Remote Heart Monitoring: A Predictive Modeling Approach for Biomedical Signal Processing

By Jiaming Chen

A Thesis Submitted in Partial Fulfillment

of the Requirements for Degree of

Master of Science

in Electrical Engineering

Northern Arizona University

August 2018

Approved:

Abolfazl Razi, Ph.D, Chair

Fatemeh Afghah, Ph.D

Bertrand Cambou, Ph.D

Remote Heart Monitoring: A Predictive Modeling Approach for Biomedical Signal Processing

Jiaming Chen

(ABSTRACT)

Smart healthcare is an emerging field with numerous research projects devoted to design electronic devices, ~~computer technologies~~ and platforms aiming at facilitate technology-based health service at lower costs. Biomedical signal can directly reflect the information regarding patient health and has therefore been ~~frequently investigated~~. The essence of ~~biomedical~~ signal analysis system is to ~~process~~ signals and build ~~statistical or machine learning~~ model to ~~provide informative result about~~ the health status of patients. While the majority of methods ~~in literature~~ focus on improving classification performance ~~on pooled dataset~~, the predictive modeling of biomedical signal is rarely emphasized. In this work, we go one step beyond the conventional methods and intend to predict potential upcoming abnormalities before their occurrence. The objective is to build a patient-specific model and identify minor deviations from the normal signal, which can be indicative of potential upcoming significant deviations.

To enable an accurate ~~prediction on different data, two spatial transformation methods in which feature space are reshaped according to designed topology are proposed together with a patient-adaptable classification framework.~~ We applied the developed algorithms on Electrocardiogram (ECG) signals and the results confirm the effectiveness of the proposed method in predicting upcoming heart abnormalities before their occurrence. For instance, the probability of observing a specific abnormality class increases by 10% after triggering a yellow alarm of the same type. This approach is general and has the potential to be applied to a wide range of physiological signals.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Heart-related mortality rate has been increasing dramatically due to the aging of population, chronic cardiovascular diseases and increasing life stress and pace of modern life [1]. According [2], cardiac diseases are the most common cause of sudden cardiac death (SCD) with 250 000 to 300 000 mortalities in the U.S. every year accounting for 14.7% of total deaths [2]. As World Health Organization reported, 31% of global deaths are related to cardiovascular diseases (CVDs) [3]. These facts fully reflect that heart diseases are threatening the general health of human beings. Since death from CVD can occur in most cases without prior warning and obvious symptoms, it is of great importance to enable a timely treatment of heart diseases. For this purpose, prevention principles and guidelines which covers age, family history and other potential risk factors causing CVD are deployed in most clinical modeling method [4]. However, these methods require complex manual analysis by trained physicians. Taking this issue into consideration, a cost-effective automatic analysis for CVD prevention based on computer is in demand. More specifically, since most CVDs

are accompanied with arrhythmia, accurate and timely recolonization of arrhythmia is a key factor for effective prevention of heart diseases.

Electrocardiogram (ECG) is the most common way of monitoring hearth functionality, which contains abundant physiological and pathological information that reflects the heart rhythm and status of various parts of the heart. ECG signal are recorded for the first time by Waller in 1887 [5]. It records signals generated by electrical activities of heart as a time series. As a noninvasive examination method, it is known to be highly reliable in reflecting functionality of heart. For this reason, ECG has become one of the most conventional technologies (ECG, clinical examination, radiation and ultrasonic inspection) in modern hospitals and clinics, serving as an important reference for doctors' diagnosis of heart diseases [6].

The traditional diagnosis based on ECG analysis are mainly performed by physicians through visual observation and interpretation.However, the approach costly and impractical when continuous monitoring of patients is required (e.g. to recognize CVD conditions).There are tremendous ECG records generated everyday, all demand for timely diagnosis and analysis. Due to the limitations in the access to experience physicians, automated ECG classification system has been introduced and became popular soon afterwards to generate real-time analysis result and provide additional information to physicians.

Several computer-based automated classification algorithms has been developed by researchers in the last decades to minimize human intervention or to assist physicians with more accurate diagnosis by reducing human mistakes [7–17]. Moreover, with the emerging application of smart health and smart ~~homes/~~cities, a constant monitoring and analysis of ECG and other physiological signals with direct experts' intervention deems impossible. Therefore, applying conventional classification algorithms on biomedical signals remains challenging, especially for applications of spontaneous disease detection.

A typical feature of cardiovascular disease is the ~~high complexity~~ of causing factors and implicit symptoms before occurrence [18, 19]. Failing to predict life threatening CVDs is the principal cause of high mortality for patients with heart disease. A timely prediction of hearth abnormalities before their actual occurrences would enable a therapeutic intervention before the condition becomes detrimental, which minimizes the risk of mortality. Nevertheless, the majority of developed conventional ECG classification systems are only able to detect abnormalities when they occur. To the best of our knowledge, no research work is devoted to the prediction of heart abnormalities ahead of time, which is the main focus of this project [20, 21].

Another important property of ECG waveforms is their inherent variability among different individual and due to ~~physical condition and~~ different environmental including but not limited to gender, age, body-mass index etc [22, 23]. Conventional classification algorithms do not easily generalize, when applying to different patients' records [14]. Due to the inter-patient variation in ECG signals and the complexity of cardiac pathological information analysis, most of the existing ECG analysis software only serve as auxiliary devices for physicians. The final results of diagnosis still depend on manual labeling by cardiologists. Recently, several novel patient-specific ECG analysis methods are proposed. Broadly speaking, in these methods systems parameters are adaptable according to individual ECG signal properties. Some algorithms combine cardiologist manual annotations with automatically generated labels and train personal classifier with updated labels for each individual. This design still requires expert assistance in order to accurately classify ECG signal. Another design in to train patient-specific classifier with a personal ECG signal. Whereas this method fails when a certain type of abnormal signal is not included in the limited personal ECG signal data.

The automatic analysis of ECG signals includes a wide range of techniques. In this work,

we focus on overcoming the two drawbacks of existing automatic ECG classification systems, namely the failure in capturing patient-specific variability and the lack of predictive power.This research aims at improving the inter-patient classification performance and prediction capability of ECG-based diagnosis methods. Our proposed method can revolutionize the current practice of healthcare service by enabling early detection of heart abnormalities, with applications to high-risk people, senior people, and athletes. It also can significantly reduce the mortality rate of SCD.

1.2 ECG and Arrhythmia

Electrocardiogram is widely used to monitor the electrical activities of heart and assist diagnosing fatal cardiac diseases. In order to design algorithms specifically for ECG analysis, it is important to develop an insightful perception of the functionality of heart and ECG waveforms.

1.2.1 Characteristics of ECG signal

A ECG signal reflects the periodical electric signals generated by a heart. Fig.1.1 demonstrates the typical signal waveform for a cardiac cycle (i.e. a heartbeat), which is usually composed of three main waves including P wave, QRS complexes and T wave. These waves corresponds to different physiological activities of the heart. P waves are generated by atrial depolarization which represents the process of pumping blood to ventricles. QRS complexes as the most significant electric activities are caused by the *Ventricular* contraction, which is the process of pumping blood to lungs and the rest of the human body. Finally, T waves are the result of *Ventricular* repolarization, which is a required recovery process before the

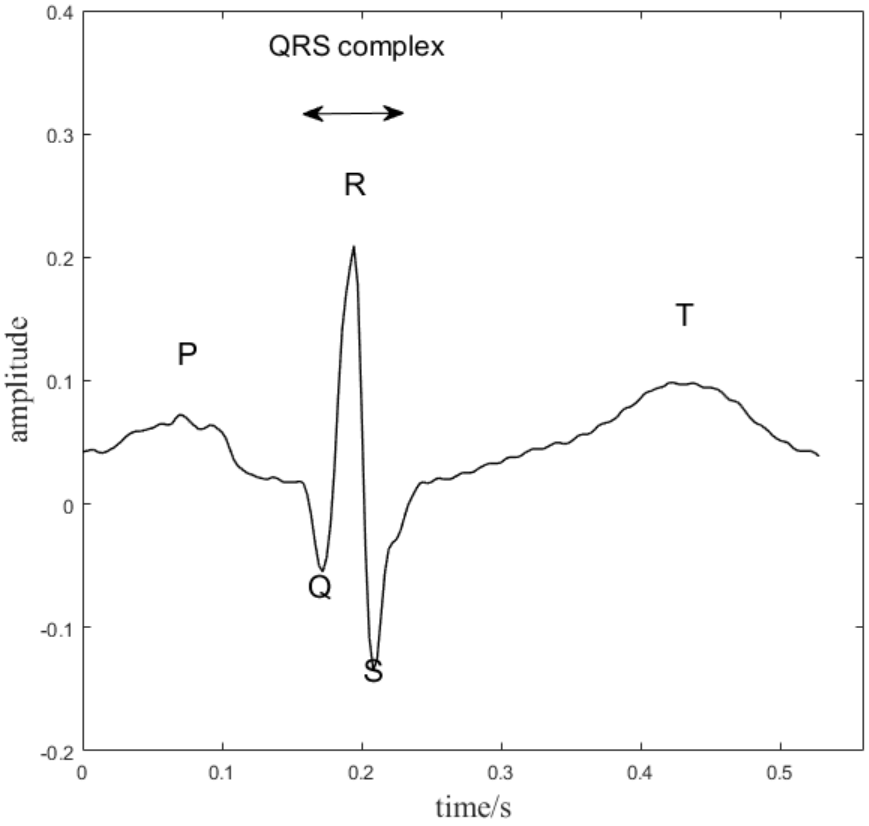

Figure 1.1: A typical cardiac cycle in ECG signal with five characteristic waves

following cardiac cycle. Accurate detection and segmentation of each wave is necessary for a profound ECG analysis. The waves are usually represented by their peak locations, also called fiducial peaks. By detecting the most significant peak within QRS complexes (e.g. R peak) automatic algorithms are able to discriminate between two adjacent cardiac cycles. The interval between two R peaks is called RR interval, which is also the inverse of heart rate. Fig.1.1 represents a typical cardiac cycle with the aforementioned intervals.

1.2.2 MIT-BIH Arrhythmia Database

Arrhythmia is related to various morbid behaviors of heart. Generally speaking, arrhythmias consist of two main categories: *supraventricular* and *ventricular*. *Ventricular* ectopic beats imply abnormal activities in the ventricles while *supraventricular* ectopic beats are related to the *atria* [24]. Both categories contains fatal abnormal beats, which may lead to death [25]. Therefore, in order to help researchers standardize the evaluation of works on ECG classifiers, Association for the Advancement of Medical Instrumentation (AAMI) has proposed recommendations for reporting ECG classifier performance [26]. According to these recommendations, MIT-BIH Arrhythmia Database (MITDB) is regarded as a standard database to train and test ECG classifiers in the last two decades. MITDB is a public database which is available on Physionet.com [27] since 1997 [28]. There are 48 records collected from 47 individuals in this database. Each record contains two channels of ECG raw signals along with annotations for each cardiac cycle. Annotated labels include 16 types, as shown in Table 1.1. Cardiac cycles are determined by the locations of R peaks. The sampling frequency of MITDB is 360Hz and the signal frequency spans from 0.1 to 100 Hz.

