complexity of the classifier would become unreasonably large. This is a problem that needs to be addressed.

2) Even when using an algorithm to select a subset of columns to make d/n as large as desired, it is difficult to find a systematic algorithm to obtain proper error correcting output code matrix. In addition, there is no practical method to test for optimality of such methods, or decide on how many columns to use, etc.

3) For some complex problems, increasing d/n ratio may not help with solving the problems. In other words, the problem may not be solved just through the choice of error correcting output code matrix, because it also depends on the data as to which binary classifier might work.

6.2 Related Work

To better make use of the ECOC framework for multiclass learning problem, many researchers devote their energies in improving the capabilities of ECOC. As shown in Fig

6.1, ECOC has three main components in the framework: encoding process, code matrix, and decoding process.

As to the encoding process, in order to gain better performance, the encoding step has to satisfy row separation and column separation as illustrated in Dietterich's study [59]. Originally, the improvements made to the ECOC mainly focused on improving the choice of ECOC matrix and revising the form of this matrix. The most known coding strategy is one-versus-all, in which each class is discriminated against the rest of classes. In [65], BCH error-correcting codes are employed as an output representation because of BCH's desirable row and column separation.

In Dietterich's study [59], other methods such as exhaustive codes, column selection from exhaustive codes, randomized hill climbing, and BCH codes were used to improved the ECOC. It was suggested that ECOC may not be best improved while a "one-size-fits-all" design is used for the ECOC matrix. As such, some studies began to design problem-dependent ECOC matrices [66,67]. These studies designed ECOC matrices by considering the distribution of the data. Another example of this type of work is [68] in which ECOC matrix is designed by building a binary tree. The binary tree is built by searching for the best partition set using Sequential Forward Floating Search (SFFS) [69], which is one of sub-optimal search strategies. Such a search strategy can reduce the time complexity of the method to a large extent. The criterion used is Fast Quadratic Mutual Information (FQMI) on the partition set.

To better separate binary classifiers, more advanced and effective methods employ classification accuracy as the criterion of designing the ECOC matrix [70,71]. In these methods, even more trees have to be built. Recently, it has been suggested that finding a large ECOC linear or nonlinear model for too complex learning problems, in which samples from multiple complex classes "diffuse" to each other's space, may not be practical. Therefore, the formation of sub-ECOC [72,73] models is proposed to deal with complex learning problems. In these methods, the complex classes are divided into subclasses using a cluster approach, in particular for the cases in which the base classifier is not capable of distinguishing the classes. Sequential Forward Floating Search based on maximizing the Mutual Information is used to generate the subgroups of problems that cannot be modeled using the original set of classes and without the need of more classifier.

For code matrix, the original ECOC uses only two values in the ECOC matrix, '1' or '-1' . In order to enhance the flexibility of ECOC, Allwein [74] introduces another element, '0', to ECOC matrix, called ternary ECOC. '0' in a column means that the corresponding class is ignored for the binary learning problem. Therefore, the code matrix can be extended to one-versus-one code matrix, dense code matrix, and parse code matrix. The common characteristic of these matrices is that they ignore some classes for each binary classifier. These methods are effective for specific dataset, not for all the dataset.

For decoding matrix, the decoding strategies of original ECOC directed towards two values ECOC matrix. It involves Hamming distance, Inverse Hamming distance [75],

Euclidean decoding. Koby Crammer [76] proposes to relax discrete code matrix into continuous code matrix. As to decoding strategies for ternary ECOC, whether the decoding strategies of the original ECOC are suitable for the ternary ECOC is discussed in [77]. It mentioned that it is not suitable to directly apply the decoding strategies of the original ECOC for the ternary one. The reason is two folds: 1) the existence of the decoding bias, which is the bias introduced by the comparison of a position coded by {-1, 1} with the position containing the zero symbol, and 2) the presence of a dynamic range bias that corresponds to the difference among the ranges of values associated with the decoding process of each codeword. In the ternary decoding framework, all the existing decoding strategies are allocated to four types according to different values of the two biases. The existing ternary decoding strategies include the Attenuated Euclidean Decoding, Loss-based Decoding, Probabilistic-based Decoding [78], etc.

The formulation of Attenuated Euclidean decoding is:

$$AED(x, y_i) = \sqrt{\sum_{j=1}^{n} \left| y_i^j \right| \left| x^j \right| (x^j - y_i^j)^2} \tag{56}$$

where x is the prediction result; y_i is the i th codeword and n is the length of codeword. From the formulation, the AED considers the influence of x and y_i. It means that the measure will not be affected by the zero symbols in y_i. Loss-based decoding is proposed by E. Allwein [74]. In this method, a Loss-function model is applied to calculate the total loss on dataset. The element of the Loss-function is expressed as an expression, called margin. The formulation of Loss-based decoding is:

$$LB(x, y_i) = \sum_{j=1}^{n} L(y_i^j \bullet f^j(x)) \quad (57)$$

where L is the loss function; f is the decision function for binary classifier and $y_i^j \bullet f^j(x)$ is margin. The final decision is achieved by assigning a label to an example when it obtains the least loss. Probabilistic-based decoding uses the probability in each position of code matrix from the training dataset. In this approach, the formulation employs exponential functions. The study conducted by Escalera [77] categorizes all these existing ternary decoding strategies to just one class of approach and proposes two different decoding strategies using two different types of the frameworks: β Density and Loss-Weighted decoding methods. All these methods, however, consider not only the influence of zero symbol but also the dynamic range by making the codeword work in the same dynamic range. The detail of Loss-Weighted Decoding will be explained in the Proposed Method. Experimental results reveal the performance of the methods in this category as well as the suitability of their decoding strategies.

6.3 Proposed Method

The different mechanisms to find the decision boundary make up different machine learning algorithms for not only binary learning problems but also multiclass learning problems. The main challenge in all types of machine learning algorithms is dealing with classes that contain data / sample overlap. The most common ways of dealing with issue include optional data deletion and extraction of extra feature. In the proposed research, attempting to deal with data overlapping problem, the proposed algorithm attempts to

decompose the learning problem, and use the ECOC as a tool to address the data overlapping. While the proposed system is still in progress, in the following sections, the general framework and the finished tasks are illustrated.

6.3.1 Proposed framework of improved ECOC

6.3.1.1 Analysis

Experimental results show that adding problem dependent coding process to the ECOC is an efficient way to obtain more accurate result. However, the higher accuracy will negatively affect the ECOC's error correcting ability. By analyzing the existing research on the ECOC, the proposed attempts to improve the ECOC while complying to the following design idea:

1) Adopt BCH. The proposed work uses a specific type of error correcting output code, BCH, as code matrix. BCH is one of the most widely used error correcting output codes. On one side, it can correct multi-error at the same time, i.e. BCH has desirable row and column separation. On the other hand, the BCH avoids the choice of code matrix while providing reasonable error correcting ability. Finally, through the theory of error correcting output codes (shown in the following part), without knowing the real code, it can be known whether or not the transmitted code is correct. This characteristic of BCH can help form a cascade structure.

2) Adopt SVM as binary classifier. Some of the prevalent phenomena in the practical datasets of learning problem include data /sample imbalance and small sample size. SVM can deal with this type of problems more effectively and will make the system more

robust toward the variations among datasets. The parameters of SVM can be set as reasonable fixed values due to the operations in the next step, which relieves from using cross validation to get the best parameters.

3) Decompose learning problem. For a learning problem, if the samples in classes, at least under some conditions, are normally distributed, the data distribution in both binary learning problems and multiclass learning problems satisfy some relatively traceable behavior. Therefore, the proposed research assumes that the data points in one class make up several normal distributions. These normal distributions have different mathematical expectations and standard deviations that separate them from each other while the union of the sub-classes is still forming the same class. After training, the proposed research allocates the original dataset into different layers (The preliminary decision is three layers). In each layer, the region is divided into two sub-regions. One of sub-regions has high level of confidence on the prediction result, while the other one has low confidence on the prediction result and will transmit the dataset into the next layer to classify.

4) Apply voting in different layers. From the horizontal perspective, the improved ECOC has three layers; while from the vertical perspective, there are data points on which a reliable decision cannot be reached. For such points, the decision is also formed using a "voting" mechanism. The voting process can happen in different layers.

5) Use a parallel strucutre. Multi binary classifiers are designed to work in a relatively independent way. This will allow the proposed improved version of the ECOC to be implemented in a parallel environment.

6.3.2 The framework of improved ECOC

Fig 6.2 is the schematic diagram of our proposed improved ECOC. The proposed improved ECOC will include three layers. In the following part, for layer one and layer two, the general process will be introduced. However, for layer three, it is not entirely finished. The details of the thir layer will be left for the future work.

1) Layer one:

The code matrix is BCH. For each binary classifier, SVM is employed as base function. Using confidence and credibility measure [79] in SVM, the proposed method divides the dataset into two groups: one is in the “sure” region, and the other one is in “unsure” region. Distance from the hyper-plane is employed to identify these regions. The unsure data points are transmitted to the next layer, and at the same time, the structure of the dataset is saved in Structure One .

Structure One is implemented just by setting threshold about confidence. The threshold can take the confidence of point, which is wrongly classified, and has the maximum value on confidence. For binary classifier, it is reasonable to have two thresholds in both ‘positive’ side and ‘negative’ side.

2) Layer two:

The data points transmitted from the previous layer is the overlapping data points. Inspired by the sub-class ECOC, the data overlapping problem is solved by the sub-class ECOC. In the process, EM clustering algorithm [80] is employed. Once the sub-class is formed, Sequential Forward Floating Search (SFFS) is used to find the best partition with the assigned criterion. The most reasonable criterion, i.e. the accuracy of the classifier, is employed. Finally, a binary tree is built to form the final code matrix. At the same time, the structure of dataset is saved in Structure Two.

Structure Two is set using a table. If the code matrix of sub-class ECOC is $k \times n$, the table will have 2^n rows. Each prediction result is allocated into the unit of the table by calculating the decimal from Binary string (prediction result).

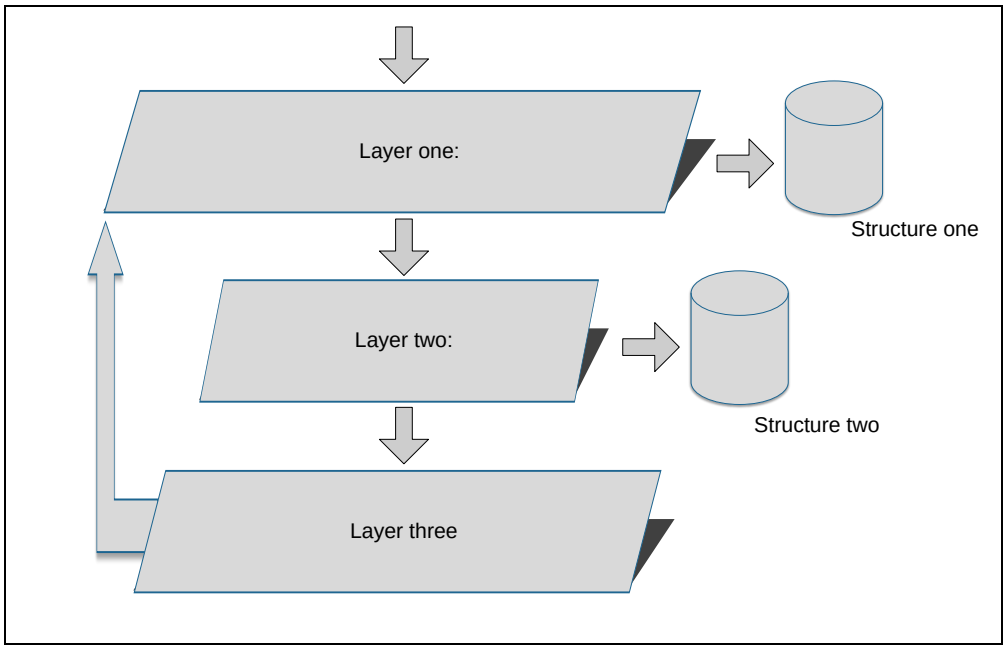


Figure 6. 2: The schematic of improved ECOC

3) Layer three: The exact design of this layer is not finalized, and is a work in progress. This topic will be one of the main focuses on the future work for this research. Therefore, the results are the preliminary results from the first two layers.

In the next sections, some of the main elements of the proposed system are introduced in detail.

6.3.3 BCH

BCH [81] is a family of cyclic codes with good error-correcting ability in which block length is specified and acceptable error value is achieved. For example, BCH(k,n,t) represents k classes, n columns and t errors, which can be corrected. The theory of BCH lies in strong mathmatical basis of group theory. The problem in using BCH is a the insufficiency of structure, because until now the BCH, the way it is used, only supplies the code with a limited length and the value of k. Therefore, in real application this utilization of the BCH results in a mismatch between the number of class labels and the number of labels that BCH can supply. However, the number of the classes in BCH may be chosen to exceed the the number of labels in the application, assuming that it is feasible to choose the same number of rows in the BCH as the labels in the problem in hand without too much manipulation of the problem. This assumption, however, needs to be verified in the future, and for every specific application.

6.3.4 Description of improved ECOC algorithm

Although the complete system is not proposed, in the following section, a detailed description of the algorithms for the first two layers is given.