Following the recommendations by AAMI, the original annotations of MITDB are further grouped into 5 major classes: class N(normal and bundle branch block beat types) class V(*Ventricular* type), class S(*supraventricular* type) and class F(fusion of normal and *Ventricular* types). The class Q which includes unclassified and paced beats are discarded due to the limited number of samples. Table.1.1 summarizes the mapping from 16 original types that include cycles of this type. Therefore, only 4 remaining types (N, V, S, F) are typically used.

Table 1.1: Mapping from 16 original types in annotation to the standard 5 types recommended by AAMI

Standard Types by AAMI	Original Types in MITDB Annotation
N	NOR, LBBB, RBBB, AE, NE
V	PVC, VE, VF
S	APC, AP, BAP, NP
F	VFN
Q	PACE, FPN, UN

1.3 Problem Statement

ECG signals are investigated broadly by researchers to design automated non-invasive diagnosis methods and real-time monitoring systems [17, 29, 30]. As described in the previous section, a majority of current methods suffer from two main challenges: i) failure to capture inter-patient variability and ii) incapability of early detection and prediction.

In conventional classification systems, the training dataset is typically composed of records collected from different patients with experts' annotations per heartbeat. In order to unify the records from different patients, most of the conventional classification algorithm mix heartbeat samples from different individual ECG records and cluster the pooled ECG dataset simply based on the annotations of heartbeats. Since the classification performance is measured based on the comparison between the predicted labels with the annotate (true) labels for each sample, the classifiers are trained to improve the performance on pooled ECG data. While ECG signals shares similar morphologies, the signals from different patients demonstrate considerable variability as shown in Fig.1.2. Ignoring this difference will lead to inconsistent classification performance between patients. Therefore it's of significant importance to adjust classifier configuration according to patient-specific characteristics.

In addition to the inter-patient variability, majority of ECG classification algorithms fail to provide a predictive capability, which refers to the power of triggering corresponding alarms

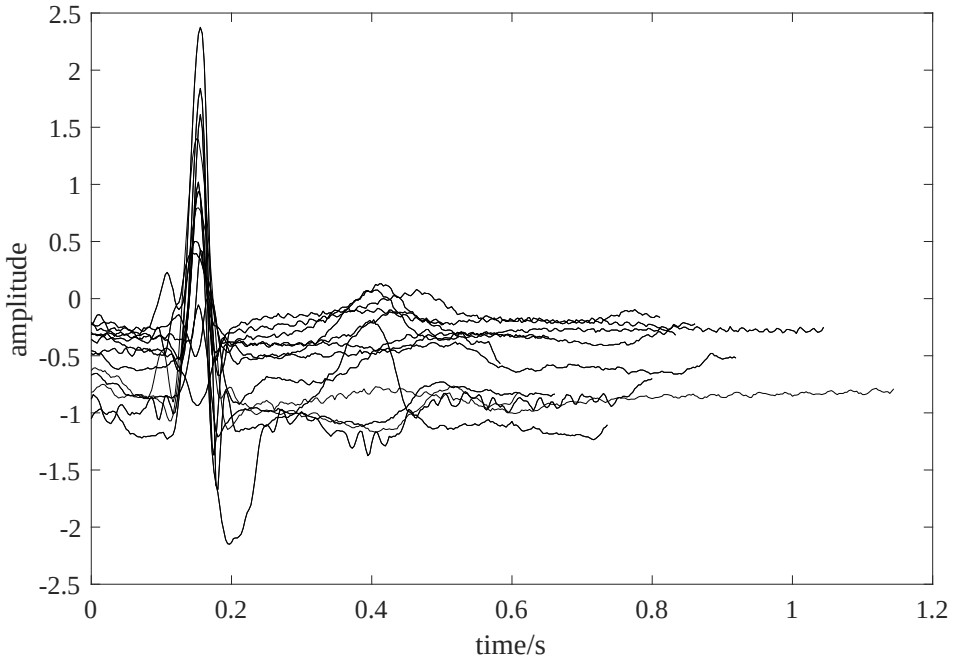

Figure 1.2: ECG signals of normal heartbeat from 15 different records in MIT-BIH reflect the inter-patient variability of ECG signal

before the occurrence of abnormalities. Typically alarms represent significant distortions in the ECG morphology which reflect a severe heart abnormality. The rest of heartbeats considered as *normal* beats. However, an abnormal beat may include mild distortions that can be indicative of a problem, while it is not severe enough to call a *red alarm*. In this work, we represent these minor deviations with *yellow alarm* and use them to predict real abnormalities as *red alarms*, before their actual occurrence.

Fig.1.3 illustrates this concept, where observation of minor deviations from the patient-specific normal trends (yellow alarms) can be indication of upcoming severe abnormalities (red alarms). Supporting results are provided in Section 3.4 and Section 4.6.Therefore, a method to quantify the level of signal similarity to abnormalities should be incorporated into the ECG classification system.

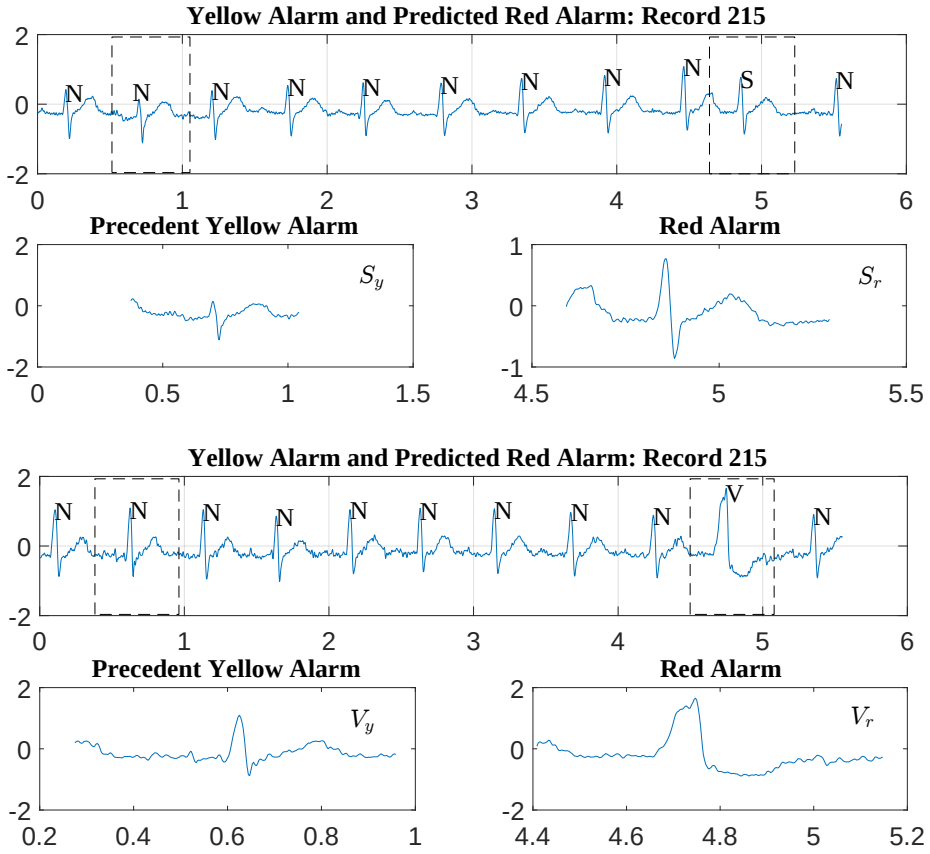

Figure 1.3: A potential latent abnormal status(yellow alarm) predicts an upcoming abnormality of the same type.

1.4 Literature Review

Automatic analysis of ECG signals refers to the entire process spanning from the acquisition of signals to the classification of samples. This process can be divided into five stages: ECG signal acquisition, preprocessing, fiducial peak detection and segmentation, feature extraction and predictive modeling (Fig. 1.4). Different research works focused on one or multiple stages of the automatic analysis system. Since the main objective of this work is addressing problems in classification algorithms, the literature reviews in this section focuses on studying existing methods proposed for stages before classification, conventional classification algorithms along with patient-specific classification systems.

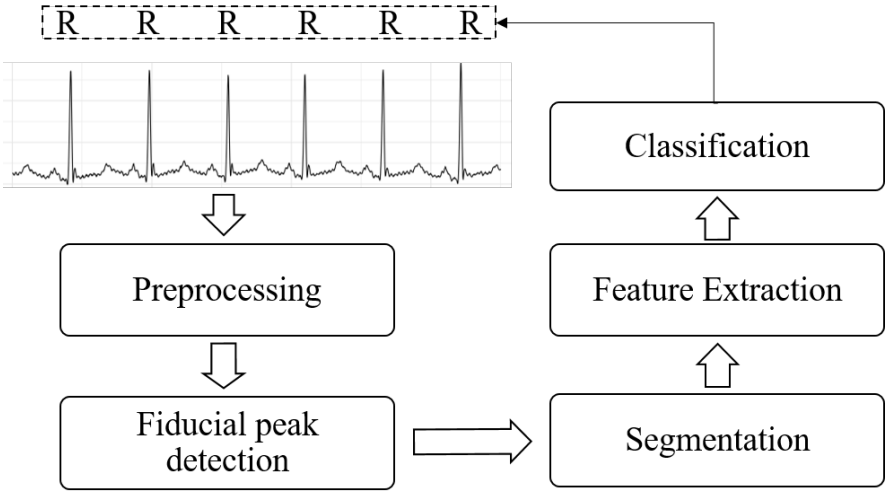

Figure 1.4: General structure of ECG analysis system

1.4.1 ECG Signal Preprocessing

During data acquisition, the ECG signal may be affected by different kinds of noise including physiological noise (e.g. myoelectricity noise, breathe interference etc.) and non-physiological noise (e.g. power-frequency interference and electrode impedance interference) [31].These noises often interfere with the informative signal and thus influence the ECG classification results. Therefore, ECG signal preprocessing mainly focuses on the suppression of noise and interference terms in the ECG signal.

The ECG signal is in millivolt (mV) level with a central frequency ranging from 0 to 40 Hz [32]. Due to the relatively low signal to noise ratio in ECG signals, signal preprocessing is a necessary step before classification. Therefore, various methods are proposed to eliminate noise and other artifacts from the ECG signal [31–38].

Generally speaking, ECG signal preprocessing methods includes finite impulse response (FIR) filtering, adaptive filtering, and modern signal processing filter methods such as wavelet transforms and neural networks [31, 34, 35, 38]. YW Bai *et al.* compared different notch filters and concluded that equiripple notch filter outperforms other methods in terms

of noise reduction and CPU time [34]. Lian *et al.* [33] designed a multiplier-free finite impulse response (FIR) filter to surpress biological and environmental noises with a low power consumption. Sayadi *et al.* proposed a modified extended Kalman filter with estimated hidden state variables to perform denoising and compression at simultaneously [35]. Park *et al.* designed a wavelet-based adaptive filter to reduce S-T segment distortion due to the baseline drift and compared its performance with general adaptive filters [36]. A general conclusion is that the performance of wavelet adaptive filtering is usually higher than generic adaptive filters. In [37], the authors combined wavelet decomposition with Wiener filtering to filter out the noise by thresholding, which is proved to outperform other thresholding denoising methods. Regarding various wavelet basis functions, Singh *et al.* studied an optimal selection of basis functions for ECG signal denoising [31]. By comparing the classification root mean square error using the same classifier and different denoising methods, they concluded that Daubechies filter of order 8 is the best choice for ECG classification system.

1.4.2 Fiducial Peak Detection and Segmentation

Fiducial Peak Detection and cardiac cycle segmentation are the basis for extracting important information from ECG signals, since a ECG record is usually a continuous time signal. This signal can be split into smaller intervals, each of which representing one cardiac cycle. Each cardiac cycle can be viewed as an independent signal and is associated with a separate label to represent the heart function during the corresponding interval. The accuracy and reliability of this stage directly determine the final performance of diagnosis and analysis.

Fiducial peak detection, which is also called ECG signal delineation, aims at localizing five characteristic peaks within one cardiac cycle. The most significant peak is the QRS complex consisting of Q, R and S peaks. The other two fiducial peaks include P wave before the QRS

complex and T wave after the QRS complex. As shown in Fig.1.5, these five characteristic waves along with the onset and offset of the QRS complex are often used to present a cardiac cycle.

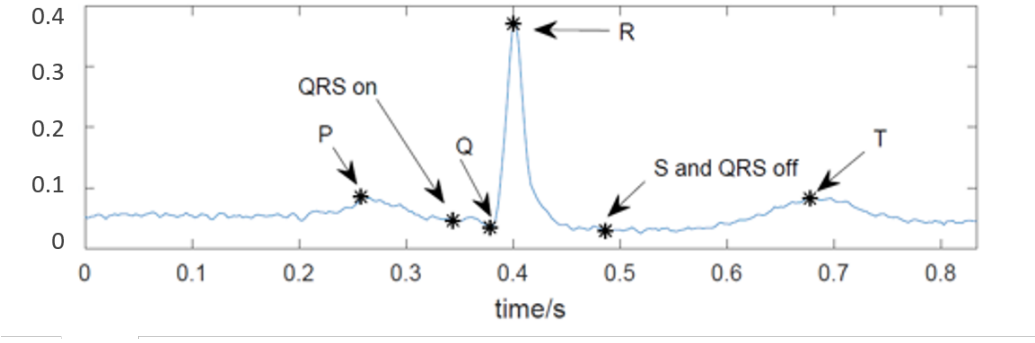

Figure 1.5: Fiducial peaks within one cardiac cycle

The QRS complex is the most prominent wave and it contains the majority of the information of a ECG signal; therefore, most of the ECG delineation methods detect QRS complex prior to the detection of other peaks. Afonso *et al.* proposed a method using filter banks to detect QRS complexes [39]. In this method, the signal is decomposed to several frequency bands. Fiducial peaks are thus detected using its morphological features in the decomposed signals. Sadhukhan *et al.* proposed a method of detecting R peak by thresholding the double difference signal of ECG data and comparing the relative amplitude within QRS region [40]. The performance of this method is validated using clinical ECG signals and has been proven to be promising. Some advanced machine learning techniques are also deployed to detect QRS complexes. In [41], *Support Vector Machine* (SVM) is used to train a predictive model for QRS complex detection and achived 99.93% accuracy. Other pattern recognition methods such as *Hidden Markov Models* (HMM) are investigated and proved to be effecient in modeling and detecting characteritic peaks in ECG signals [42]. Wavelet decomposition is also frequently adopted for signal delineation due to the morphological similarity between wavelet basis functions and QRS complexes. As the QRS complex power

spectrum is centered at the range of 5 to 30 Hz, the wavelet coefficients of the corresponding scale levels are frequently used for delineation purpose. In [43], QRS complexes are detected by thresholding wavelet coefficients at scales 1 to 4, then onset, offset and individual waves within QRS complexes are detected using the morphological characters of coefficient at scale 2. T and P waves are detected at scale 3 with a similar method approach. In the literature, some improvements have been proposed to eliminate false detection of R peaks by adding a fixed searching window of 160ms [44].

1.4.3 Feature Extraction and Classification

After localizing the fiducial peaks within a cardiac cycle, we proceed with the next step of extracting informative features of the signal, which collectively convey meaningful information about the signal properties. Since the objective of designing an automatic classification system is to precisely predict types of sample signals, feature selection is usually performed to obtain a better performance and reduce the computation cost [7–11].

As the most significant wave within a ECG signal, the information of QRS complexes are proved to be the most important features for ECG classification systems. Lagerholm *et al.* decompose QRS complexes with a set of Hermite basis functions and the decomposition coefficient are deplyed as ECG features to train a Self-Organizing Map (SOM), which achieved an average error rate of 1.5% for 16 ECG types [7]. Prasad *et al.* used discrete wavelet transform (DWT) to extract RR intervals between the current beat and previous or next beats. The two RR-intervals serve as input for training a neural networks, which achieves the average accuracy of 96.77% in classifying 13 different arrhythmia types. De Chazal *et al.* proposed two set of features: morphology and heartbeat interval features. They used different combinations of these features combined with Linear Discriminant Analysis to clas-

sify ECG signal into five arrhythmia types and selected the optimal feature set according to classification performance [9]. The result shows that the sensitivity of detecting two major arrhythmia types can be improved by feature selection. R. Ceylan *et al.*included RR interval as the only ECG feature to train a fuzzy clustering neural network that achieved an average detection rate of 98.35% [10] . Osowski *et al.* proposed two set of features including *Higher Order Statistics* (HOS) and Hermite characterization of QRS complex to classify ECG signals with Support Vector Machine. Their final average error rate is at 1.82% [11].

1.4.4 Patient-Specific ECG Classification

The main drawback of the majority of aforementioned methods mentioned in the last section is the lack of inter-patient model adjustment. In order to generalize the ECG classification system to clinical applications, several methods which are more robust to inter-patient signal variation are proposed to address this issue [12–17].

Hu *et al.* proposed a patient-specific Mixture of Experts (MOE) classifier by incorporating personalized annotations provided cardiologists in the local classifier [12]. The methods achieves patient-adapting capacity but requires further input from human experts. This MOE approach achieved an accuracy of 94.0% for distinguishing *Ventricular* beats from the other non *Ventricular* types. Following the design of MOE, de Chazal and B. Reilly proposed an improved patient-adapting classifier by reducing the requirement of manual annotations to as few as 10 beats for training adaptive local classifier [13]. And Llamedo et al. designed an automatic classification system, which uses experts' assistance, but does not fully depend on the experts and can work independently [14]. By implementing a special block-based neural networks (BbNNs), Jiang et al. achieved accuracies of 98.1% and 96.6% in distinguishing *Ventricular* ectopic beats and *supraventricular* ectopic beats from other types [15].

In [16], particle swarm optimization (PSO) is combined with a neural network to optimize the network structure using patient-specific training data. Based on 1-D convolutional neural networks (CNN), Kiranyaz et al. proposed a flexible algorithm, which adjusts its parameters using information extracted from individual signals [17]. The classifier demonstrates consistent performance over different ECG records achieving an accuracy between 98% and 99% for distinguishing VEBs from non-VEBs. (Acc = 98.9% Sen = 95.9% Spe = 99.4%). While this approach outperforms the aforementioned classification algorithms as it does not require expert further annotations, its performance reduces for some rare abnormal classes.

1.5 Contributions

A crucial drawback of these patient-specific classification systems, recently proposed in the literature, is their failure in predicting abnormalities in advance. These methods aim at improving classification performance by comparing generated labels with ground truth for each beat and ignore the relationship between the generated labels and upcoming abnormalities. While in many common applications this approach generates satisfying results, it does not meet the needs of SCD prediction.

One of the main objectives of this work is to address the problem of forecasting by proposing the concept of yellow and red alarms and proving the fact that yellow alarms can be indicators of upcoming red alarms. Yellow alarms are defined through a novel deviation analysis which assesses the tendency of deviant normal alarms to one of the red alarm. In order to realize such a deviation analysis, symmetry of different abnormality classes in feature space is desired. We propose a novel controlled nonlinear transformation that maps the original feature space into a new space that presents the desired symmetry. To elaborate more on the symmetry of abnormal classes in the feature space, we assume that there are one nor-

mal class and multiple abnormal classes for a signal while latent states exists for some of the normal samples that represent slight deviations towards abnormalities.By distinguishing latent states, the designed automated system is capable to generating a yellow alarm which indicates a high probability of the presence of some upcoming abnormalities (red alarms) of the same type. Therefore, the contribution of this work can be summarized as:

- Propose a novel self-configuring patient-adaptive framework which incorporates a personal classifier into the predictive modeling;
- Utilize a kernel-based method as a spatial transformation with parameters optimized using multi-objective particle swarm optimization (MOPSO) for the purpose of deviation quantification;
- Design a controlled spatial transformation with deterministic mapping function to optimize cluster topology for predictive analysis
- Propose a deviation quantification method based on cosine similarities, which is capable of generating red alarms for upcoming abnormalities

1.6 Organization of Thesis

In the following chapters, details of the proposed classification framework are presented after reviewing the introductory concepts and related works. Chapter 2 provides a general information about the utilized ECG dataset and the general framework used in proposed classification system. Chapter 3 describes the details of nonlinear transformation with kernel methods and presents the experimental results using kernel transformation. With the concept of nonlinear transformation, chapter 4 introduces an optimized spatial transformation with a

novel deterministic mapping function. The experimental result for the spatial transformation method is presented section 4.6. Finally, the experimental results for the proposed methods in chapter 3 and chapter 4 are studied and compared. More importantly, the predicting capacity of the proposed system is studied and analyzed in this chapter. Based on the experimental results, we introduce some potential directions to further improve the system in terms of classification and predicting accuracy.

Chapter 2

Patient-Adaptable ECG Classification Framework

2.1 Introduction

For decades, automatic ECG signal analysis has been a controversial research topic. Several academic research projects have proved that the design and implementation of automatic ECG analysis methods is beneficial for timely detection and therapeutic intervention of heart diseases. However, some major challenges should to be resolved, and automatic ECG analysis should reach a level of maturity and reliability before getting ready for clinical use. One of the most typical challenges is the inherent inter-patient variation of ECG waveforms, which leads to inconsistent performance of ECG classification systems when applied to different patients under different conditions. In this chapter, the details of the patient-adaptable ECG classification method, as the core of the proposed framework are presented.