Table 6.3, depicts the ideas of the improved ECOC. The starting point of the improved ECOC is to solve the decision boundary, which presents challenges for the the training process. The clustering step is only used in the uncertain region, where the wrongly predicted data points are located. In this aspect, it is different from the original subclass idea [72]. Another aspect, which is distinguishes the proposed algorithm, is that the clustering is employed instead of dividing the dataset into two datasets continuously until certain conditions are satisfied. Table 6.4 illustrates the design idea in the second layer. In the proposed algorithm, two inner algorithms, Column Code Binary Tree (CCBT) and Sequential Forward Floating Search (SFFS), are adopted from the work method proposed by O. Pujol [68]. However, unlike Pujol's, in ours algorithm and when using the SFFS, the mutual information between the feature data and its class label is replaced with the accuracy. The reason for doing so is that SFFS achieves a higher level of enery by choosing features based on the specific mechamism/measures and since the overall purpose is to have higher accuracy, we are directly using accuracy as the measure to be increased in the SFFS.

Table 6. 3 Algorithm for Improved ECOC

Improved ECOC: Given a coding matrix G of BCH according to the number of labels in the application. 1) Scale training and testing dataset simultaneously. 2) For each binary classification problem, which correspons to one column of G ▪ Construct new training dataset. ▪ Predict the accuracy $A1$ for original training dataset and predicted

labels $Pvector1$ for original testing dataset.

- Predict the accuracy $A2$ for original training dataset and predicted labels $Pvector2$ for original testing dataset from the second layer.
- If $A1$ is bigger than $A2$, $Pvector1$ is the final predicted labels; otherwise, $Pvector2$ is the final predicted labels.

3) Constrcut the codeword for each sample of original testing dataset.

4) Decoding using hamming distance.

Table 6. 4 Algorithm for the second layer

The second layer: Input: predicted labels for original training dataset, original training dataset, original testing dataset. Output: the accuracy $A2$ for original training dataset and predicted labels $Pvector2$.

1) Calculate two values ,*Tpos* and *Tneg,* to decide safety area and unsafety area.

2) If any of Tpos and Tneg indicates errorless for predicting original training dataset, exit; otherwise, enter next step.

3) Project the data points into higher dimension space according to the used kernel function. These data points will be used in the following step.

4) For data points of original training dataset in unsafety area

- Produce clusters using EM algorithm.
- Produce new code matrix G' using CCBT and SFFS.
- Produce performace matrix H.

5) For each binary classification problem, which correspons to one column of G'

- Construct new training dataset.
- Predict original training dataset and original testing dataset in unsafety area.

6) Construct the codeword using the predicted label both in safety area and

unsafety area for original training dataset and original testing datast.

7) Decoding using Loss-Weighted strategy.

6.3.5 Loss-Weighted Decoding

Loss-Weighted Decoding is proposed by S. Escalera [77]. In this work, using training samples, a performance matrix H is constructed as follows:

$$H(i, j) = \frac{1}{m_i} \sum_{k=1}^{m_i} \varphi(h^j(\rho_k^i), i, j), \tag{58}$$

$$\varphi(x^j, i, j) = \begin{cases} 1, & if \quad x^j = y_i^j \\ 0, & otherwise \end{cases} \tag{59}$$

where (i, j) is the position of a performance matrix H; the performance matrix H has the same size with the code matrix M; h_i is the binary classifier; i corresponds to the i th class; j corresponds to the j th column of the code matrix; m represents the number of training samples in each class.

The performance matrix H is further normalized to M_w, which is considered as a discrete probability density function. The equation (60) is used for the normalization function.

$$M_w(i, j) = \frac{H(i, j)}{\sum_{j=1}^{n} H(i, j)} \qquad \forall i \in [1, \cdots, N], \forall j \in [1, \cdots, n] \tag{60}$$

The following pseudo code describes the algorithm for Loss-Weighted Algorithm. Given a coding matrix M,

1) Calculate the performance matrix H:

$$H(i,j)=\frac{1}{m_i}\sum_{k=1}^{m_i}\varphi(h^j(\rho_k^i),i,j), \tag{61}$$

based on: $$\varphi(x^j,i,j)=\begin{cases}1, & if \quad x^j=y_i^j\\ 0, & otherwise\end{cases} \tag{62}$$

2) Normalize H : $\sum_{j=1}^{n}M_w(i,j)=1, \quad \forall_i=1,\cdots,N$:

$$M_w(i,j)=\frac{H(i,j)}{\sum_{j=1}^{n}H(i,j)} \qquad \forall i\in[1,\cdots,N],\forall j\in[1,\cdots,n] \tag{63}$$

3) Given a test data sample ρ , decode based on:

$$LW(\rho,i)=\sum_{j=1}^{n}M_w(i,j)L(y_i^j\bullet f(\rho,j)) \tag{64}$$

where L is the defined loss function, and f is the output of the binary classifier.

Throughout the analysis of the Loss-Weighted algorithm, the following characteristics are satisfied.

1) The decoding strategy considers not only decoding bias but also the dynamic range bias. The decoding bias is compensated for by the zero in performance matrix; while the dynamic range bias is compensated for by the normalization.

2) The performance matrix can be treated as the performance of each binary classifier. It can also be treated as the probability of the prediction.

3) The formulation (64) shows that the Loss-Weighted decoding fully integrates Loss-based decoding, margin of the binary classifier, and the probability of the prediction together.

6.4 Summary

In this chapter, the new model for multiclass ECOC learning system was proposed and discussed. The chaper also providede the advantages and disadvantages of ECOC as well as the reasons to use and improve the ECOC. In order to gain a more efficient classification tool, the framework of the proposed improved ECOC is introduced with detailed explanations.

{CHAPTER 7 Results and Discussion}

7.1 Results for Preprocessing

7.1.1 Adaptive noise filter

7.1.1.1 Adaptive notch filter

One objective of the proposed system is the removal of unwanted frequencies around 0Hz as well as 60Hz. As the frequencies around zero are excluded the filter acts as a high pass filter. In order to lessen the influence of the time varying components, first, one needs to set a suitable parameter N to obtain a desirable level of time-varying component, β/N. Fig 7.1. shows the value of the time-varying component β/N for different values of N.

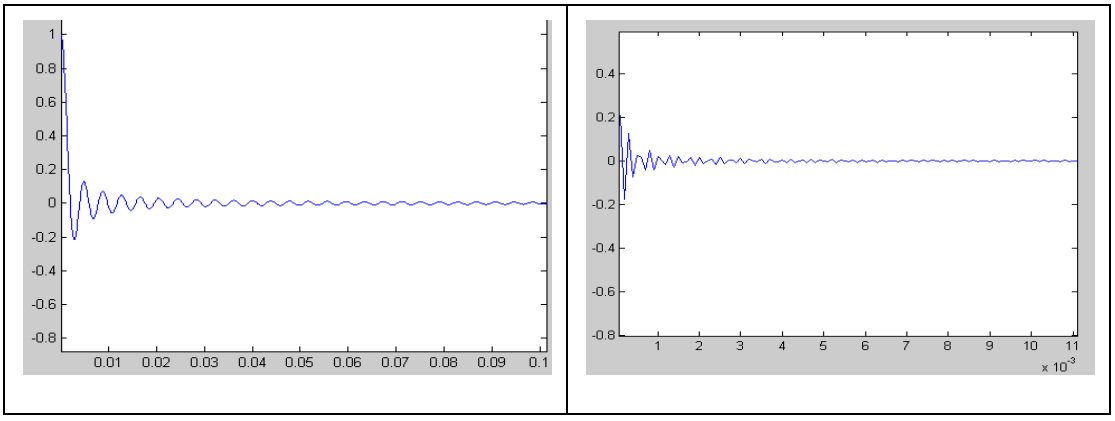


Figure 7. 1. The resulting value of β/N (left: N =256; right: N =4096)

Figure 7.1 indicates that the value of N determines the degree at which the time varying component influences the filter. In general, with the increase in the value of N, this

influence decreases gradually. In this study, the value of N was set to 10000. The parameter ∂ identifies whether or not the adaptation converges. The value of ∂ should be greater than 0 but less than the reciprocal of the largest eigenvalue, λ, of the matrix R, which is defined as the correlation matrix of signal. In this study, the value of ∂ was set to 0.0001. The bandwidth of the filter can be approximated using the following equation:

$$BW = \frac{N\partial C^2}{2T}(rad/s) \tag{65}$$

Fig 7.2 shows the transfer function of the resulting adaptive notch filter, and as expected, this filter acts as a high-pass filter. Note that the value of C provides yet another degree of freedom for this filter design process, and hence, Figure 7.2 presents the transfer function for two different filters formed using two different values of *C*, each resuing to a very different bandwidth. A main advantage of the adpative noth filter used here is that changing the values of parameters N, ∂ and C can provide a wide spectrum of desired filters with diverse shapes of transfer function.

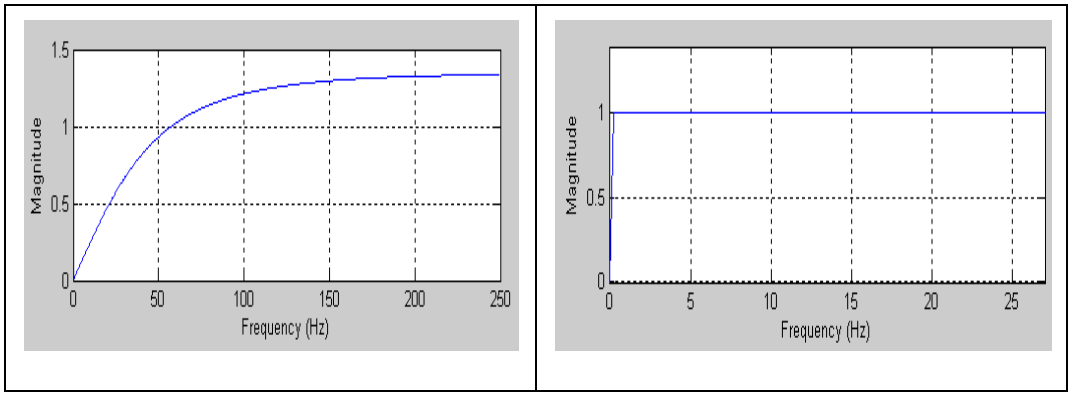


Figure 7. 2. Transfer function for two choices of adaptive notch filters (left: C =1; right: C =0.01)

Adaptive notch filter for frequencies around 60Hz is designed similarly. The parameter N was to 2048, ∂ to 0.001 and C to 0.1. Figure 7.3 depicts the transfer function of the resulting adaptive notch filter.

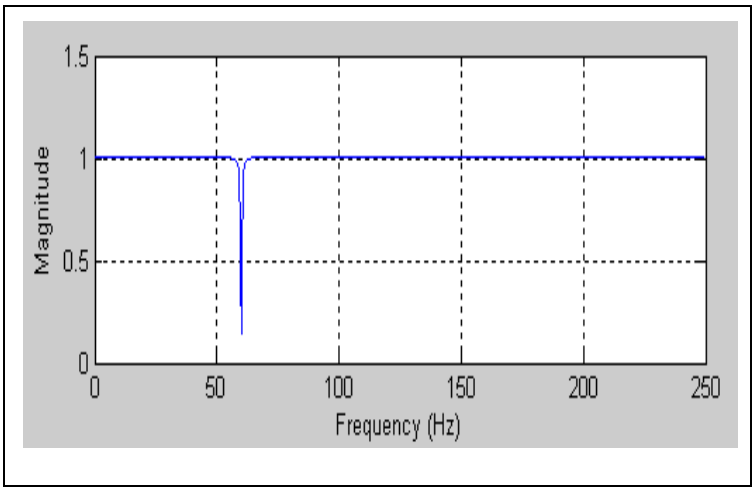


Figure 7. 3. Transfer function of the adaptive notch filter around 60Hz

7.1.1.2 Experimental Results and Problems Analysis

First the results of the reference method are manually examined in all 91 subjects, and a unified 'span' value in the reference method [82] can provide desirable results in removing the baseline wander. This value for all experimental results was 1500. The details of the results are shown for 72 out of 91 subjects in Table 7.1; for these subjects the proposed algorithm can achieve almost the same results as the reference method. The results remaining 19 subjects will be discussed separately.

In Table 7.1, 'subject' is the code of the subject undergoing LBNP; 'shift' and 'elevation' are the values for adjustments from the original independent component (baseline wander) to form the new baseline wander in the horizontal and vertical directions;

'$error_1$' represents the difference between the old baseline wander ($sig1$) and the baseline wander (sig) from the reference method calculated as: $error_1 = (sig_1 - sig)^2 / n$; and finally '$error_2$' represents the difference between the new baseline wander (sig_2) and the baseline wander (sig) from the reference method calculated as: $error_2 = (sig_2 - sig)^2 / n$.

Table 7. 1 Experimental results of removing the baseline wander

subject	shift/ elevation	error1	error2	subject	Shift/ elevation	error1	error2
A013S	290/0	2.0996	0.7847	A159S	290/1	19.8874	0.8131
A026S	250/1	28.1832	2.7037	A160S	290/1	14.3623	2.6499
A096S	300/4	193.9524	3.4495	A161S	260/1	8.8582	6.4787
A106S	300/1	24.3905	1.0727	A162S	300/4	135.0286	3.0056
A109S	300/3	89.1358	3.6282	A163S	290/1	29.5551	2.7541
A110S	290/2	17.9017	1.1614	A164S	300/1	43.3923	3.0052
A111S	300/1	28.955	1.0623	A165S	290/0	51.0465	6.9241
A112S	200/0	107.7542	13.4439	A166S	290/0	31.9213	5.4646
A114S	300/2	203.8138	4.0846	A167S	290/1	9.7597	1.3328
A115S	290/2	81.7942	2.2818	A168S	270/1	22.7897	1.3598
A116S	290/2	256.3747	8.7264	A170S	290/2	93.5265	1.899
A120S	300/2	41.0977	2.4223	A171S	290/0	8.7422	1.5607
A121S	260/1	44.2238	2.279	A173S	350/6	892.829	74.3034
A122S	260/2	101.7592	2.3317	A176S	300/3	209.4986	6.3436
A123S	310/2	700.1481	101.429	A177S	300/3	60.5121	2.6645
A124S	290/1	12.7575	1.3522	A178S	290/1	3.7123	0.9486
A125S	290/1	45.6429	2.6412	A179S	290/4	247.2271	5.138
A127S	310/0	36.8833	11.8224	A180S	250/1	32.0128	2.8609
A128S	290/1	9.1224	1.88	A182S	310/0	20.1471	1.336
A130S	290/2	181.3923	23.0193	A183S	310/0	5.2858	4.0839
A131S	300/2	25.4492	2.6421	A185S	290/0	7.1664	0.9526
A132S	370/4	252.4353	8.5616	A186S	300/1	35.4656	0.8932
A133S	260/2	304.7066	7.4637	A187S	290/1	10.9895	0.8653
A134S	290/1	116.9048	3.77	A189S	300/3	115.7327	4.8387
A135S	300/1	16.3922	1.05	A191S	300/1	26.7803	0.7141
A136S	290/0	3.4748	0.6671	A192S	290/2	9.3222	2.8809
A137S	300/2	144.4347	1.8579	A193S	290/1	16.9436	0.9469
A140S	290/0	23.9724	2.1641	A195S	300/0	27.7014	1.7794
A142S	290/1	14.5089	0.2205	A207S	290/1	55.1891	4.9226
A143S	290/1	155.3859	3.6707	A208S	310/6	620.3234	8.0999
A148S	300/1	50.6959	2.6757	A210S	400/2	23.6969	0.5595

A149S	300/1	27.2665	1.101	A212S	290/2	36.63757	1.4766
A151S	300/1	56.7045	1.999	A216S	290/2	241.5044	11.7279
A154S	290/2	324.4399	10.0313	A217S	290/1	5.5229	0.3386
A155S	300/1	42.6266	0.8791	A219S	290/2	173.1734	7.0318
A156S	290/2	539.7357	31.3238	A220S	300/2	77.4627	3.2468

As it can be seen in Table 7.1, for all cases $error_2$ is significantly smaller that $error_1$ which shows the impact that the method in "purifying" the baseline wander and creating a better estimate of the drift. In order to better assess the performance of the proposed method in removing the baseline wander, more analysis are conducted on the results.

Figure 7.4 and Figure 7.5 show that the shift and elevation for all 72 subjects. As it can be seen, both of these variables, are almost the same for all subjects and do not change across different subjects (x-axis) or vary in a small scope. This observation illustrates the reason to adjust the parameters between the old baseline wander and the new baseline wander.

Figure 7. 4 Value of 'shift' that adjust the old baseline wander to form the new one for all 72 subjects

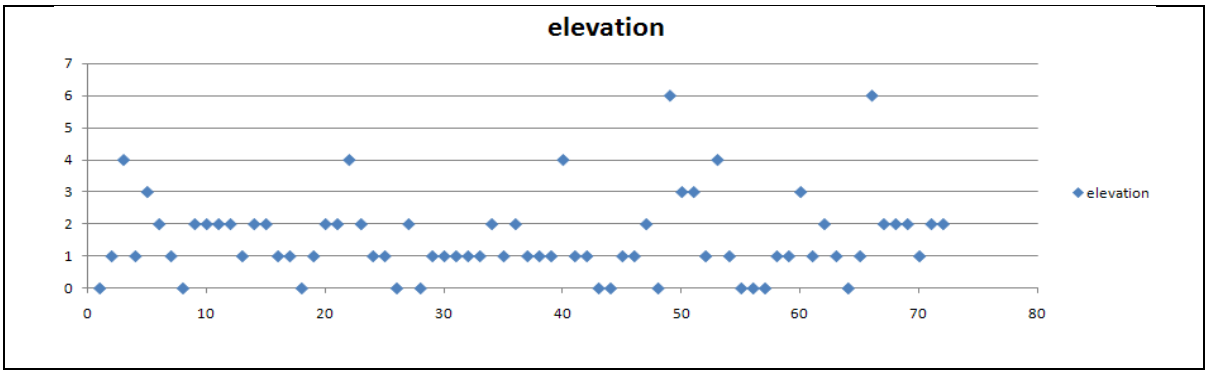


Figure 7. 5 Value of 'elevation' that adjust the old baseline wander to form the new one for all 72 subjects

Figure 7.6 shows the error reduction in 72 subjects after adjusting shift and elevation value. It can be seen that in all these cases the errors decreases significantly after adjusting the baseline wander compared with the baseline wander. The average percentage of error reduction $AverE$ reaches up to 90.13%. The formulation of the average percentage of error reduction is shown in the following:

$$percentage\,(i) = \frac{(error_1 - error_2)}{error_1} \qquad i \in [1, n] \tag{66}$$

$$AverE = \sum_{i=1}^{i=n} percentage\,(i) \tag{67}$$

where i is the index of subject, and n is the total number of subjects.

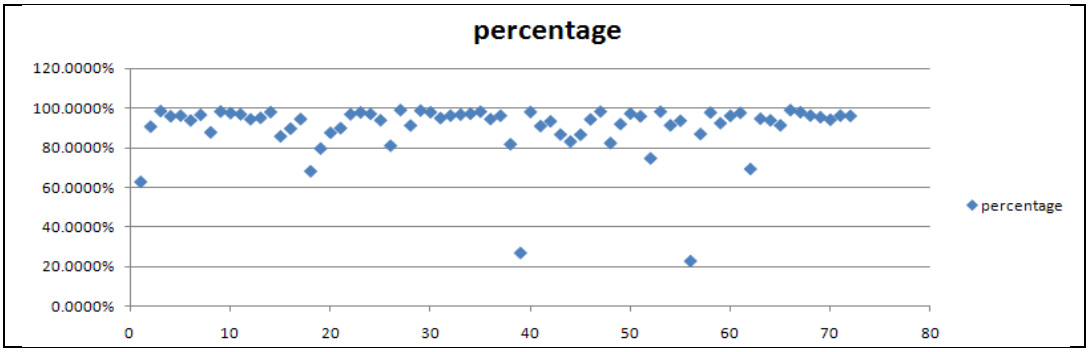


Figure 7. 6. Improved percentages of error after adjustment

As mentioned above, on the ECG of 19 subjects, the results of the proposed method and the reference method are not close, i.e. the value of $error_2$ is a big value. As mentioned above we have visually inspected all 91 cases. By examining the signals for these 19 cases, the high value of $error_2$ does not come from the inability of the proposed method to remove the baseline wander. Slightly adjusting some parameters, such as the number of independent components, can alleviate the situation.

7.1.2 QRS Complex detection

In order to verify the effectiveness of our algorithm, the MIT dataset is used as the testbed. The MIT dataset [83] contains the total of 48 subjects. Each subject has an ECG recording of about thirty minutes (sample rate: 360Hz). The variation of ECG signals is due to presence of arrhythmia. Subjects 106, 231 and 232 have the condition in which the RR interval changes heavily while for subjects 104, 106, 108, 114, 116, 200, 203, 207, 208, 215, 222, 228 and 231, the amplitude of R waves varies significantly. As to the other subjects, their ECGs are often noisy. Table 7.1 presents the result of our approach. The technique results in an accuracy, sensitivity, and specificity of 99.89%, of 99.89%, of 99.92%, respectively. Moreover, the point insertion process corrects the placement of R in 97.27% of the originally missed R waves.

For further validation of our approach, the results were compared with those of other QRS complex detection algorithms. For comparability, several issues were considered: 1) each algorithm uses all the subjects in the MIT dataset; 2) each algorithm is a different but major type of detection algorithm (e.g. wavelet, filter, or filter bank); 3) the

preprocessing of ECG signal is also done differently. Table 7.2 represents the statistical comparison of different algorithms.

Table 7. 2: The performance of the proposed QRS complex detection algorithm (1st column: subject name; 2nd column: #total beats; 3rd column: #wrongly detected beats; 4th column: #missing detected beats; 5th column: period controller; 6th column: amplitude controller; 7th column: # total beats before insertion))

subject	total	Wrong	missing	A	b	beats before insertion
100	2264	0	0	0.2	1.8	2265
101	1861	0	0	0.2	1.8	1861
102	2179	0	0	0.2	1.8	2179
103	2075	0	2	0.2	1.8	2072
104	2218	0	4	0.2	3	2052
105	2567	0	2	0.2	1.8	2564
106	2006	0	3	0.2	3	1952
107	2114	1	10	0.2	3	2105
108	1754	0	1	0.2	1.8	1743
109	2522	0	0	0.2	1.8	2512
111	2126	10	0	0.2	3	2063
112	2531	1	0	0.2	1.8	2531
113	1788	0	0	0.2	1.8	1788
114	1879	5	1	0.2	3	1112
115	1945	0	0	0.2	1.8	1943
116	2376	0	4	0.2	1.8	2364
117	1531	0	0	0.2	1.8	1531
118	2269	0	0	0.2	1.8	2269
119	1981	0	0	0.2	3	1981
121	1854	0	1	0.2	1.8	1853
122	2466	0	0	0.2	1.8	2467
123	1512	0	0	0.2	1.8	1512
124	1612	1	0	0.2	3	1582
200	2582	2	3	0.2	3	2577
201	1919	4	8	0.2	3	1905
202	2122	0	0	0.2	1.8	2109
203	2657	5	20	0.2	3	2384
205	2624	1	0	0.2	1.8	2623
207	2027	0	1	0.2	1.8	1916
208	2912	0	11	0.2	1.8	2889
209	2997	0	0	0.2	3	2997
210	2588	1	8	0.2	3	2527
212	2739	0	0	0.2	1.8	2739
213	3239	0	1	0.2	1.8	3230
214	2246	0	2	0.2	3	2239
215	3340	0	1	0.2	3	3344
217	2184	0	5	0.2	1.8	2136
219	2144	1	0	0.2	1.8	2140
220	2039	0	0	0.2	1.8	2039
221	2416	0	0	0.2	3	2405
222	2397	3	0	0.2	3	2300
223	2588	1	4	0.2	3	2435
228	2026	15	23	0.3	3	878
230	2248	2	0	0.2	1.8	2248
231	1592	28	0	0.2	3	1566

232	1787	5	0	0.2	3	1777
233	3056	0	2	0.2	1.8	3043
234	2744	0	0	0.2	1.8	2743
	108643	86	116			105490

Table 7. 3: The statistical comparison of different QRS complex detection algorithms

Method	# TP	#FP	#FN	#total errors
[29]	90909	406	374	780
[26]	116079	58	166	224
[77]	116004	133	174	307
[24]	109019	248	340	588
[25]	109375	240	239	479
Our algorithm	108521	86	116	202

7.1.3 Systole&Diastole detection

In order to verify the effectiveness of our algorithm, LBNP is used as the testbed. In LBNP dataset, 93 subjects are employed. Each subject has an ABP recording of about forty seven minutes (sample rate: 500Hz). The variation of ABP signals is due to presence of volume loss in the upper body. Table.7.3 is the result of our algorithm. The accuracy is 99.95%.