The goal of automatic ECG analysis is to determine the arrhythmia types for each signal

sample. Continuous ECG signal is firstly segmented into individual segments, each of which represents a heartbeat cycle, to be processed by the further stages of the proposed algorithm. Section 2.1 in this chapter focuses on the data preparation stage, which includes four steps: signal preprocessing, delineation, segmentation, and feature extraction. Following the data preparation, Section 2.2 elaborates on the details of the proposed two-stage hierarchical classifier. The proposed classification system introduces a novel method for patient-adaptation by gradually capturing the normal range for each individual. More specifically, in Section 2.3, the dynamic normal cluster shaping method to achieve rationalization property is discussed. One feature of this method is that the cluster can track a patient's ECG waveform changes and dynamically adapt to it. In many applications, the physicians need to monitor a long-term real-time heart activity. This dynamic adaptation is able to address the issue of intra-patient temporal signal variation as well.

2.2 ECG Signal Processing

2.2.1 Preprocessing

Biomedical signals, such as ECG signals, composed of a sequence of signal segments, that can be presented by a set of statistical, morphological, and spectral features. Since the signal properties change over time, the traditional Fourier transform is not suitable for this type of non-stationary signals since it is unable to capture time-varying statistics of the signal. Wavelet decomposition solves this problem by scaling and translating mother wavelet to constitute its basis functions, which capture both spectral and temporal properties of the signal. Given a time series, wavelet decomposition decomposes the signal into linear combinations of basis functions. Thus, basis functions with larger scales are smoother than

those with smaller scales and consequently correspond to lower frequency components of the signal. Similarly, the coefficients of the decomposition correspond to higher frequencies when the scale is smaller. Using wavelet decomposition, we extract both time and frequency features of the signal.

In this work, wavelet analysis is applied to ECG signals in MITDB with a sampling frequency of 360 Hz. Daubechies wavelet of order 8 ($db8$) is selected as mother wavelet for denoising stage in this work for its optimal performance [31]. The decomposition coefficients and their corresponding frequency components are presented in Fig.2.1. Consistently use Fig. or Figure without dot "." throughout the thesis.

Low frequency noise or baseline wander between 0.15 to 1 Hz, caused by respiration and body movement, can be removed by deducting the approximation coefficient of level 8 (A_8) from the signal. Since the power of ECG signal is mainly concentrated in the frequency band from 1 to 40 Hz, higher frequency terms are more likely to represent noise terms including electromyogram induced noise and mechanical forces acting on the electrodes. These terms can be removed by discarding the detail coefficient of level 1 (D_1).

2.2.2 Segmentation

Most machine learning (ML) algorithms operate on input vectors and are not directly applicable to continuous-signals. Therefore, biomedical signals are typically converted to a representative vector before incorporating to ML algorithms. We follow the common trend of translating a signal segment into a vector of representative features. In this regard, we first need to split the signal in time domain into smaller segments. To obtain more relevant results, we choose the segments as one or multiple consecutive cardiac samples, noting the fact that each cardiac cycles is associated with a label based on experts observation.

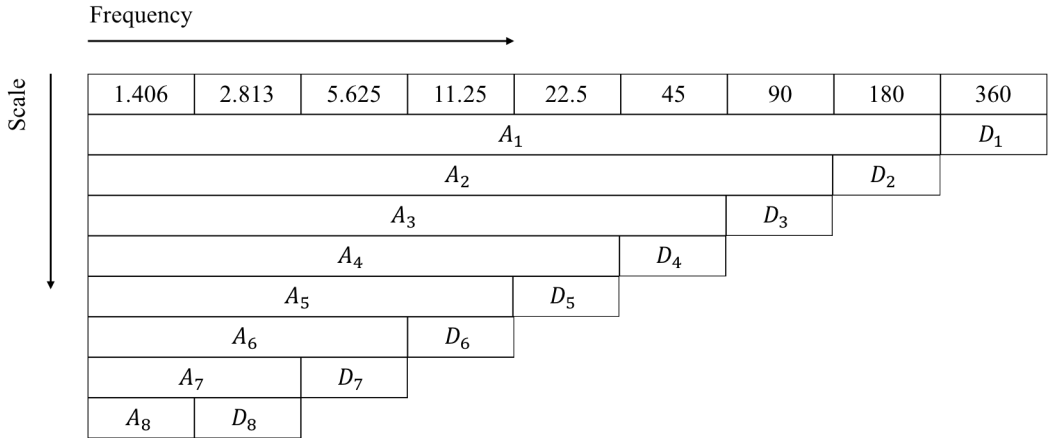

Figure 2.1: Frequency band of wavelet decomposition coefficients for MITDB signals

Most existing methods use wavelet analysis to detect the highest peak (R wave) in a cardiac cycle as a reference point and then use signal morphology and typical properties of other waves to determine boundaries between consecutive intervals [REF]. In this study, we use this common approach. As was earlier mentioned in Section 1.4.2, a cardiac cycle consists of five basic characteristic peaks: P, Q, R, S, and T. Among them, the QRS complex is the most significant peak in one cycle. The energy of ECG signal in one cardiac cycle is mainly concentrated within the QRS complex. The QRS complex also conveys important information that reflects the arrhythmia category [45]. Accurate detection of the QRS complex is of crucial importance for subsequent analysis. The energy of the QRS complex is generally within the range of 5-25 Hz. For ECG signals with sampling frequency of 360Hz, the QRS complex can be extracted from the detail coefficients of level 5 (D_5) and level 6 (D_6)

The mother wavelet *db*4 is utilized at this stage due to its morphological similarity to QRS complexes. By superimposing D_5 and D_6, the QRS complex information in the ECG signal can be characterized in a one-dimensional recombined signal ($QRS\_DET = D_5 + D_6$).Other Fiducial peaks (P, QRS onset, Q, S, QRS offset and T waves for each cardiac circle) are localized according to an algorithm proposed in [45]. The accuracy of peak detection and

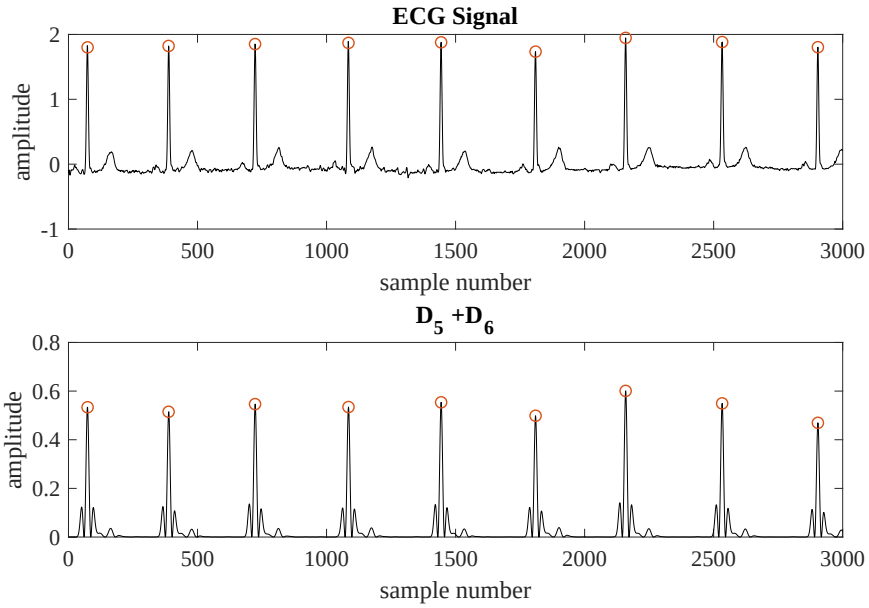

Figure 2.2: QRS detection with wavelet coefficients of level 5 and level 6. Add dot to all captions. Also, it is good to use subfigure with captions (a) and (b) and then include the full description in the caption.

its coincidence with the true signal is shown in Fig.2.2.

With the empirical values described in [45], we use 15% as the detection threshold. Since the width of most of the QRS complexes does not exceed 160ms, here we use a sliding window with a width of 160ms to detect the peaks in the QRS_DET. The window step size is set to 200ms, given that the time lag between the two adjacent heartbeat cycles does not exceed 200ms. Figure.2.3 shows a typical QRS_DET waveform along with the corresponding window of width 160ms. The false peaks are eliminated using a 160ms time window, as seen in Figure.2.3.

The T and P waves are outside the QRS window. Through scanning the region spanning the start of a QRS window to the end of its neighboring QRS I think you mean: the end of a cardiac interval to the beginning of QRS window in the next interval, the P-wave is located as the highest positive peak in this region. In a similar way, the position of the T wave is

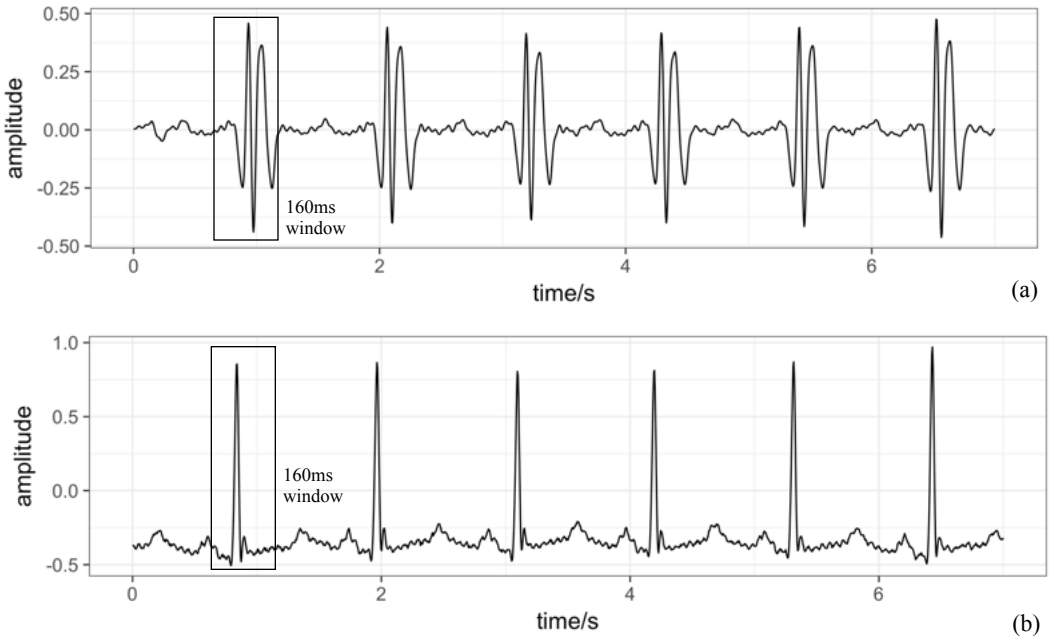

Figure 2.3: Window for detecting R peaks within QRS complexes

obtained by finding the maximum of the signal in the region between the from the end of the QRS in the current cardiac interval to the beginning of the next cardiac interval. In an alternative way, the two peaks between the two consecutive QRS windows, respectively represent the T and P waves. In short, an ECG signal in one heartbeat segment can be fully described by 7 feature points: P, QRS on, Q, R, S, QRS off and T.