Table 7. 4: The performance of the proposed Systole&Diastole detection

Subject	Total S	Missing S	False S	Total D	Missing D	False D
A103S	3871	0	0	3871	0	0
A026S	2798	0	0	2798	0	0
A096S	3676	0	0	3676	0	0
A106S	2700	0	0	2700	0	0
A109S	3481	3	0	3480	3	0
A110S	2097	1	0	2096	1	0
A111S	2871	1	0	2870	1	0
A112S	1797	1	0	1797	1	0
A113S	3092	4	25	3092	4	25
A114S	3138	0	0	3138	0	0
A115S	3468	0	0	3468	0	0
A116S	2113	0	0	2114	0	0
A117S	3331	0	0	3331	0	0
A120S	3000	4	0	2999	4	0
A121S	2666	0	0	2666	0	0
A122S	2795	0	0	2794	0	0
A123S	2429	0	0	2429	0	0
A124S	2654	2	1	2654	2	1
A125S	3453	0	1	3453	0	1
A126S	2859	0	0	2860	0	0
A127S	3039	2	0	3039	2	0
A128S	3081	0	0	3081	0	0

A130S	3105	0	0	3104	0	0
A131S	2843	0	0	2843	0	0
A132S	2540	0	0	2540	0	0
A133S	2925	0	0	2924	0	0
A134S	1751	0	0	1751	0	0
A135S	3132	0	0	3132	0	0
A136S	3183	0	1	3182	0	1
A137S	1798	0	0	1799	0	0
A140S	2092	0	1	2091	0	1
A141S	3610	0	0	3609	0	0
A142S	2942	0	0	2942	0	0
A143S	2185	4	1	2185	4	1
A144S	4089	5	0	4089	5	0
A147S	2245	0	0	2245	0	0
A148S	2352	0	0	2351	0	0
A149S	2666	0	0	2666	0	0
A150S	2249	0	0	2248	0	0
A151S	2116	0	0	2115	0	0
A153S	2451	1	1	2451	1	1
A154S	2951	0	0	2951	0	0
A155S	3314	0	0	3313	0	0
A156S	2560	0	0	2559	0	0
A157S	2145	0	0	2145	0	0
A158S	3490	0	0	3490	0	0
A159S	2375	0	0	2375	0	0
A160S	2353	0	0	2353	0	0
A161S	2486	0	0	2485	0	0
A162S	2739	0	0	2739	0	0
A163S	3356	0	16	3356	0	16
A164S	2738	6	0	2737	6	0
A165S	2684	5	0	2683	5	0
A166S	2215	0	0	2215	0	0
A167S	2276	0	0	2277	0	0
A168S	2484	0	0	2484	0	0
A170S	2745	0	2	2745	0	2
A171S	3831	0	0	3832	0	0
A172S	2108	0	0	2108	0	0
A173S	2395	0	0	2394	0	0
A174S	3824	0	0	3824	0	0
A175S	2490	0	0	2490	0	0
A176S	3430	57	4	3430	57	4
A177S	3385	6	0	3385	6	0
A178S	3079	0	0	3079	0	0
A179S	2325	0	0	2325	0	0
A180S	4015	0	10	4014	0	10
A181S	3641	0	0	3641	0	0
A182S	2503	2	0	2502	2	0
A183S	3903	2	3	3903	2	3
A184S	2529	0	1	2529	0	1
A185S	3588	0	0	3588	0	0
A186S	3066	5	4	3065	5	4
A187S	2840	0	0	2840	0	0
A189S	2591	2	0	2590	2	0
A190S	2458	0	0	2458	0	0
A191S	2713	0	0	2713	0	0
A192S	2184	0	0	2184	0	0
A193S	2846	0	0	2846	0	0
A194S	3001	0	0	3000	0	0
A195S	2490	2	46	2490	2	46

A207S	4156	0	0	4156	0	0
A208S	2305	0	0	2304	0	0
A209S	2119	0	0	2119	0	0
A210S	2291	0	0	2291	0	0
A212S	1453	0	0	1453	0	0
A213S	3386	0	0	3386	0	0
A214S	3744	1	0	3744	1	0
A216S	1931	0	0	1930	0	0
A217S	1276	0	0	1275	0	0
A218S	2214	4	0	2214	4	0
A219S	2544	7	1	2545	7	1
A220S	2242	10	0	2241	10	0
	258490	137	118	258468	137	118

7.2 Results for Feature Extraction

After the completion of QRS complex detection and Systole&Diastole detection, most of features can be extracted. In the following part, the result of improved Lempel-ziv is presented.

7.2.1 Conventional Lempel-Ziv

The dataset used for this study is also the LBNP dataset. 93 subjects are employed. Each subject has an ECG recording of about forty seven minutes (sample rate: 500Hz). The variation of the ECG signals is due to the presence of the simulated blood loss. For the LBNP dataset, there are eight stages. The objective of calculating Lempel-ziv is to distinguish the mild, moderate, and severe stages from these 8 LBNP stages for each of the subjects.

The calculation of Lempel-Ziv obeys the following rules: 1) The raw ECG signal is used to calculate Lempel-Ziv; 2) In each stage, the window used to extract a portion of signal to be analyzed has a fixed size; 3) There is no overlap between different windows; 4) The threshold used is the medium value of the signal in the window. Fig 7.7 is mean-variance

plot of the result of Lempel-Ziv in different stages. X-axis stands for the LBNP stage, and Y-axis stands for the scope of Lempel-Ziv with mean and variance. Fig 7.8 is mean-variance plot of the result of Lempel-Ziv in three stages (Mild, Moderate and Severe) on raw signal (left) and RR signal (right). From the figure, it is shown that Lempel-Ziv cannot distinguish different LBNP stages from each other.

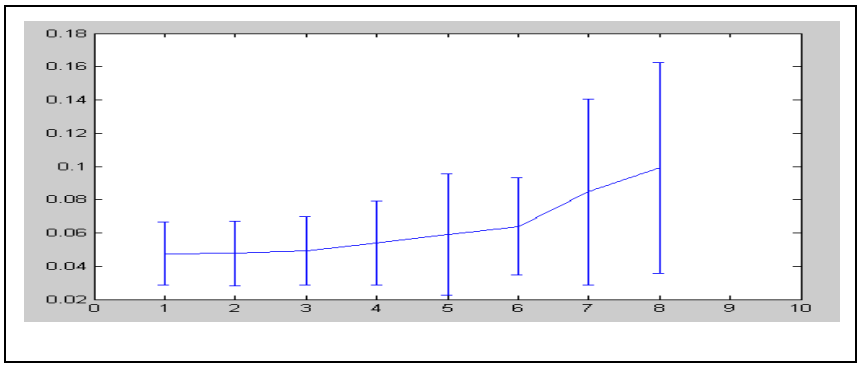


Figure 7. 7: The result of Lempel-ziv in different stages

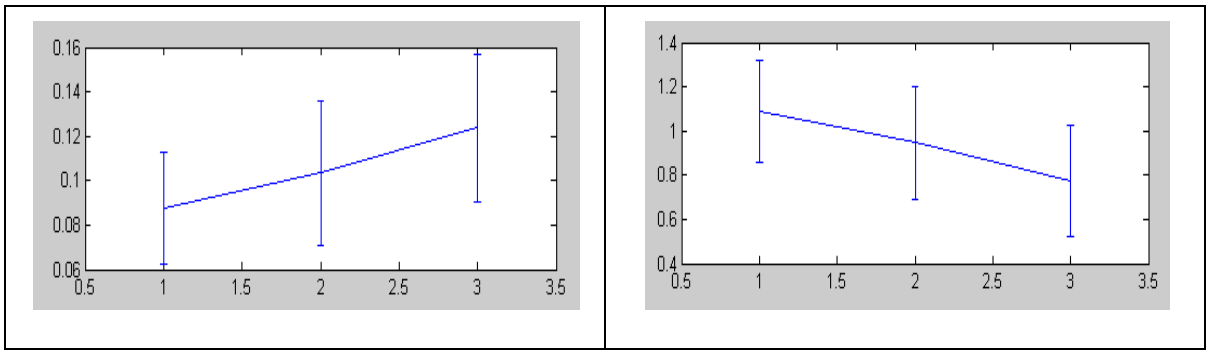


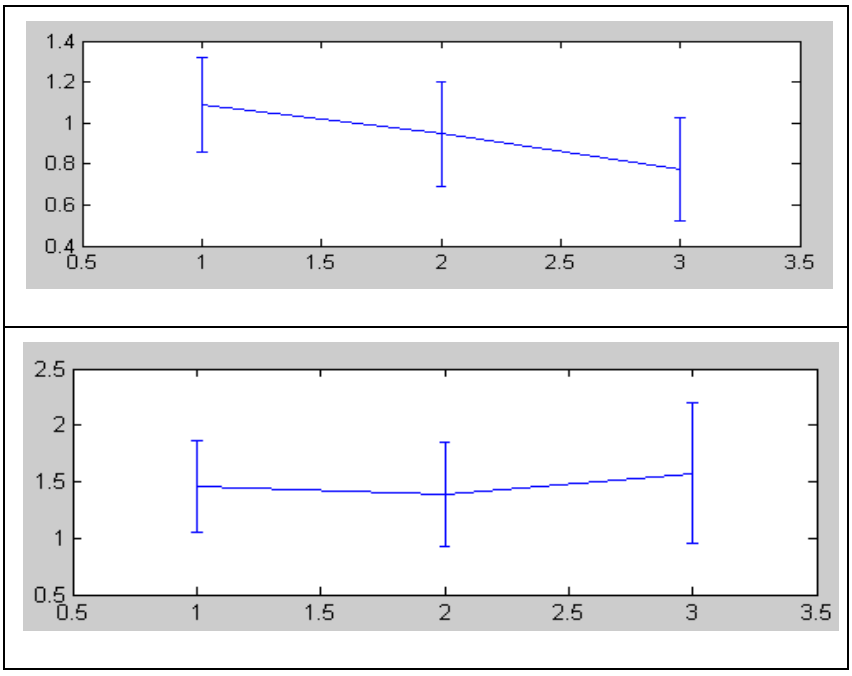
Figure 7. 8 Mean-variance plot of Lempel-ziv in three stages on raw signal (left) and RR signal (right)

7.2.2 Improved Lempel-Ziv

To verify the effectiveness of improved Lempel-Ziv, Both mean-variance plot of improved Lempel-Ziv and accuracy are presented.

Fig 7.9 shows the mean-variance plot of improved Lempel-ziv in three stages based on RR signal using different elements. From top to bottom, they are elements with one digit, five digits, twenty one digits, and forty six digits, respectively. Seen from the continous figures, it is evident that the captured information increases as the size of elements enlarges.

To better illustrate the effectiveness of improved Lempel-Ziv, the accuracy is also compared with the conventional Lempel-Ziv. The calculation of accuracy in classifying mild, moderate and servere is made based only on one feature, that is the conventional Lempel-Ziv or the improved Lempel-Ziv. Several classifiers are employed to test the result, as shown in Table 7.5. From the table, it can be seen that the improved Lempel-Ziv can better represent the patterns of volume loss.



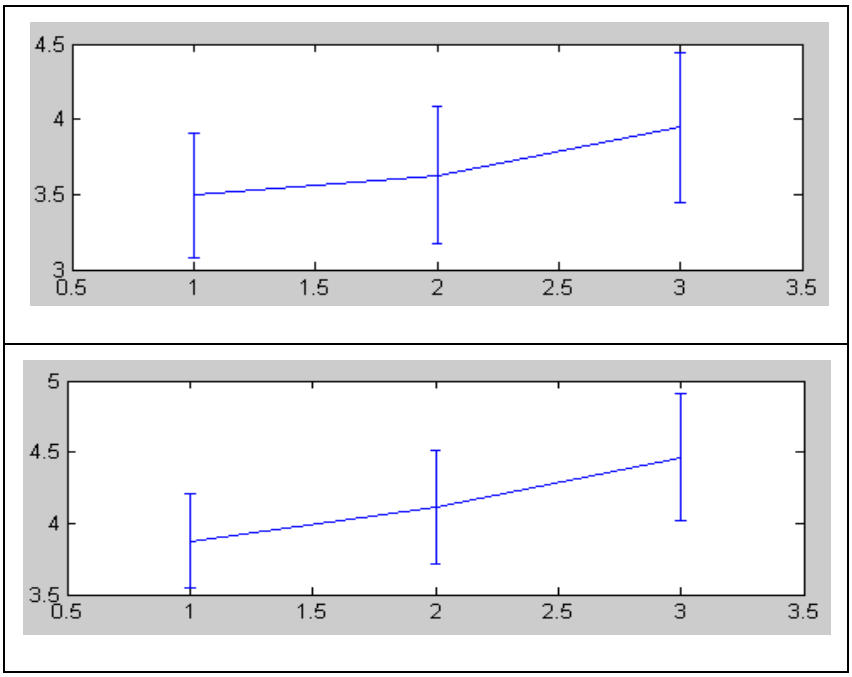


Figure 7. 9. Mean-variance plot of improved Lempel-Ziv using different element

Table 7. 5. Accuracy comparison between improved Lempel-Ziv and conventional Lempel-Ziv

Type	J48	Multilayer Perceptron	ByesNet	DecisionTable	Decision Stump	RBFNetwork
Improved Lempel-Ziv	63.6	63.2	63.3	63.7	62.3	63.2
Conventional Lempel-Ziv	64.2	54.2	59	63.3	54.9	54.8

7.3 Results for Error Correcting Output Codes

The improved version of Error Correcting Output codes includes three layers. The design of the third layer is still in process. In the following sections, the intermediate results verifying the value in the proposed Error Correcting Output Codes are presented.

7.3.1 Cascade Structure

The results of our previous research show that the building of a cascade classifier can improve accuracy of the ECOC, and will not result in overfitting. One of the reasons for such improvements can be the property of BCH, that is, without comparing the predicted and actual class labels, it can be identify whether testing samples are classified correctly or not.

For testing and validation, the cascade structure is tested against the real datasets from the SVM website. Table.7.6 shows the attributes in the datasets. In order to verify effectiveness of the proposed method, the results of the following two ECOC classifiers are compared with each other: 1) The ECOC with just BCH matrix (classifier#1); 2) The proposed ECOC with BCH matrix equipped with the decoding step and cascade structure (classifier#2). The two classifiers have the same BCH and both of them are combined with the binary classifier support vector machine (SVM), which uses the same parameters in different datasets. Error correcting output codes used are shown in Table.7.7.