After localizing the feature points, we can use the location of R peaks to determine the boundaries of a cardiac cycle. By processing a large number of ECG signals, we realize that R peaks approximately split cardiac cycles into two pieces with 1/3 and 2/3 of the entire signal. Add REF if any. Based on this observation in the morphology of a typical ECG signal, we define the starting point of a cardiac cycle as the point which divides the distance between the two consecutive R waves into two sections, where the length of the first section is half the length of the second section, as depicted in Fig. fig:interval. In this figure, the distance (RR) between the two adjacent R waves is denoted by h, therefore R wave is located

in distance $h/3$ with respect to the beginning of the cycle. The advantage of this method is its low computational complexity and easy generalization to different patients, noting that a slight change in cardiac boundaries do not significantly alter the properties of a cardiac cycle as long T and P waives remain in the correct interval.

The quality of ECG signals provided by most of the portable ECG measuring instruments is very unstable and may include transient noises. Signals transmitted through wireless communication systems will exhibit even more unstable waveforms. These transient effects appear in the resulting feature vector and consequently they negatively affect the prediction accuracy of the subsequent ML method. In order to eliminate and smooth out these transient terms, we use the concept f *segmentation* here, where each segment includes multiple cardiac intervals, as shown in Fig. 2.4. Note that the number of cycles in each segment is a modeling parameter shown by XXX. Also, we can arbitrarily slide segments equivalent to YYY cycles to generate the new feature vector. The parameters XXX and YYY can be tuned to improve the overall performance of the method. Based on intensive simulations, we choose $XXX = xxx$ and $YYY = yyy$, which yield the best classification accuracy. In Fig. 2.4, we have $XXX = xxx$ and $YYY = yyy$.

This is not well connected to the text before and after. Make it a separate Section titled as 2.2 Utilized Dataset and move it to the beginning of this section. For the purpose of training and evaluating classifier, MITDB is split into test (DS2) and training (DS1) set by balancing the four classes according to [9].

The ECG signals in MITDB dataset are annotated and labels are provided for each cycle. However, we define a segment, which may include more than one sample as a sample. Therefore, we need to translate per-cycle labels into per-segment samples. In this regard, a new label for each segment is generated by integrating all annotations of the beats within the segment. The segmend is labeled as *normal*, if all member beats are annotated as N;

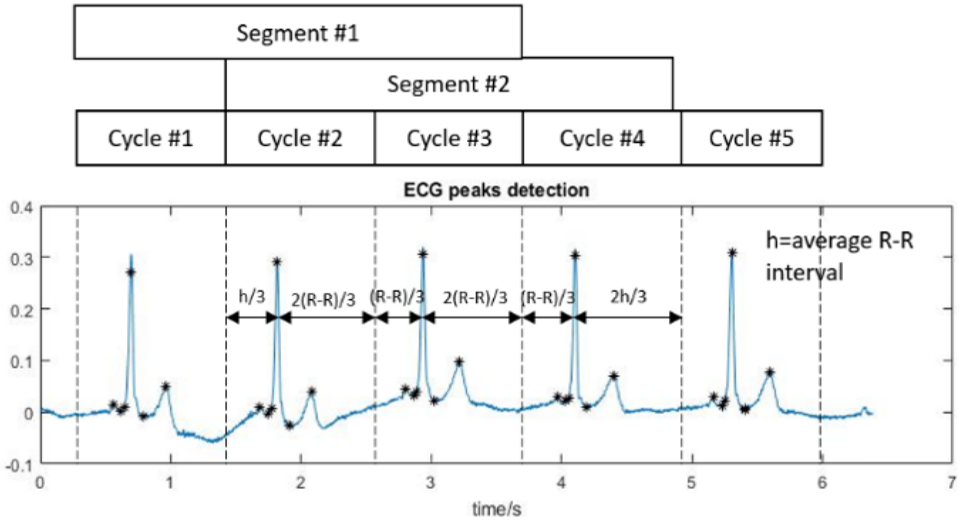

Figure 2.4: Segment samples correspond to three consecutive cardiac cycles

otherwise, this segment will be labeled as the abnormality type of its member beats if there is only one abnormality type. If more than one abnormality types are present within the same segment, the segment is excluded. For instance, segments with members labels "NNN", "SSS", "NVV", "VVV" are respectively mapped to "N","S", "V", and "V", and a segment with member label "NVS" is discarded. After segmentation and new annotation, the total number of samples in the training and test set are obtained as summarized in Table.2.1

2.3 Feature Extraction

The feature extraction step plays a crucial role in the diagnosis of heart disease and has a great influence on the performance of the subsequent automated classification systems. De

Table 2.1: Training and test datasets in MITDB.

	Number of segments per AAMI class				
Evaluation Dataset	N	V	S	F	Total
DS1:Training	11721	2356	862	256	15195
DS2:Test	12633	2053	550	121	15357
Total	24354	4409	1412	377	30552

Chazal et al. have studied the impact of using waveform morphological features on classification results [9]. As discussed in [20], the combination of three types of characteristics (i.e. temporal, morphology, and frequency domain features) can provide a better discriminative power for the classification algorithm to distinguish between different types of arrhythmia.

According to several studies different types of ECG signals mainly differ in the power level of the frequency band ranging from 5 Hz to 15 Hz [**?**]. Likewise, some other morphological features (such as the duration between the Q wave and the T wave, P The distance to the R wave, etc.) present different levels of correlation with a specific signal classes [**?**]. Therefore, we choose to use combination of temporal, morphological and spectral features as detailed in Table 2.2. Also, to account for segment-level characteristics as well as cycle-level properties, the extracted features include both periodic-based features (SET 1) and segment-based features (SET 2), where SET1 includes the average and standard deviation of the corresponding features of the three cardiac cycles within a segment, and SET2 contains the overall characteristics of the time signal within the segment, so it is calculated only once per segment. Therefore, we have a total of $8 \times 2 + 6 = 22$ features per segment as shown in Table 2.2. Therefore, each feature vector is a 221 vector with zero-mean unit-variance elements after a proper normalization. In Table .2.2, mean $(R_{i+1} - R_i)$ refers to the mean of the time lag between two adjacent R waves, while $(R_i - R_{avg})$ is the length of each cardiac cycle and the average cardiac cycle duration of the patient.

From Table.2.2, one may notice that these 22 feature are not completely independent of each other. Also, some of the features may not be as relevant as the others. Therefore, we reduce the number of features to obtain a more robust predictive modeling [**?**]. We use Principal Component Analysis (PCA) or principal component analysis (PCA): choose one notation as a commonly used the common dimension reduction method instead of explicit feature selection for its improved performance in biomedical signal processing [**?**]. We keep the 8

Table 2.2: Features extracted from ECG signal. Change column width such that each line represents one feature.

Feature Type	SET1	SET2
Temporal Features	QRS duration, QT duration, PR duration	mean$(R_{i+1} - R_i)$, mean$(R_i - R_{avg})$
Morphological Features	max positive peak to second peak ratio	signal average energy, max positive peak, max negative peak, peak to energy ratio
Frequency Domain Features	signal power level at 7.5Hz, 10Hz, 12.5Hz, 15Hz	

dominant directions of the signal after PCA as the most informative 8 features.

2.4 Classification Framework

part of the following text is related to global classifier, and part of it to deviation analysis (including personalized normal cluster, ...). Use subsections as needed to be more clear.

In this section, we elaborate on the details of the proposed methodoology to perform the classification and prediction tasks using the pre-processed ECG data. Based on our previous study [29] as well as similar prioir works on developing patient-specific classifiers [12–14], we propose a two-staged structure which includes a global classifier to capture general properties of different classes followed by a personalized classifier to capture patient-specific properties [12–14, 29]. Moreover, the proposed algorithm incorporates a novel deviation analysis module with details presented in the following sections.

The flowchart in Figure 2.5 presents the overall operation of the proposed system. The

global classifier is trained using the whole training dataset as the first classification step. It facilitates the subsequent analysis in the system by identifying samples with severe morbidity. Depending on the application (properties of signals, the utilized labels, and the choice of features), different classification algorithms can be utilized [**?**]. Two important considerations include the classification accuracy and the computation complexity of the method [**?**]. Any abnormal label generated by the global classifier is considered as a *red alarm* and does not require further processing. However, samples labeled as *normal* go through the subsequent personalized classification step.

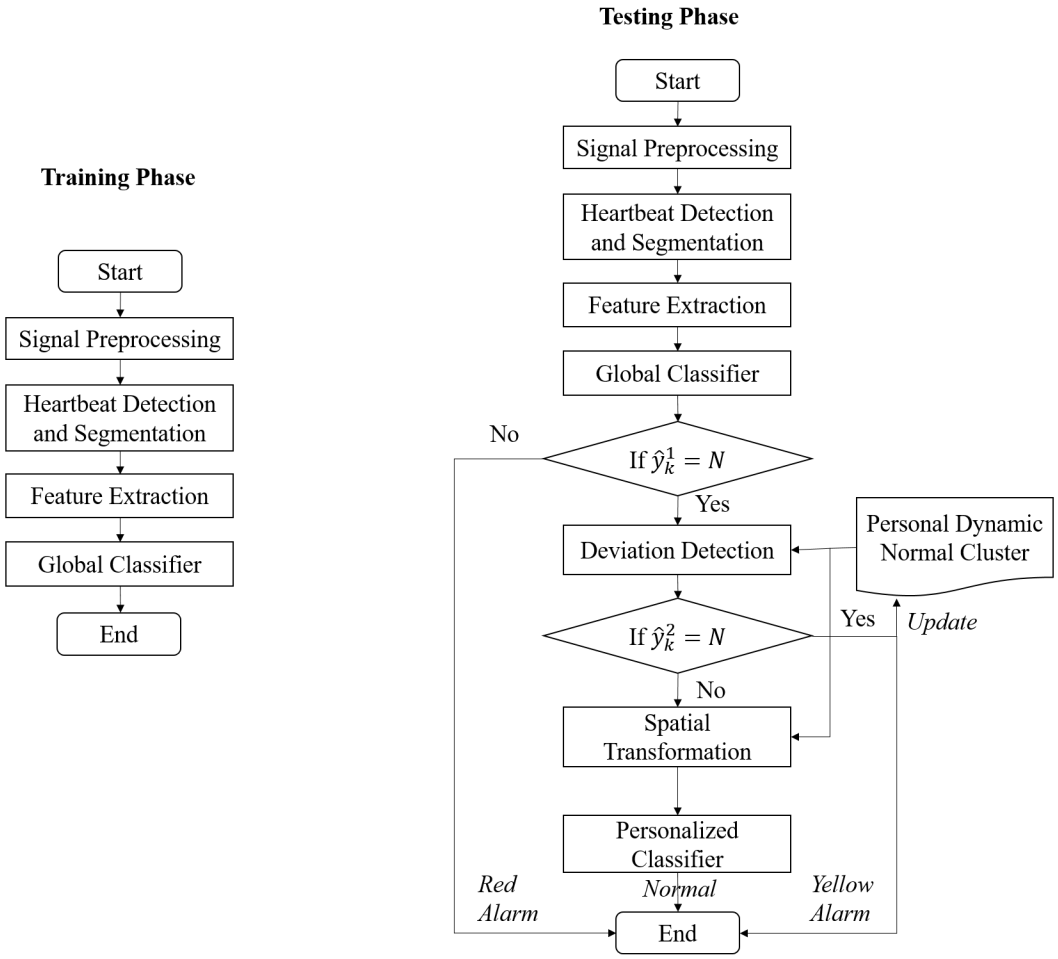

Figure 2.5: The general flowchart of proposed framework

Since, one objective of this study is to identify the fuzzy states between normality and ab-