Table 7. 6: Experimental dataset

Name	class	features	Training samples	Testing samples
satimage	6	36	4,435	2,000
Shuttle	7	9	43,500	14,500
Usps	10	256	7,291	2,007
Letter	26	16	15,000	5,000
Dna	3	180	2,000	1,186
Vowel	11	10	528	462

Table 7. 7: Error correcting output codes

Name	Classifier#1	Classifier#2
satimage	BCH(31,6)	BCH(31,6,7)
shuttle	BCH(63,7)	BCH(63,7,15)
usps	BCH(511,10)	BCH(511,10,121)
letter	BCH(63,30)	BCH(63,30,6)
dna	BCH(15,5)	BCH(15,5,3)
vowel	BCH(31,11)	BCH(31,11,5)

Table.7.8 shows error correcting performance of two different classifiers ((1): Minimum Hamming distance, (2): Minimum error correcting ability, and (3): Automatic error correcting ability)

Table 7. 8: Error correcting performance

Name	Classifier#1			Classifier#2		
	(1)	(2)	(3)	(1)	(2)	(3)
satimage	15	7	0	15	0	7
shuttle	31	15	0	31	0	15
usps	255	127	0	255	0	121
letter	13	6	0	13	0	6
dna	7	3	0	7	0	3
vowel	11	5	0	11	0	5

Table.7.9 shows the accuracies of classifier#1 and classifier#2. From the table, it can be seen that the accuracies obtained in all five datasets for the proposed method are equal or higher than those of the other method, although the degree of improvement of accuracy may not be significant. The difference among the two methods is in their capacity to deal with error. Specifically, when the strength of noise and error is more than the maximum error correcting ability of the method, changes / noise would result in decoding a new code into a wrong original codeword.

Table 7. 9: Accuracy of two classifiers

Name	Classifier#1	Classifier#2
satimage	91.45%	91.50%
shuttle	99.67%	99.67%
usps	95.02%	95.07%
letter	89.94%	81.80%
dna	91.65%	91.74%
vowel	55.19%	56.49%

7.3.2 Confidence Measure

In our previous work, we focused on the observation by the community that prediction of classification confidence as the degree of reliability is a much needed step. In the

proposed improved ECOC, decomposing original dataset into two parts is a step employed. Confidence can be used as the criterion to split dataset.

The prediction with confidence is called conformal prediction. The original conformal prediction assumes that the training and testing samples come from the same distribution. Each sample has an associated conformity score or non-conformity score. Both conformity and non-conformity scores can be interpreted as measures of 'supportiveness' of the sample. As such, the 'supportiveness' is also called confidence. The confidence is formed through calculating n times if there are n testing samples and each time just one testing sample is supplied to the classifier. However, if there are n testing samples and all of them are supplied at the same time, it will take 2^n times to get all the confidence measure for all the testing samples.

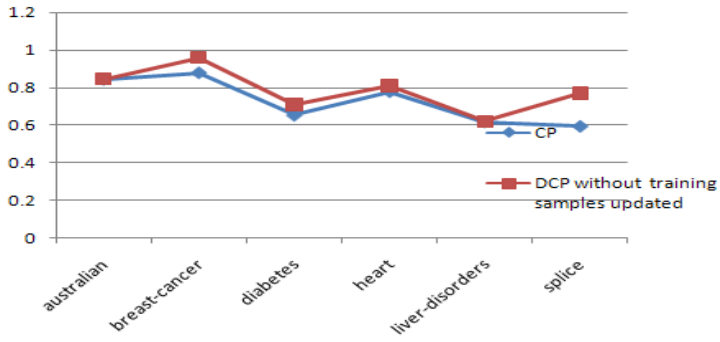
As discussed in previous work, the original conformal prediction has high computational complexity. In order to provide not only higher accuracy and lower computational complexity but also more confidence in the resulting predictions, we proposed the concept of dynamic conformal prediction (DCP), and designed a new form of confidence. The experimental results verify the effectiveness of the dynamic conformal prediction.

Both clinical and non-clinical datasets are employed to test and verify the effectiveness of our proposed method. Clinical datasets (94088 samples with 18 features) are based on MIT-BIT to detect arrhythmia. Non-clinical datasets (five datasets 'australian' (690/14), 'breast-cancer' (683/10), 'diabetes' (768/8), 'heart' (270/13), 'liver-disorders' (345/6), 'splice' (1000/60)) from Lib-SVM website were used. Non-clinical and clinical datasets are

employed to test the proposed methods. For base classifier, SVM, the best parameters are set, which are selected based on 10-fold cross validation.

7.3.2.1 Non-clinical applications

The nature of data has no time-sequence, thus the DCP without updated training samples was used. The results (Fig.7.10) of accuracy, time complexity, and average errors are presented in (a), (b) and (c), respectively. The purpose of average errors (errors: each time, use the confidence of a data point as the current confidence. The error is the number of samples if correctly classified if confidences are below its current confidence, and wrongly classified if confidences are above its current confidence) is to test how reliable the confidence measure is. The results show that the accuracies, running time and average errors using proposed the method are superior to the original method.



(a) accuracy

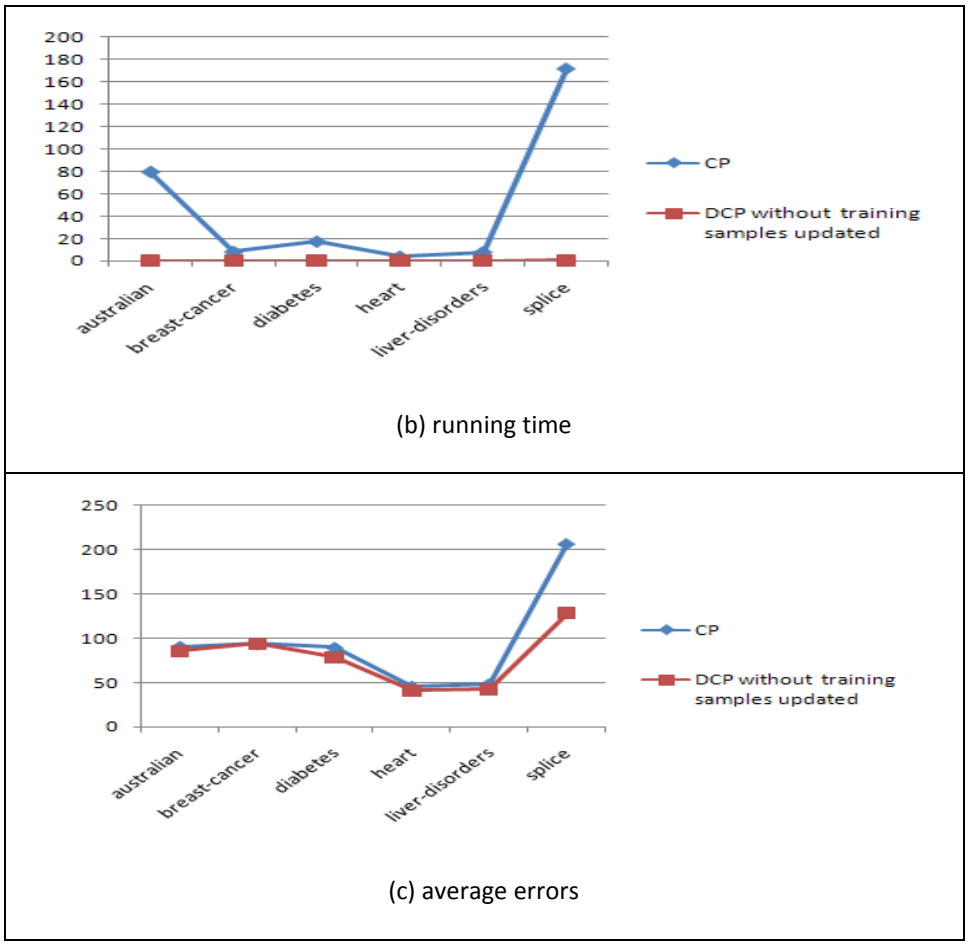


Figure 7. 10: The result of accuracy, running time and average errors

In Fig.7.11, horizontal axis stands for conformity or non-conformity scores; and the vertical axis shows confidence. Fig.7.11 shows the confidences curves of all data for both the original method (Fig 7.11.a and 7.11.c) and our proposed method (Fig 7.11.b and 7.11.d). Five datasets ('australian', 'breast-cancer', 'diabetes', 'heart', 'liver-disorders') in six have relatively similar results for both methods because the confidence increases as conformity (or non-conformity) score increases. However, one dataset ('splice') shows abnormal behavior in the original method (Fig 7.11.c) as when conformity (or non-conformity) score is around the same value, confidence does not show a clear pattern.

However, our measure for the same dataset (Fig. 7.11.d) shows a clear and well-accentuated pattern of confidence.

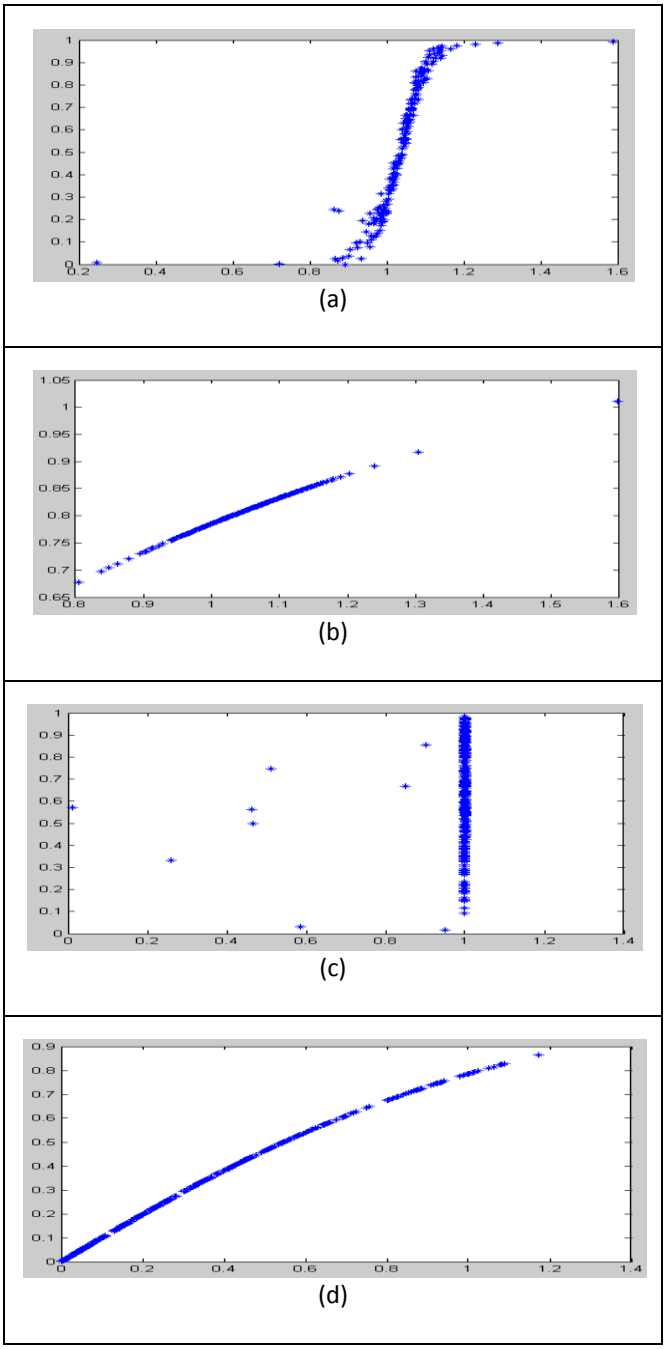


Figure 7. 11: The shape of confidences for all data in one dataset

7.3.2.2 Clinical application

The total number of ECG beats in the MIT-BIT dataset is 40,000 beats, which includes 28987 healthy beats and 11,013 arrhythmia beats. First, feature extraction is performed on each beat using the methodology described in [84,85], and totally 17 features are extracted. Then using the extracted features, DCP with updated training samples is used for classification. The resulting accuracy is 98.23% which is significantly higher than that of the original CP which is 94.75%. Also, the time complexity in our proposed method is significantly reduced from 2,750 seconds (in CP) to 54,063 seconds.

7.3.3 Solving Data Overlapping using sub-ECOC

Although it is not finished for the design of the whole system, the finished part has the characteristics of the first two layers, as mentioned above. To illustrate the result of the finished parts, in the experiment, two datasets [83] (vowel and satimage) are adopted. As a comparision, conventional ECOC and direct multiclass classifier, libsvm, from weka are employed. In order to level the two methods, the parameters in libsvm are the same as the parameters in binary classifiers of conventional ECOC and the improved ECOC. They use three degree polynomial kernel function with cost *C* and gamma *G* changing in different values.

7.3.3.1 Comparision based on accuracy

I. Vowel

The dataset has eleven classes. In each subject, it includes 10 features. And, there are 528 training data points and 462 testing data points. The employed BCH is BCH

(15,11,1). The cost *C* changes from 0.01 to 1000 and gamma *G* change from 0.0001 to 10.