normalities, the subsequent analysis is focused on processing samples classified as *normal*(N) use a consistent notation: you can choose to use italic font for "normal" and "abnormal", "red", "yellow", etc in their first appearance or for all appearances. Also, define labels N, S, ... on their first appearances. to check whether or not they show considerable deviations to one of the abnormality classes. For this purpose, a one-layer classifier is not sufficient due to multiple reasons: i) the numbers of samples in the normal and different morbid classes are not balanced in the training set DS1 as shown in Table.2.1, ii) the patient-specific normal trend is not known, iii) determining yellow alarms require a new set of decision rules to determine whether or not the deviations are worthy of calling a yellow alarm. Therefore, a deviation detection module is added after the Global Classifier to identify latent has a special meaning in ML, which is different than what you mean., I recommend change "latent" to a different word yellow alarms using patient-specific normal cluster. In order to develop a ground for patient-specific normal functionality and adapt the classifier accordingly, the first 20% of total Normal samples of each patient are selected as the initialization of *personalized dynamic normal cluster* $\mathcal{N}_0^k$. To do this, we use a binary classification of N versus non-N, where the second includes all abnormal classes. We firstly calculate the following distance metrics:

$$R_i^{\max} = \max_{\mathbf{x}_j \in \mathcal{N}_i^k, \mathbf{x}_k \in \mathcal{N}_i^k} \{\sqrt{(\mathbf{x}_j - \mathbf{x}_k)^2}\}, \tag{2.1}$$

$$D_{\mathcal{X}}(\mathbf{x}_k(i)) = \underset{\mathbf{x} \in \mathcal{X}}{\text{median}} \{\sqrt{(\mathbf{x}_k(i) - \mathbf{x})^2}\}, \tag{2.2}$$

$$D_{\mathcal{N}}^{\max}(\mathbf{x}_k(i)) = \max_{\mathbf{x} \in \mathcal{N}_i^k} \{\sqrt{(\mathbf{x}_k(i) - \mathbf{x})^2}\}, \tag{2.3}$$

The following conditions are then examined to verify if the deviation of a sample is within the range defined by α. Since some rare abnormalities are unlikely to be observed within the

limited initialization period, therefore abnormal clusters: $\mathcal{S}, \mathcal{V}, \mathcal{F}$, which include abnormal beats for all patients in DS1, are deployed as the training dataset when developing the *personalized dynamic normal cluster* $\mathcal{N}_i^k$ as follows. I think here in $\mathcal{N}_i^k$, you mean that $i = 0$ represents the normal cluster and $i = 1, 2, ...$ represent different abnormality clusters. This is not obvious and you need define it before its first appearance.

$$\begin{cases} D_{\mathcal{N}}^{\max}(\mathbf{x}_k(i)) \leq \alpha R_i^{\max}, \\ D_{\mathcal{N}}(\mathbf{x}_k(i)) < D_{\mathcal{X}}(\mathbf{x}_k(i)) \quad \text{for } \mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\} \end{cases} \tag{2.4}$$

If a sample already called normal by the global classifier, is again confirmed as N in this module, it will be used to update the $\mathcal{N}_i^k$. Otherwise, the system assumes that the sample has a considerable deviation towards one of the abnormal clusters and hence will pass it to the subsequent *personalized classifier*. The *personalized classifier* uses controlled transformation with optimized parameters to discern the deviation to different morbid types regardless of the cluster topology within the original feature space , as detailed in Section XXX.

After performing both global and personalized classification steps, a given sample x_k at time k is mapped to label $\hat{y}_k \in \{N, Y_V, Y_S, Y_F, R_V, R_S, R_F\}$, where N stands for normal status, Y_X stands for a ventricular yellow alarm of type "X", and R_X stands for a ventricular red alarm of type "X".

2.5 Personal Classifier

I suggest you use the term personalized classifier, although personal classifier is correct too.

Here, we provide the core idea behind the design of the personalized classifier, and the details

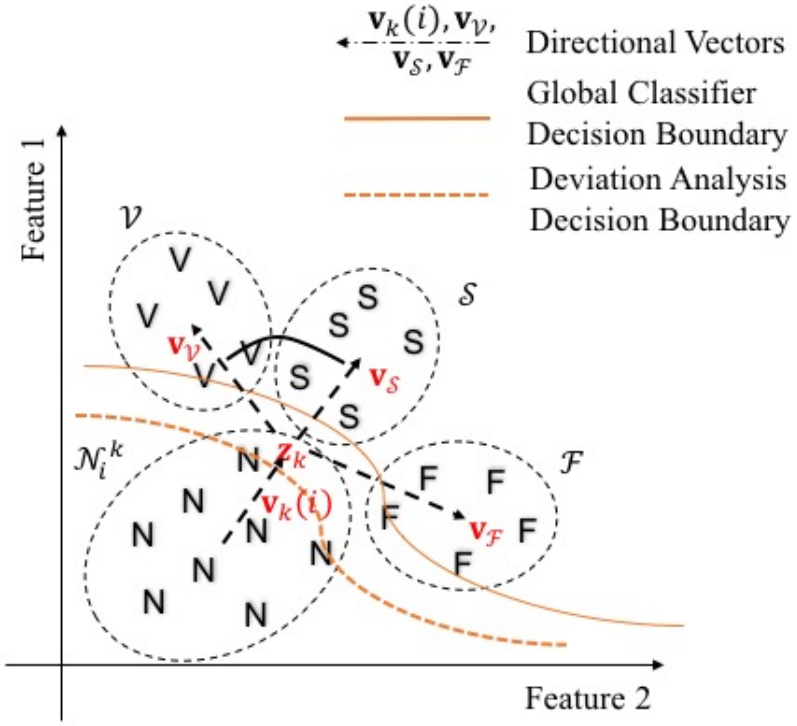

Figure 2.6: The deviation analysis boundary restrict on latent status between normal and abnormal samples compared with the Global Classifier boundary

of implementation will be discussed in Chapters 3 and 4. Since the proposed system aims at predicting subsequent abnormalities by analyzing a signal's deviation from the patient's normal functionality of the sample signal, it is vital to quantify the deviations using topological characteristics of the training dataset. For most of the ECG applications, ECG signal is analyzed by its representative feature vectors. A natural choice for deviation analysis in the high-dimensional feature space, is Cosine Distance *cosine distance* as defined in Eq.2.5) to quantify the distance between two vectors $\mathbf{v}$ and $\mathbf{w}$.

$$d(\mathbf{v}, \mathbf{w}) = 1 - \frac{\mathbf{v}^T\mathbf{w}}{|\mathbf{v}||\mathbf{w}|} = 1 - \frac{\mathbf{v}^T\mathbf{w}}{\sqrt{(\mathbf{v}^T\mathbf{v})(\mathbf{w}^T\mathbf{w})}} \tag{2.5}$$

Consequently, relative deviations of a sample from normal cluster(N) to other abnormal clusters (V, S, F) are defined by the cosine distance between the vector $\mathbf{v}_k(i)$ (defined in Eq.2.6) and the three vectors $\mathbf{v}_{\mathcal{X}}(i) = \mathbf{c}_{\mathcal{X}} - \mathbf{x}_k(i)$ where $\mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\}$. In this case, a smaller cosine distance represents a higher alignment between the vector from the normal cluster centroid $\mathbf{v}_k(i)$ to the current sample $\mathbf{x}_k(i)$ and the reference vector from the normal cluster centroid to abnormal centroids $\mathbf{c}_{\mathcal{X}}$.

$$\mathbf{v}_k(i) = \mathbf{x}_k(i) - \mathbf{c}_N^k(i) = \mathbf{x}_k(i) - \sum_{\mathbf{x} \in \mathcal{N}_i^k} \mathbf{x} / |\mathcal{N}_i^k|, \tag{2.6}$$

Therefore, the classification result of the Personal Classifier $\hat{y}_k^2(i)$ is determined as follows:

$$\hat{y}_k^2(i) = \underset{\mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\}}{\operatorname{argmin}} \{d(\mathbf{v}_k(i), \mathbf{v}_{\mathcal{X}}(i))\} \tag{2.7}$$

The relevance of cosine distance depend on the topology of clusters in the feature space. The topology in feature space, itself is inherited from the feature extraction and feature selection methods. For example, as shown in Fig.2.7, in the original feature space overlaps and alignment of abnormal clusters may lead to inaccurate results of deviation quantification. In order to eliminate the deviation analysis errors that arise from the asymmetry in the clustering topology, it is desired to transform the original topology into a more symmetric topology, where cosine distances directly reflect the amount of deviations.

For this purpose, two different spatial transformation methods are proposed in the next two chapters.

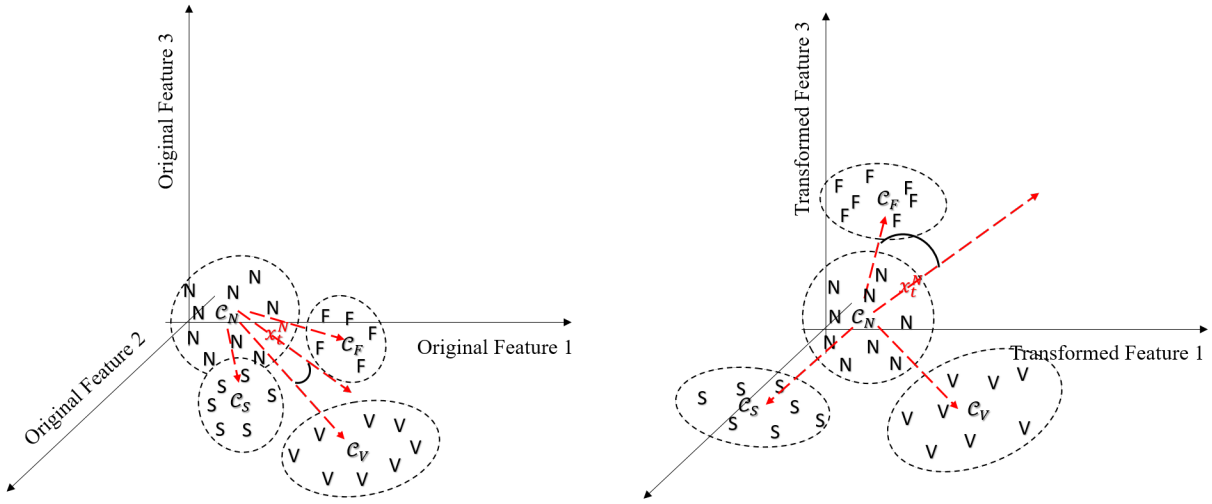

Figure 2.7: Left: illustration of the cluster topology in the original feature space; Right: illustration of the cluster topology in the transformed feature space using a nonlinear mapping function.