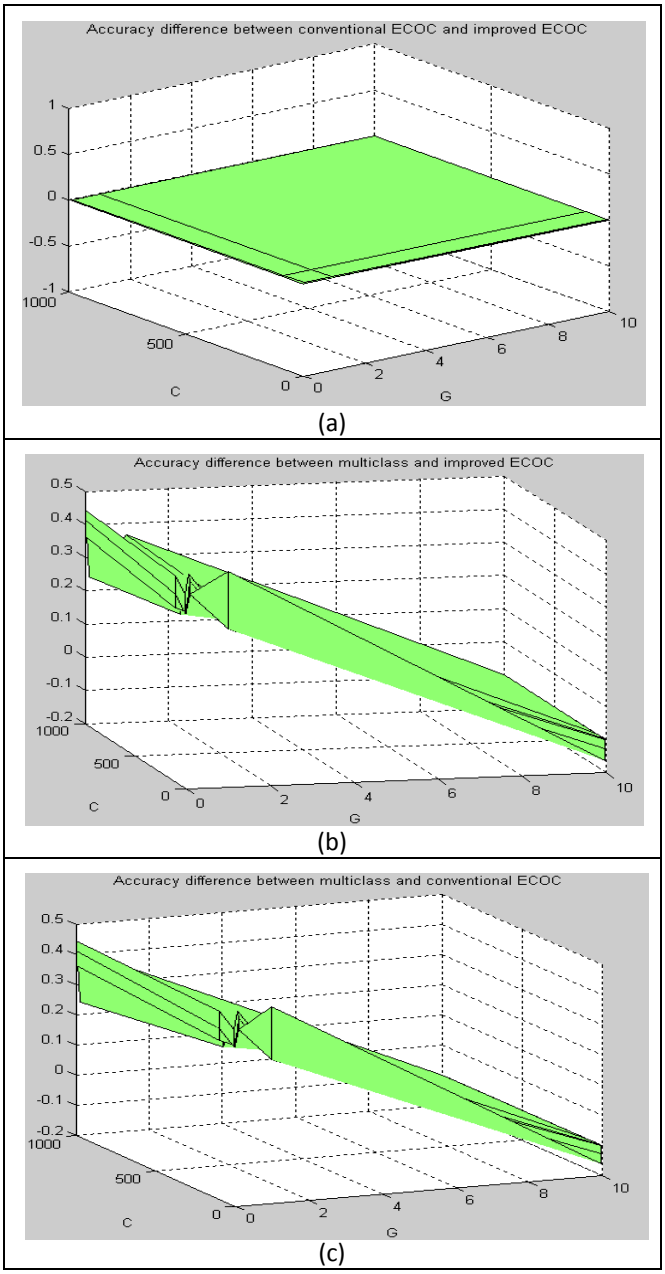


Figure 7. 12. Accuracy comparision between two methods (Vowel)

In figure.7.12, it shows the accuracy comparision. X-axis represents cost *C*; Y-axis indicates gamma *G*; and Z-axis shows the accuracy difference between two methods. As shown in figure.7.12, the sub figure (a), (b), and (c) give the accuracy comparision between conventional ECOC and improved ECOC, between multiclass classifier and conventional ECOC, and between multiclass classifier and improved ECOC, respectively. From the figures, it tells us that, based on Vowel dataset, the conventional ECOC and improved ECOC get most the same results. At the same time, the accuracy presents difference between conventional ECOC/improved ECOC and multiclass classifier according to each combination of *C* and *G*.

In order to see more clearly about the result, table 7.10 and table 7.11 show the accuracy value for each combination of *C* and *G* (row: *C*, column: *G*).

Table 7. 10 accuracy of improved ECOC and conventional ECOC

0.3463	0.0952	0.0909	0.0909	0.0909	0.0909
0.3117	0.2641	0.0909	0.0909	0.0909	0.0909
0.2965	0.3398	0.0909	0.0909	0.0909	0.0909
0.2965	0.3463	0.0952	0.0909	0.0909	0.0909
0.2965	0.3117	0.2641	0.0909	0.0909	0.0909
0.2965	0.2965	0.3398	0.0909	0.0909	0.0909

Table 7. 11 accuracy of multiclass classifier

0.1840	0.5450	0.4760	0.4240	0.4220	0.4220
0.1840	0.5410	0.4760	0.4240	0.4220	0.4220
0.1970	0.6260	0.5130	0.4180	0.4220	0.4220
0.1970	0.6490	0.5390	0.4740	0.4160	0.4220
0.1970	0.6620	0.5690	0.5130	0.4570	0.4160
0.1970	0.6620	0.5820	0.5350	0.5040	0.4550

II. Satimage

The dataset has six classes. In each subject, it includes 36 features. And, there are 4435 training data points and 2000 testing data points. The employed BCH is BCH (31,6,7). The cost *C* changes from 0. 1 to 100 and gamma *G* change from 0.001 to 1.

In figure.7.13, it shows the accuracy comparision. X-axis represents cost *C*; Y-axis indicates gamma *G*; and Z-axis shows the accuracy difference between two methods. As shown in figure.7.13, the sub figure (a), (b), and (c) give the accuracy comparision between conventional ECOC and improved ECOC, between multiclass classifier and conventional ECOC, and between multiclass classifier and improved ECOC, respectively. From the figures, it tells us that, based on Satimage dataset, the conventional ECOC and improved ECOC produce different results. At the same time, the accuracy presents difference between conventional ECOC/improved ECOC and multiclass classifier according to each combination of *C* and *G*.

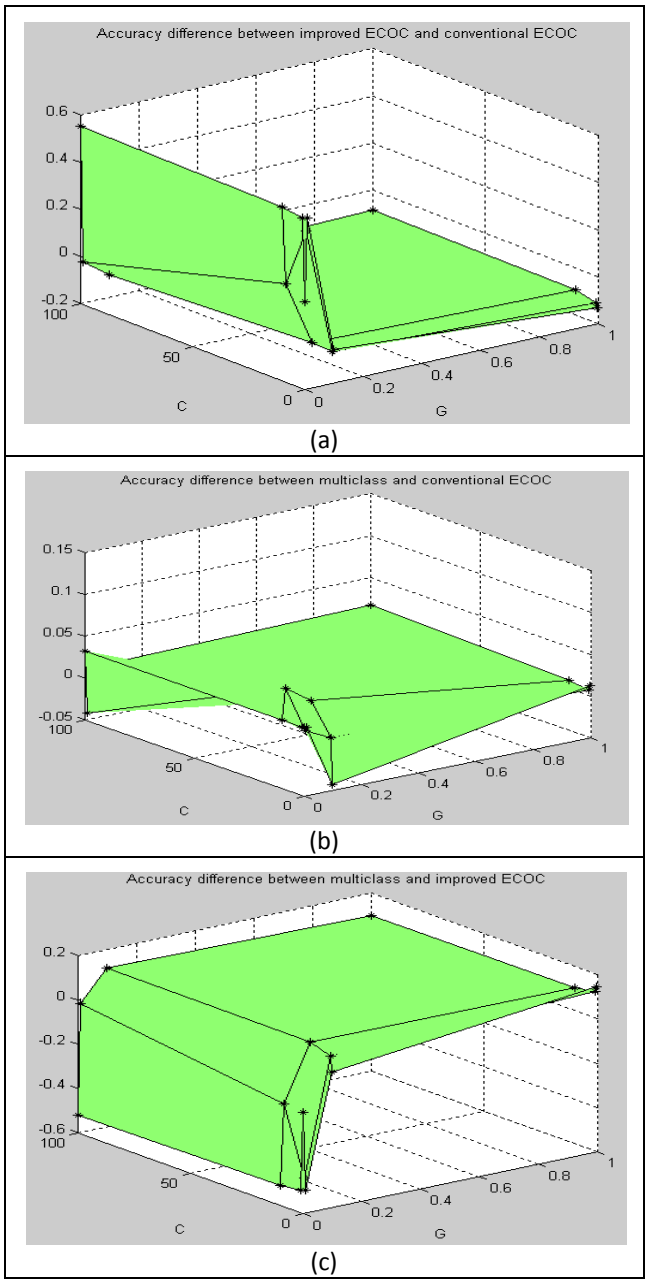


Figure 7. 13 Accuracy comparisions between two methods (Satimage)

Again, in order to see more clearly about the result, table 7.12, table 7.13 and table 7.14 show the accuracy value for each combination of *C* and *G* (row: *C*, column: *G*). These tables show that the improved ECOC works.

Table 7. 12 accuracy of improved ECOC

0.7495	0.7390	0.7255	0.3710
0.7710	0.7495	0.7390	0.7255
0.7745	0.7710	0.7495	0.7390
0.7750	0.7750	0.7710	0.7495

Table 7. 13 accuracy of conventional ECOC

0.8810	0.7955	0.1985	0.1985
0.8870	0.8165	0.2685	0.1985
0.8690	0.8350	0.5370	0.1985
0.8570	0.8810	0.7955	0.1985

Table 7. 14 accuracy of multiclass classifier

0.8930	0.7530	0.2310	0.2310
0.8930	0.8290	0.2960	0.2310
0.8790	0.8840	0.6070	0.2310
0.8740	0.8930	0.7530	0.2310

7.3.3.2 Comparision based on time consumption

According to the experiments without figures supplied, compared with conventional ECOC and direct multiclass classifier, the time consumption in the proposed ECOC is extremely large, which is a obvious disadvantage.

7.4 Results for Hemorrhage Prediction

7.4.1 Separate results for ABP and ECG

To verify the effect of ABP and ECG towards hemorrhage, the results are given separately. For ECG signal, the dataset include 91 subjects all together. Each subject contains 49 features. For ABP signal, the dataset include 91 subjects all together. Each subject contains 50 features. Two methods, ten fold cross validation and leave one out cross validation, are used to get the average accuracy. The classifier employed is libsvm, which is excuted through Matlab and weka interface. They use radial basis kernel function with *C* changing from 0.01 to 10000, gamma *G* changing from 0.00001 to 10.

Before implementing ten fold cross validation, all instances, which are the individuals of each stage are shuffled. Table 7.15 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) for ECG signal. The maximum accuracy is 94.13% as shown.

Table 7. 15 accuracy of ECG using ten fold cross validation

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.4461	0.6656	0.6243	0.4470	0.4461	0.4461	0.4461
0.1	0.8406	0.7105	0.6674	0.6290	0.4477	0.4461	0.4461
1	0.9277	0.7875	0.6872	0.6674	0.6288	0.4477	0.4461
10	0.9413	0.8910	0.7259	0.6752	0.6673	0.6288	0.4477
100	0.9407	0.9284	0.7775	0.6991	0.6744	0.6672	0.6289
1000	0.9405	0.9360	0.8504	0.7332	0.6865	0.6743	0.6672
10000	0.9405	0.9301	0.9005	0.7676	0.7098	0.6858	0.6738

Before implementing leave one subject out cross validation, all instances, which are the patients, are shuffled. Each time, ten percentage of patients are chosen to be testing dataset, and the rest are picked as training dataset. Table 7.16 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) with standardization for ECG signal. The maximum accuracy is 84.75% as shown.

Table 7. 16 accuracy of ECG using leave one subject out cross validation

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.7941	0.8219	0.7747	0.4454	0.4452	0.4452	0.4452
0.1	0.8168	0.8371	0.8212	0.7756	0.4455	0.4452	0.4452
1	0.8378	0.8432	0.8369	0.8214	0.7755	0.4455	0.4452
10	0.8240	0.8434	0.8444	0.8351	0.8213	0.7756	0.4455
100	0.8175	0.8325	0.8459	0.8446	0.8349	0.8211	0.7755
1000	0.8174	0.8174	0.8400	0.8457	0.8426	0.8349	0.8210
10000	0.8174	0.7946	0.8290	0.8475	0.8458	0.8420	0.8349

The same operations are taken on ABP signal. Table 7.17 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) using ten fold cross validation. The maximum accuracy is 91.74% as shown. Table 7.18 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) using leave one subject out cross validation with standardization. The maximum accuracy is 86.73% as shown.

Table 7. 17 accuracy of ABP using ten fold cross validation

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.4604	0.6994	0.6421	0.4461	0.4461	0.4461	0.4461
0.1	0.7654	0.7379	0.7055	0.6496	0.4461	0.4461	0.4461
1	0.9033	0.8038	0.7319	0.7057	0.6505	0.4461	0.4461
10	0.9174	0.8751	0.7619	0.7292	0.7057	0.6501	0.4461
100	0.9174	0.9124	0.8069	0.7501	0.7293	0.7059	0.6504
1000	0.9174	0.9121	0.8524	0.7718	0.7463	0.7293	0.7058
10000	0.9174	0.9089	0.8869	0.8045	0.7594	0.7447	0.7293

Table 7. 18 accuracy of ABP using leave one subject out cross validation

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.6705	0.8225	0.6788	0.4452	0.4452	0.4452	0.4452
0.1	0.8231	0.8474	0.8232	0.6889	0.4452	0.4452	0.4452
1	0.8462	0.8633	0.8485	0.8233	0.6903	0.4452	0.4452
10	0.8334	0.8652	0.8641	0.8469	0.8235	0.6903	0.4452
100	0.8337	0.8459	0.8673	0.8621	0.8470	0.8235	0.6904
1000	0.8337	0.8263	0.8626	0.8663	0.8599	0.8471	0.8233
10000	0.8337	0.8241	0.8475	0.8658	0.8660	0.8602	0.8472

7.4.2 Results for whole dataset

Through the steps, preprocessing and feature extraction, the hemorrhage dataset is formed, which will be used to predict the severity of hemorrhage. The dataset include 91 subjects all together. Each subject contains 99 features from ECG signal and ABP signal. To verify the result of the extracted dataset. Two methods, ten fold cross validation and leave one out cross validation, are used to get the average accuracy. The classifier employed is libsvm, which is excuted through Matlab and weka interface. They use radial basis kernel function with cost *C* changing from 0.01 to 10000, gamma *G* changing from 0.00001 to 10.