Chapter 3

Kernel-Based Nonlinear Spatial Transformation

3.1 Introduction

As proposed in last chapter, objective of the system is to predict subsequent abnormalities by quantifying the deviation of sample. Whereas the original cluster topology in feature space Ω^d depends on the feature extraction stage $g()$ and may lead to poor performance or even failure of predicting. An optimization using spatial transformation is addressed to solve this problem. More specifically, the designed method is able to maximize angles between vectors of normal cluster centroid to each abnormal cluster center centroids. In this chapter, a kernel based nonlinear spatial transformation is proposed to reshape the feature space to a symmetric topology which has the following features:

- Vectors pointing from normal centroid to different abnormal clusters centroids has no overlapping or cross with each other.

- Vectors pointing from normal centroid to different abnormal clusters centroids has maximum mutual cosine distance.

- The overlapping part of all clusters is minimized.

3.2 Kernel Method

First, given that clusters do not meet the symmetric features mentioned above, it's assumed in this chapter that new feature space obtained by non-linear reshaping is noted as $\Phi^{d'}$. This reshaping process is included in Personal Classifier stage (as shown in Fig.2.5) of the ECG classification system as described in chapter 2. The nonlinear mapping projects each sample $\mathbf{x}_k$ onto a new vector $\mathbf{z}_k$ in a higher dimension space, which can be considered as the union of subspace for each type: $\Phi^{d'} = \{\mathcal{N}, \mathcal{S}, \mathcal{V}, \mathcal{F}\}$, where $d' > d$

Kernel method has been widely applied in the machine learning algorithms, among which nonlinear support vector machine (SVM) has been deployed in numerous applications recently [46]. Nonlinear kernel method is most efficient when there is a nonlinear relationship between input and output variable. This nonlinear relationship assumption is usually true for ECG classification due to the complex feature vectors for ECG analysis. Therefore, it is practical to apply this method in the ECG analysis.

In nonlinear SVM, the algorithm minimize the expected error $E[L(y, f(x)]$ so as to obtain classification function f, where L is the designated loss function, such as square error $(y - f(x))^2$ [47]. Based on input space $x_i \in \Omega$ and classifier mapping function f, we first define loss function through following formula:

$$\frac{1}{m}\sum_{j\in\mathbb{N}} L(y_j, f(x_j) + \gamma||f||_K^2 \tag{3.1}$$

where, $||f||_K^2$ is the norm of f on H_K. If H_K is the Hilbert space of linear functions, $f = w'x,\ w \in \mathbb{R}^d$. Therefore the loss function can be written as:

$$\frac{1}{m}\sum_{j\in\mathbb{N}} L(y_j, w'x_j) + w'w \tag{3.2}$$

Positive constant γ known as regularization parameter, controls the balance between training error and the complexity (smoothness) of solution. SVM and other machine learning methods deploy kernel method with selected loss function [48]. In general, the solution to Eq.3.2 is as the following:

$$f(x) = \sum_{j\in\mathbb{N}} c_j K(x_j, x) \tag{3.3}$$

where c_j is a real number, K is kernel function, such as polynomial function: $K(x,t) = (x't)^r,\ x,t \in \mathbb{R}^d$

Hibert space H_K is usually defined base on feature mapping concept, $\Psi : \Phi^d \to \Phi^{d'}$ where $\Phi^{d'}$ is Hilbert space and $<,>_\Phi$ represents inner product. Therefore, Eq.3.3 can be also written as:

$$f(x) = \sum_{j \in \mathbb{N}} c_j \Psi(x_j, x) \tag{3.4}$$

For nonlinear kernel function method, the selection of kernel function plays a decisive role. An effective kernel function generally needs to satisfy Mercer theorem [49]. This means the matrix defined by function $\Psi(x_j, x)$ is symmetric positive semi-definite. In general, the selection of kernel function depends on all observations in input space. However, in effect, kernel functions with simple expression, such as polynomial kernel function, Gaussian kernel function and exponential kernel function are usually preferred than complicated kernel function for its simplicity and consistency. Polynomial kernel function is usually applied on normalized data for its explicit expression and steady performance. Whereas degree of freedom of polynomial kernel function is comparatively high which leads to a large of parameters. Gaussian kernel function, a very classic robust radial function, has high robustness when the data includes strong noise. Nevertheless, since Gaussian kernel actually projects samples to a infinite dimensional space, it's difficult to visualize projected observations and interpret the result. Moreover its performance is greatly affected by parameter selection.

In this work, the representative polynomial kernel is selected for the purpose of validating the method and explaining the effect of nonlinear kernel method and optimization on the feature space reshaping. The kernel functions can be written in the following format:

$$\mathbf{z}_k = \mathbf{\Psi}_{\mathbf{w}}(\mathbf{x}_k) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d'} \end{bmatrix} \circ \begin{bmatrix} \psi_1(\mathbf{x}_k) \\ \psi_2(\mathbf{x}_k) \\ \vdots \\ \psi_{d'}(\mathbf{x}_k) \end{bmatrix} \tag{3.5}$$

The example above shows that regardless the selection of kernel function, parameter optimization will play a critical role. For instance, in this work, the process of spatial topology optimization is accomplished by adjusting the coefficients of fixed polynomial basis functions $\psi()$. Since the number of parameters to adjust is very large when the order of polynomial functions is high, exhaustive algorithms are not practical for parameter optimization. Therefore, it's necessary to implement a heuristic optimization algorithm, in which parameters are obtained by maximizing or minimizing a loss function. More specifically, the nonlinear reshaping in this work aims to adjust mapping coefficient $\mathbf{w} = [w_1, w_2, \dots w_d]^T$ to achieve the ideal symmetry of clusters in the reshaped feature space while maintaining the distance between clusters.

3.3 Multiobjective Optimization

3.3.1 Objective Functions

To illustrate the optimization problem, if we assume the original feature space is the 2-dimension space Ω^2, then the mapping kernel function may adopt second-order polynomial function as follows:

$$\mathbf{x} = [x_1\ x_2]^T, \quad \mathbf{w} = [w_1\ w_2\ \ldots\ w_5]^T, \quad d = 2, \quad d' = 5$$

$$\psi_1(\mathbf{x}) = x_1, \psi_2(\mathbf{x}) = x_2, \psi_3(\mathbf{x}) = x_1^2, \psi_4(\mathbf{x}) = x_2^2, \psi_5(\mathbf{x}) = x_1 x_2 \tag{3.6}$$

Similar to the loss function in the standard nonlinear kernel method, the following objective functions can be used to represent the symmetrical structure to be obtained:

$$o_1(\mathbf{w}) = \frac{1}{\min\limits_{c,d=2,\ldots,p \text{ and } c\neq d} \{d(\mathbf{v}_{\mathcal{X}_c}, \mathbf{v}_{\mathcal{X}_d})\}} \tag{3.7}$$

$$o_2(\mathbf{w}) = \frac{SW}{SB} = \frac{\sum_{c=1}^{C} \sum_{\mathbf{z}\in\mathcal{X}_c} (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})^T (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})}{\sum_{c=1}^{C} \sum_{d=1,d\neq c}^{C} (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})^T (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})}$$

The maximization of pair cosine distance between vectors $\mathbf{v}_{\mathcal{X}_c}$ connecting the centroid of the normal class to the abnormal classes $\mathcal{X}_c$ can be gained by minimizing $o_1(\mathbf{w})$. In the formula, vectors used to measure symmetry is defined by abnormal sample sets in training set DS1 and the personal normal cluster as follows:

$$\mathbf{v}_{\mathcal{X}_i} = \mathbf{c}_N^k - \mathbf{c}_{\mathcal{X}_i} \tag{3.8}$$

On the other hand, $o_2(\mathbf{w})$ represents the ratio of within cluster variance to between-cluster variance, hence controls the separation between clusters. Cosine distance is defined by Eq.2.5 and these objective functions are deduced from discrimination function of personal classifier in Eq.2.7. By minimizing these two objective functions at the same time, the algorithm

eliminates the ambiguity while applying cosine distance and hence improve the predicting capacity.

3.3.2 Multiobjective Particle Swarm Optimization

It should be noted that $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ are not necessarily independent to each other. Thus the optimization problem here means the jointly minimizing of $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ subject to constraint condition $|w|_2 = 1$. Since this is a non-convex multiobjective optimization problem, closed form solution or optimization methods for convex function are not suitable for this problems. Therefore, this paper adopts multiobjective particle swarm optimization (MOPSO) algorithm to solve it.

Particle Swarm Optimization (PSO) has the advantage of fast-converging, heuristic searching and easy implementation [50, 51]. Therefore, researchers have already started investigating in extending PSO to multiobjective optimization problems. In the framework of MOPSO, the goal is to solve the typical Pareto optimization problem with the structure of PSO. In other word, it aims at solving optimization problem with two or more conflicting objective functions by approximating the Pareto front.

Among all MOPSO algorithms in literature, the algorithm proposed by Coello Coello and Lechug facilitates the implementation and improved the performance compared to other methods [50]. Therefore, this algorithm is adopted in this work. One featured design of this algorithm is the external repository in which all Pareto optimized particles in each swarm is recorded for each iteration. The configuration of repository members are stored and used as an optimal approximation of the Pareto front of the problem as they converge to the actual Pareto front as proved in [50]. Fig.3.1 generated by optimizing $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ demonstrates the repository members are Pareto optimal than other particles in the

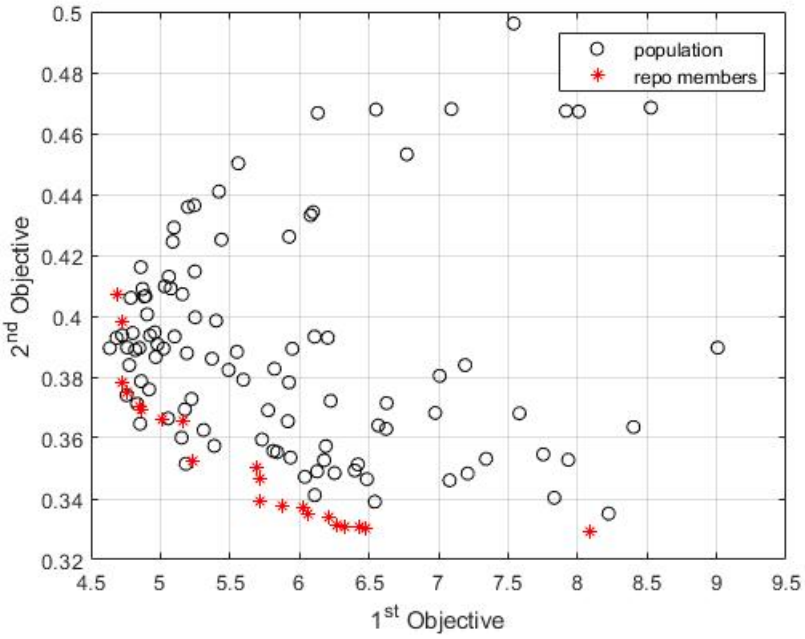

Figure 3.1: Particles stored in external repository approximate the Pareto front

objective function space and they converge to the Pareto front.