Before implementing ten fold cross validation, all instances, which are the individuals of each stage are shuffled. Table 7.19 is the accuracy of each combination of cost *C* (row) and gamma *G* (column). The maximum accuracy is 96.99% as shown.

Table 7. 19. Results on ten fold cross validation

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.4461	0.7018	0.6744	0.5117	0.4461	0.4461	0.4461
0.1	0.4685	0.7976	0.7292	0.6812	0.5245	0.4461	0.4461
1	0.9470	0.9365	0.7722	0.7279	0.6818	0.5255	0.4461
10	0.9494	0.9693	0.8619	0.7504	0.7282	0.6819	0.5257
100	0.9494	0.9695	0.9328	0.7973	0.7456	0.7282	0.6819
1000	0.9494	0.9699	0.9574	0.8641	0.7740	0.7450	0.7280
10000	0.9494	0.9699	0.9590	0.9195	0.8108	0.7707	0.7450

Before implementing leave one out cross validation, all instances, which are the patients, are shuffled. Each time, ten percentage of patients are chosen to be testing dataset, and the rest are picked as training dataset.

Table 7. 20. Results on leave one out cross validation without standardization

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.4452	0.6773	0.6661	0.5117	0.4452	0.4452	0.4452
0.1	0.4452	0.7052	0.6998	0.6730	0.5252	0.4452	0.4452
1	0.5829	0.6863	0.6990	0.7022	0.6737	0.5264	0.4452
10	0.5827	0.6743	0.6916	0.6922	0.7022	0.6739	0.5266
100	0.5827	0.6786	0.6701	0.6964	0.6894	0.7022	0.6738
1000	0.5827	0.6790	0.6706	0.6925	0.6907	0.6891	0.7022
10000	0.5827	0.6790	0.6787	0.6698	0.6930	0.6897	0.6897

Table 7. 21. Results on leave one out cross validation with standardization

	10	1	0.1	0.01	0.001	0.0001	0.00001
0.01	0.4452	0.8330	0.8000	0.4905	0.4452	0.4452	0.4452
0.1	0.7265	0.8606	0.8405	0.8011	0.4997	0.4452	0.4452
1	0.8108	0.8726	0.8649	0.8409	0.8015	0.5006	0.4452
10	0.8144	0.8708	0.8735	0.8615	0.8407	0.8015	0.5008
100	0.8144	0.8553	0.8733	0.8699	0.8595	0.8407	0.8015
1000	0.8144	0.8549	0.8685	0.8727	0.8694	0.8591	0.8407
10000	0.8144	0.8549	0.8510	0.8744	0.8694	0.8700	0.8589

Table 7.20 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) without standardization. The maximum accuracy is 70.52% as shown. Table 7.21 is the accuracy of each combination of cost *C* (row) and gamma *G* (column) with standardization. The maximum accuracy is 87.44% as shown.

7.5 Computational time for hemorrhage prediction

This section is to evaluate the computational time of the system. The configurations of the computer are as follows: four cores computer with 2.0 GHz for each of them. The computational time of the system contains two parts. One is dataset collecting time. The other one is prediction time (the classifier is connected in weka through Matlab). In the research, totally, 91 subjects are used, and 9443 instances are collected. Total running time for these instances is 2.1327e+04 seconds. Therefore, for each instance, the feature collecting time is 2.2585 seconds. For prediction time, the number of the combination of *C* and *G* is 49, and the cross validation is ten fold cross validation. Total running time for training and predicting is 6.1915e+0.4 seconds. Therefore, for each trainning dataset and testing dataset, the prediction time is 126.3578 seconds. This number is closely related to the connection between Matlab and Weka, and many factors. It means that this number can be improved to a large extent.

7.6 Summary

In this Chapter, the results are presented in subpart of the proposed computer aided decision making system. First, the performance of adaptive notch filter is given. Then, two detection algorithms, QRS complex detection algorithm and Systole&Diastole detection

algorithm, are tested using real datasets (MIT dataset and LBNP dataset). In the following part, it is given the trend for using Lempel-ziv to predict the stages of hemorrhage shock. Then, the results from parts of the improved ECOC are presented, which are used to combine in the framework of the proposed improved ECOC. Finally, the results of the whole system, the results about the stage estimation of hemorrhage, are given.

{CHAPTER 8 Conclusions and Future Work}

8.1 Conclusion

Trauma is a serious problem threatening human life in current society. Hemorrhage, in particular, is an important factor encountered in many traumatic injuries and need to be addressed using effective diagnostic methods. In this research, a computer-aided decision making system is proposed for this purpose. Specifically, the proposed computer-aided decision making system help physicians decide the presence and severity of hemorrhage. The accurate and timely information supplied by the system can greatly promote the efficiency of diagnosis and treatment.

In the proposed research, the following tasks are accomplished.

1) In the preprocessing step, an adaptive notch filter is analyzed and deduced. Compared with existing notch filter, the effect of the adaptive notch filter can be adapted just by adjusting the adaptation constant. The implementation of the adaptive notch filter can improve the overall performance of the adaptive filter. Also, blind source separation is employed to remove baseline wander.

2) In the preprocessing step, a novel QRS complex detection algorithm is proposed. The design of QRS complex detection algorithm is an effective way to detection peaks in ECG signal and can greatly improve the accuracy of any computer aided decision making based on ECG.

3) In the preprocessing step, another novel Systole&Diastole detection algorithm is proposed. The novelty of the Systole&Diastole detection algorithm lies in effectively decomposing the ABP signal into components, and detecting Systolic and Diastolic peaks using the relationship among these components. The experimental results reveal the high accuracy of the proposed Systolic and Diastolic peaks.

4) In the feature extraction, a new version of Lempel-ziv is designed and implemented. Although the original idea of Lempel-ziv may not provide a very effective method of detecting hemorrhage, an improved version of Lempel-ziv is desired to achieve this aim. Also, based on ECG and ABP, a couple of features are designed, and verified to effectively represent hemorrhage.

5) In the decision making section, a general framework for the improved version of Error Correcting Output Codes is proposed. Some parts of this framework are finished and tested. Error Correcting Output Codes is a flexible model for multiclass learning problems.

8.2 Future Work

Many tasks have been accomplished in the proposed research. However, there are still some future works that need to conduct:

1) In the preprocessing step, a new method to remove baseline wander is tried. Experimental results verified the correctness of the direction. However, many factors are needed to explore to improve the effectivenss of the method. For example, construct multisignals to better represent the statistical informatioin; choose more effective G

function; locate the individual component, baseline wander, directly to save time consumption, etc.

2) In the preprocessing step, while the experimental results show the effectiveness of the proposed QRS detection algorithm and Systole&Diastole detection algorithm, as discussed above, if it is desired to achieve even better results. As such, some parts of the QRS detection algorithm and Systole&Diastole detection algorithms need further improvements.

3) In the feature extraction step, multi model analysis will be explored. By looking through related research, there is no such a research to explore and detect the characteristic of hemorrhage. The success of multi model analysis for hemorrhage can greatly help the overall objectives of this research.

4) In decision making step, the third layer in the framework of the proposed extension of ECOC will be designed. At the same time, many other factors may be added to improve the current version of the Error Correcting Output Codes.

8.3 Summary

In this chapter, first a brief conclusion of the finished research is provided. Then, the main future work to be conducted as the continuation of the project is outlined.

Literature Cited

[1] Andriy I. Batchinsky, William H. Cooke, Tom A. Kuusela, Bryan S. Jordan, Jing Jing Wang, Leopoldo C. Cancio, "Sympathetic nerve activity and heart rate variability during severe hemorrhagic shock in sheep", Autonomic Neuroscience: Basic & Clinical, 2007, pp. 43-51

[2] S.Y. Ji, K. Ward, K. Ryan, C. Rickards, V. Convertino, N. Vyas, J. Stivoric, and K. Najarian, "Prediction of Hypovolemia Severity Using ECG Signal with Wavelet Transformation Analysis From a Mobile Armband", Circulation,120:S1441,2009, Best Trauma Abstract Award.

[3] S.Y. Ji, A.A.R. Bsoul, K. Ward, K. Ryan, C. Rickards, V. Convertino, and K. Najarian, "Incorporating Physiological Signals to Blood Loss Prediction Based on Discrete Wavelet Transformation", Circulation, 2009.

[4] James Philips. "Introducing Health Sciences: Trauma, Repair and Recovery". Oxford University Press. ISBN13:978-0-19-923734-0/ISBN10:0-19-923734-4.

[5] Eric A. Finkelstein, Phaedra S. Corso and Ted R. Miller. "Incidence and Economic Burden of Injuries in the United States". Oxford University Press. ISBN13:978-0-19-517948-4/ISBN10:0-19-517948-X.

[6] Robert D. Barraco, William C. Chiu, Michael R. Bard, Faran Bokhari, Anthony Borzotta, Osbert Blow, Jeannette Capella, Marie Dieter, Thomas Z. Hayward III, William S. Hoff, Col. John C. Holcomb, John F. McCarthy, Michael Moncure, Donna A. Nayduch. "Practice Management Guidelines for the Appropriate Triage of the Victim of Trauma".

[7] Nasim Karim, Jahan Ara Hasan, Syed Sanowar Ali. "Heart Rate Variability--A Review". Journal of Basic and Applied Sciences. Vol.7, No.1, 71-77, 2011

[8] Marek Malik, "Heart rate variability-Standards of measurement, physiological interpretation and clinical use". European Heart Journal, 1996, pp.354-381

[9] William H. Cooke, Jose Salinas, Victor A. Convertino, David A. Ludwig, Denise Hinds, James H. Duke, Fredrick A. Moore, John B. Holcomb. "Heart Rate Variability and Its Association with Mortality in Prehospital Trauma Patients". The journal of Trauma injury, infection, and critical care. 2005, pp.363-370

[10] Nan Liu, Zhiping Lin, Zhixiong Koh, Guang-Bin Huang, Wee Ser, Marcus Eng Hock Ong, "Patient Outcome Prediction with Heart Rate Variability and Vital Signs", J Sign Process Syst, 2011, pp.265-278

[11] W. H. Cooke, J. Salinas, J. G. McManus et al., "Heart period variability in trauma patients may predict mortality and allow remote triage", Aviation Space and Environmental Medicine, vol. 77, no.11, pp. 1107-1112, 2006

[12] A. I. Batchinsky, J. Salinas, T. Kuusela, C. Necsoiu, J. Jones, and L. C. Cancio, "Rapid prediction of trauma patient survival by analysis of heart rate complexity: impact of reducing data set size," Shock, vol. 32, no.6, pp.565-571, 2009

[13] Mark L. Ryan, Chad M. Thorson, Christian A. Otero, Thai Vu, Kenneth G. Proctor, "Clinical Applications of Heart Rate Variability in the Triage and Assessment of Traumatically Injured Patients", Anesthesiology Research and Practice, Volume 2011, pp.1-8

[14] William H. Cooke, Kathy L. Ryan, and Victor A. Convertino, "Lower body negative pressure as a model to study progression to acute hemorrhagic shock in humans", Journal of Applied Physiology, 2004, pp.1249-1261

[15] Gray M. Friesen, Thomas C.Jannett, Manal Afify Jadallah, Stanford L.Yates, Stephen R.Quint, H.Troy Nagle, “A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms”. IEEE Transactions on Biomedical Engineering, Vol.37, No.1, 1990, pp.85-98

[16] Shaowei Feng, Jianjun Ding, “Introduction of Wiener Filter, Advance Digital Signal Process Final Tutorial.”, from the notes of the course at National Taiwan University.

[17] Tony Lacey. “The Kalman Filter, Tutorial”, from the notes of the CS7322 course at Georgia Tech-Notes taken from the TINA Algorithms’ Guide by N.Thacker, Electronic Systems Group, University of Sheffield.

[18] Bernard Widrow, Michael Lehr, Francoise Beaufays, Eric Wan, and Michel Bilello. “Adaptive Signal Processing”, from the papers of Widrow at Standford University.

[19] John R. Glover, JR. “Adaptive Noise Cancelling Applied to Sinusoidal Interferences”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-25, No.6, December 1977, pp. 484-491

[20] Raimon Jane, Pablo Laguna, Nitish V. Thakor and Pere Caminal, “Adaptive Baseline Wander Removal in the ECG: Comparative Analysis With Cubic Spline Technique”, Computers in Cardiology 1992. Proceedings, pp.143-146

[21] Pablo Laguna, Raimon Jane, Olivier Meste, Peter W. Poon, Pere Caminal, Herve Rix, Nitish V.Thakor, “Adaptive filter for Event-Related Bioelectric Signals Using an Impulse Correlated Reference Input: Comparison with Signal Averaging Techniques”, IEEE Transactions on Biomedical Engineeing, Vol. 39, No.10, October,1992, pp.1032-1044

[22] J.Fraden and M.R.Neumann, “QRS wave detection,” Med.Biol.Eng.Compt., vol. 18, pp.125-132, 1980

[23] M.Bahoura, M.Hassani, M.Hubin, “DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis”, Computer Methods and Programs in Biomedicine, 1997, pp.35-44

[24] R.K.Sunkaria, S.C.Saxena, V.Kumart, A.M.Singhai. “Wavelet based R-peak detection for heart rate variability studies”, Journal of Medical Engineering & Technology, vol.34, 2010, pp.108-115

[25] JIAPU PAN, WILLIS J.TOMPKINS. “A Real-Time QRS Detection Algorithm”, IEEE Transactions on Biomedical Engineering, 1985, pp.230-236

[26] PATRICK S. HAMILTON, WILLIS J.TOMPKINS. “Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database”. IEEE Transactions on Biomedical Engineering, 1986, pp.1157-1165

[27] Ivaylo I Christov. “Real time electrocardiogram QRS detection using combined adaptive threshold”. BioMedical Engineering OnLine, 2004

[28] Yun-Chi Yeh, Wen-June Wang. “QRS complexes detection for ECG signal: The Difference Operation Method”. Computer Methods and Program in Biomedicine, 2008, pp.245-254

[29] G.Vijaya, Vinod Kumar, H.K.Verma. “ANN-based QRS-complex analysis of ECG”. Journal of Medical Engineering & Technology, 1998, pp.160-167

[30] Laurence Keselbrener, Michel Keselbrener, Solange Akselrod. “Nonlinear high pass filter for R-wave detection in ECG signal”. Medical Engineering & Physics, 1997, pp.481-484

[31] Valtino X.Afonso, Willis J.Tompkins, Truong Q.Nguyen, Shen Luo. “ECG Beat Detection Using Filter Banks”. IEEE Transactions on Biomedical Engineering, 1999, pp.192-202

[32] Khandaker Abul Kalam Azad. “Fetal QRS complex detection from abdominal ECG: A fuzzy approach”. IEEE Nordic Signal Processing Symposium, 2000

[33] D.Benitez, P.A.Gaydecki, A.Zaidi, A.P.Fitzpatrick. “The use of the Hilbert transform in ECG signal analysis”. Computers in Biology and Medicine, 2001, pp.399-406

[34] Panagiotis trahanias, Emmanuel skordalakis. “Syntactic Pattern Recognition of the ECG”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, pp. 648-657

[35] P.E.Trahanias. “An Approach to QRS Complex Detection Using Mathematical Morphology”. IEEE Transactions on Biomedical Engineering, 1993, pp. 201-205

[36] Yan Sun, Kap Luk Chan, Shankar Muthu Krishnan. “ECG signal conditioning by morphological filtering”. Computers in Biology and Medicine, 2002, pp. 465-479

[37] Yan Sun, Kap Luk Chan, Shankar Muthu Krishnan. “Characteristic wave detection in ECG signal using morphological transform”. BMC Cardiovascular Disorders, 2005

[38] Remi Dubois, Pierre Maison-Blanche, Brigitte Quenet, Gerard Dreyfus. “Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators”. Computer Methods and Programs in Biomedicine, 2007, pp.217-233

[39] Kleydis V.Suarez, Jesus C.Silva, Yannick Berthoumieu, Pedro Gomis, Mohamed Najim. “ECG Beat Detection Using a Geometrical Matching Approach”. IEEE Transactions on Biomedical Engineering, 2007, pp.641-650

[40] Thorsten Last, Chris D Nugent, Frank J Owens. “Multi-component based cross correlation beat detection in electrocardiogram analysis”. BioMedical Engineering OnLine, 2004

[41] Steven E. Dobbs, Neil M.Schmitt, Haluk S. Ozemek. “QRS Detection By Template Matching Using Real-Time Correlation On A Microcomputer”. Journal of clinical engineering, 1984, pp.197-212

[42] Aapo Hyvarinen, 'New approximations of differential entropy for independent component analysis and projection pursuit', In Advances in Neural Information Processing Systems, volume 10, pp.273-279, 1998

[43] Aapo Hyvarinen, 'Fast and Robust Fixed-Point Algorithms for Independent Component Analysis', IEEE Transactions on Neural Networks, vol.10, pp.626-634, 1999

[44] Athanasios Papoulis, S. Unnikeishna Pillai, “Probability, Random Variables, and Stochastic Processes”, McGraw-Hill Series in Electrical and Computer Engineering, fourth edition, pp.408-409

[45] Athanasios Papoulis. “Probalility, Random Variables, and Stochastic Processes (3rd)”. pp. 327

[46] Masahiko Okada. “A digital filter for the QRS complex detection”, IEEE Trans. Biomed. Eng, vol.26,1979, pp.700-703

[47] D.Gustafson, “Automated VCG interpretation studies using signal analysis techniques”, R-1044 Charles Stark Draper Lab, Cambridge

[48] P.Morizet-Mahoudeaux, C.Moreau, D.Moreau, J.J.Quarante, “Simple microprocessor-based system for on-line ECG arrhythmia analysis”, Med.Biol.Eng.Comput, vol.19, no.4, pp.497-501

[49] Bryan S. Morse, Brigham Young University, “Thresholding”, Leture note 4 in school of informatics, Brigham Young University, 1998-2000

[50] Abraham Lempel, Jacob Ziv, “On the Complexity of Finite Sequences”, IEEE transactions on information theory, 1976, pp.75-81

[51] Y.L.Orlov, V.N.Potapov, “Complexity: an internet resource for analysis of DNA sequence complexity, Nucleic Acids Research, 2004, pp.W628-W633

[52] Xu-Sheng Zhang, Rob J.Roy, Erik Weber Jensen, “EEG Complexity as a Measure of Depth of Anesthesia for Patients”, IEEE transactions on biomedical engineering, Vol.48, 2001, pp.1424-pp.1433

[53] Vladimir D.Gusev, Lubov A.Nemytikova, Nadia A. Chuzhanova, “On the complexity measures of genetic sequences”, Bioinformatics, 1999, pp.994-999

[54] Daniel Abasolo, Roberto Hornero, Carlos Gomez, Maria Garcia, Miguel Lopez, “Analysis of EEG background activity in Alzheimer’s disease patients with Lempel-ziv complexity and central tendency measure”, Journal of Medical engineering & Physics, 2006, pp.315-322

[55] Mateo Aboy, Roberto Hornero, Daniel Abasolo, Daniel Alvarez, “Interpretation of the Lempel-ziv Complexity Measure in the Context of Biomedical Signal Analysis”, IEEE Transaction on Biomedical Engineering, vol.53, No.11, 2006, pp.2282-2288

[56] Nagarajan R. “Quantifying physiological data with Lempel-ziv complexity- Certain issues”. IEEE Transaction on Biomedical Engineering, 2002, pp.1371-1373

[57] Jean-Philipe Thiran, Ferran Marques, Herve Bourlard, “Multimodal Signal Processing”, ISBN: 978-0-12-374825-6

[58] Terrence J. Sejnowski, Charles R. Rosenberg, “Parallel Networks that Learn to Pronounce English Text”, Journal of Complex Systems, 1, 1987, pp.145-168

[59] Thomas G.Dietterich, Ghulum Bakiri, “Solving multiclass learning problems via error-correcting output codes”, Journal of Artificial Intelligence Research 2, 1995, pp.263-286

[60] Eun Bae Kong, Thomas G.Dietterich, “Error Correcting Output Coding Corrects Bias and Variance”, pp.1-9

[61] Eun Bae Kong, Thomas G.Dietterich, “Probability Estimation via Error Correcting Output Codeing”, IASTED International Conference: Artificial Intelligence and Soft Computing, Banff, Canada, pp.1-6

[62] Divid W. Aha, Richard L. Bankert, “Cloud Classification Using Error-Correcting Output Codes”, Natural Resources, Agriculture, and Environmental Science, 1996, pp.1-24

[63] Satomi Suzuki, Koji Oguri, "Cuffless Blood Pressure Estimation by Error-Correcting Output Coding Method Based on an Aggregation of AdaBoost with a Photoplethysmograph Sensor", 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, 2009, pp.6765-6768

[64] Yurong Luo, Kayvan Najarian, "Employing decoding of specific error correcting codes as a new classification criterion in multiclass learning problems", International Conference of Pattern Recognition, 2010, pp.4238-3241

[65] Thomas G.Dietterich, Ghulum Bakiri, "Error-Correcting Output Codes: A General Method for Improving Multiclass Inductive Learning Programs", pp.1-6

[66] W. Utschick, W. Weichselberger, "Stochastic Organization of Output Codes in Multiclass Learning Problems", Neural Computation, vol.13, no.5, pp.1065-1102, 2004

[67] K. Crammer, Y. Singer, "On the Learnability and Design of Output Codes for Multi-Class Problems", Machine Learning, vol.47, pp.201-233, 2002

[68] O. Pujol. P. Radeva, and J. Vitria, "Discriminant ECOC: A Heuristic Method for Application Dependent Design of Error Correcting Output Codes", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.28, no.6, pp.1001-1007, June 2006

[69] P. Pudil, F.J.Ferri, J.Novovicova, J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions", Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol.2, 1994, pp.279-283

[70] Sergio Escalera, Oriol Pujol, Petia Radeva, "Boosted Landmarks of Contextual Descriptors and Forest-ECOC: A novel framework to detect and classify objects in cluttered scenes", Journal of Pattern Recognition Letters, vol.28, 2007, pp.1759-1768

[71] Oriol Pujol, Sergio Escalera, Petia Radeva, "An incremental node embedding technique for error correcting output codes", Journal of Pattern Recognition, 2007, pp.713-725

[72] Sergio Escalera, David M.J.Tax, Oriol Pujol, Petia Radeva, Robert P.W.Duin, "Subclass Problem-Dependent Design for Error-Correcting Output Codes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, No.6, 2008, pp.1041-1054

[73] Nikolaos Arvanitopoulos, Dimitrios Bouzas, Anastasios Tefas, “Subclass Error Correcting Output Codes using Fisher’s Linear Discriminant Ratio”, International Conference on Pattern Recognition, 2010, pp.2953-2956

[74] E. Allwein, R. Schapire, Y. Singer, “Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers”, J. Machine Learning Research, vol.1, 2002, pp.113-141

[75] T. Windeatt, R. Ghaderi, “Coding and Decoding for Multi-Class Learning Problems”, Information Fusion, vol.4, 2003, pp.11-21

[76] Koby Crammer, Yoram Singer, “Improved Output Coding for Classification Using Continuous Relaxation”, in Proc. NIPS, 2000, pp.437-443

[77] Sergio Escalera, Oriol Pujol, Petia Radeva, “On the Decoding Process in Ternary Error-Correcting Output Codes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, No.1, 2010, pp.120-134

[78] A. Passerini, M. Pontil, P. Frasconi, “New Results on Error Correcting Output Codes of Kernel Machines”, IEEE Trans. Neural Networks, vol.15, no.1, pp.45-54

[79] Yurong Luo, Abed Al-Raoof Bsoul, Kayvan Najarian, “Confidence Based Classification with Dynamic Conformal Prediction and Its Applications in Biomedicine”, 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011

[80] A. P. Dempster, N. M. Laird, D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM algorithm”, Journal of the Royal Statistical Society, vol.39, 1997, pp.1-38

[81] Stephen B.Wicker, "Error Control Systems for Digital Communication and Storage", ISBN 0-13-200809-2, 1995, pp.68-109

[82] William S. Cleveland, 'Robust Locally Weighted Regression and Smoothing Scatterplots', Journal of the American Statistical Assocaition, vol.74, pp.829-836, 1979.

[83] http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html

[84] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-

e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

[85] A.A.R. Bsoul, S.Y. Ji, K. Ward, and K. Najarian, Automatic Prediction of Arrhythmia Severity Using Time and Frequency Domain Features, Circulation 122:A213, 23 November 2010

VITA

Yurong Luo was born in Hebei, China, on March 08, 1980. She holds Chinese citizenship. She received her Bachelor degree and Master degree from Computer Science department in Hebei Normal University and Beijing Technology of University respectively in 2004 and 2008. From August 2009, she was employed as a Graduate Research Assistant in the Department of Computer Science at Virginia Commonwealth University. Her research fields include biomedical image and signal processing, machine learning, pattern recognition.

List of Relevant Publications:

Abstracts in Journals

[1] Yurong Luo, Soo-Yeon Ji, Kevin Ward, Kayvan Najarian, "A Novel Method for Detection of QRS Complex with an Adaptive Points Insertion", AHA-Circulation, 2010, 122:A273

[2] Soo-Yeon Ji, Kevin Ward, Abed Al Raoof Bsoul, **Yurong Luo**, Kathy Ryan, Caroline Rickards, Victor Convertine, "Developing a Hypovolemia Monitoring System by Integrating Physiological Measures and P-QRS-T Waves", AHA-Circulation, 2010, 122:A272

Conference papers

[3] Yurong Luo, Kayvan Najarian, “Employing decoding of specific error correcting codes as a new classification criterion in multiclass learning problems”, 20th International Conference on Pattern Recognition, 2010

[4] Yurong Luo, Abed Al-Raoof Bsoul, Kayvan Najarian, “Confidence-Based Classification with Dynamic Conformal Prediction and Its Applications in Biomedicine”, submitted to: 33 rd International Conference of the IEEE Engineering in Medicine and Biology Society, 2011

Pending Patent/Invention disclosure:
Yurong Luo, Soo-Yeon Ji, Kevin Ward, Kayvan Najarian, "A Novel Method for Detection of QRS Complex with an Adaptive Points Insertion"