With the concept of Pareto front,we further demonstrate the impact of applying kernel functions in this multiobjective optimization problem by comparing Pareto front before transformation with kernel method and the Pareto front with nonlinear reshaping.

As shown in Fig.3.2, the Pareto front in feature space after nonlinear reshaping is superior obviously to that of Pareto front with linear combination of original data. The kernel function used in this comparison is a third-order polynomial kernel function as formulated in Eq.3.6. The result shows that nonlinear kernel method possess a higher degree of freedom to in multiobjective optimization. In other words, kernel method combined with multiobjective particle swarm optimization algorithm can improve the spatial topology of clusters.

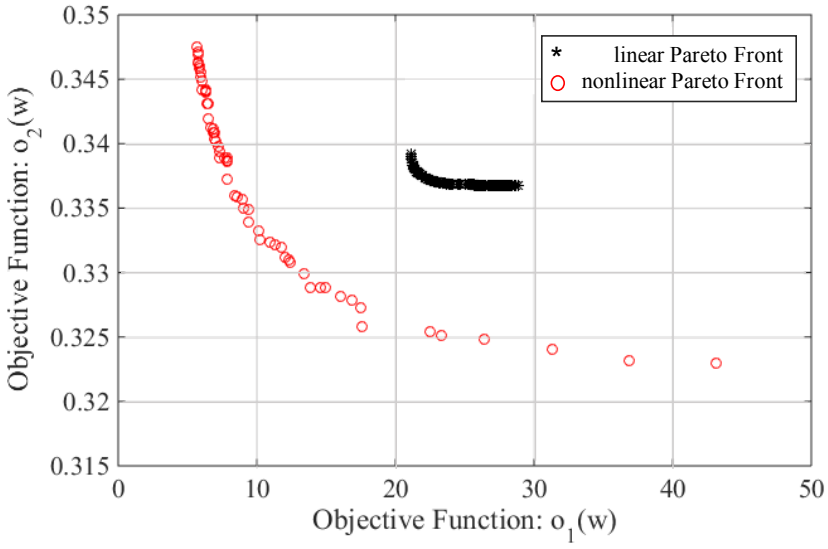

Figure 3.2: Increase of degree of freedom in optimization is proved by comparing the Pareto fronts generated by linear and nonlinear basis function

3.4 Experimental Results

This section will introduce the experimental result generated with test set DS2 of MITDB to show the performance of the proposed method in this Chapter. Here, we firstly map the original 22-dimensional feature vectors representative of cardiac segments into a 8-dimensional vectors $\mathbf{x}_{8\times1}$ using principal component analysis (PCA). In order to realize the non-linear transformation in (3.5), we use polynomial function of order 3. Taking the computational cost into account and also to avoid over-fitting of higher-order data samples, we only retain 32 terms (8 second order pure terms x_i^2, 8 third order pure terms x_i^3, 8 second order cross terms $x_i x_j$, and 8 pure third order cross terms $x_i x_j^2$) and randomly discard the rest of cross terms. Therefore, the mapped vectors $\mathbf{z}_{32\times1}$ include a total of 32 terms as follows.

$$\mathbf{z} = \{x_i^2 | i = 1, 2 \dots 8\} \cup \{x_i^3 | i = 1, 2 \dots 8\} \cup \tag{3.9}$$
$$\{x_i x_j | i, j = 1, 2 \dots 8, i \neq j\} \cup \{x_i^2 x_j | i, j = 1, 2 \dots 8, i \neq j\}$$

The performance of the proposed system in this paper is tested with DS2 excluding record 232, for this record has only 7 normal samples $y_k = N$. In total, 21 records are tested.

Table 3.1, shows the performance of the proposed method in classification of ECG signal segments. In order to average the results over all recordings, we present the median, interquartile range (IQR), mean and standard deviation of accuracy (AC), sensitivity (SE) and specificity (SP). The results are promising and the median of accuracy for all classes are in the range of $88\% - 99\%$. Sensitivity and specificity of the proposed method exhibits similar ranges. The mean accuracy is at least 86% excluding class V. Therefore, this system is very less likely to miss an important alarm or to report false alarms.

Table 3.1: Classification results of the proposed method.

Class N	median(%)	IQR(%)	mean(%)	std (%)
AC	94.8	19.52	86.62	18.55
SE	97.21	17.36	87.47	19.26
class V	median(%)	IQR(%)	mean(%)	std (%)
AC	86.11	27.54	76.41	22.81
SP	99.71	11.22	90.18	18.52
class S	median(%)	IQR(%)	mean(%)	std (%)
AC	99.28	2.24	98.29	2.57
SP	99.64	22.17	97.56	6.06
class F	median(%)	IQR(%)	mean(%)	std (%)
AC	97.91	8.2	93.85	7.84
SP	100.00	0.03	99.12	3.6

More importantly, the predicting ability of the proposed method is worthy of evaluating separately. In order to quantify the conditional probability of observing a red alarm after a

Table 3.2: Predictive power of yellow alarms: A yellow alarm increases the chance of observing a red alarm of the same type.

	Number of next abnormality				Probability of next abnormality (%)			
secondary abnormalities	V_y	S_y	F_y	Total	V_y	S_y	F_y	Total
True V	38	23	35	96	75	75	61	67
True S	11	10	8	29	21	29	14	20
True F	2	2	14	18	4	6	25	13

preceding yellow alarm of similar type in (2.7), we count the number of predicted samples as follows:

$$P(\hat{y}_{k+i} = X_r | \hat{y}_k = X_y) = \frac{\#\text{ of } y_{k+i} = X \text{ after } \hat{y}_k = X_y}{\#\text{ of true alarms after } \hat{y}_k = X_y}$$
$$P(\hat{y}_{k+i} = X_r) = \frac{\#\text{ of true alarm of type } X \ (y_k = X)}{\#\text{ of all true alarms}} \tag{3.10}$$

The summary of results for all 21 test records is presented in Table. 4.4. The last 4 columns of the Table. 4.4 show the probability of having a subsequent true abnormality of any type after observing a yellow alarm. These results confirm the predictive power of yellow alarms. For instance, the absolute probability of observing a segment with abnormal classes V, S, and F is respectively $\frac{96}{96+29+18} = 67\%$, $\frac{29}{96+29+18} = 20\%$ and $\frac{18}{96+29+18} = 13\%$, based on their relative frequencies. However, these probabilities after observing a yellow alarm of type Vp are respectively $\frac{38}{38+11+2} = 75\%$, $\frac{11}{38+11+2} = 21\%$ and $\frac{2}{38+11+2} = 4\%$. This means that the probability of observing a red alarm of type V is $75\% - 67\% = 8\%$ higher than its absolute probability. The same trend holds for other yellow alarms as well. The results suggest a more in-depth study of the concept of yellow alarms for heart monitoring.

3.5 Conclusions

This chapter explains the nonlinear reshaping with kernel function and multipurpose particle swarm optimization method. With the kernel method adopted in SVM, we utilize a group of nonlinear kernels to reshape the input feature space and map it to high-dimensional feature space, which meets two conditions, namely, maximum separation between cluster and maximum cosine similarities between abnormal clusters.

In this chapter we adopt multipurpose particle swarm optimization method to optimize parameters of kernel function. Result shows that Pareto front produced by kernel method in the objective function space is apparently optimal to that produced by linear combination of features. The results verify the accuracy of the proposed method with a classification accuracy range of $88\% - 99\%$ for different ECG records in publicly available MIT-BIH database.

Above all the proposed methodology demonstrates a potential to add detailed information about sample deviation upon conventional machine learning algorithm. We tested this system with ECG signal and observed promising results, but this method is not bound to this application.

Chapter 4

Controlled Spatial Transformation With Deterministic Mapping Function

4.1 Introduction

In Chapter 2, we elaborated on the details of patient adaptable ECG classification framework. In Chapter 3, the specific spatial topology of normal and abnormal clusters was analyzed in order to model the trajectory of ECG samples evolving from normal status to abnormal. These two methods share a common concept of modeling the intermediate states from annotated normal to abnormal state.

While the aforementioned systems demonstrated capacity of predicting upcoming abnormalities, it's challenging to interpret the mechanisms of the systems and thus hindering the generalization of predictive warning to other applications of biomedical signals. Therefore, the main object of this chapter is developing a classification system with abnormality predicting capacity based on spatial topology studied in [29].

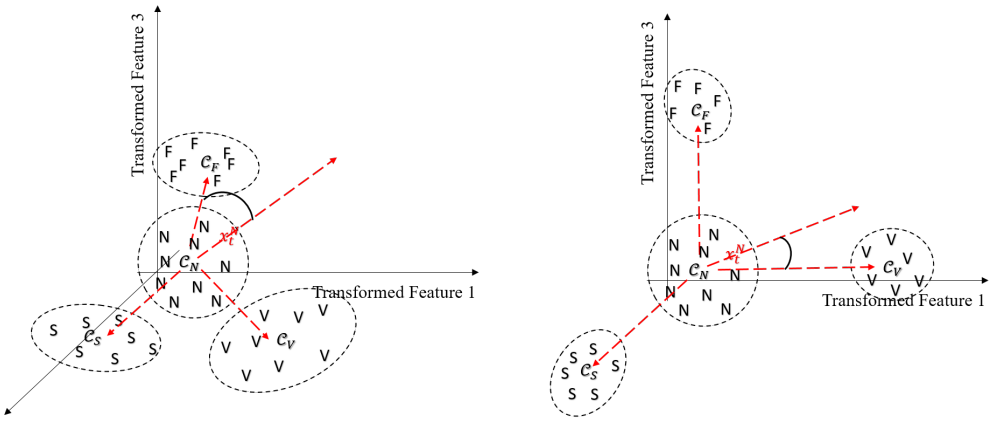

Figure 4.1: Left: illustration of cluster topology in feature space transformed with simple mapping function; Right: illustration of cluster topology in feature space transformed with optimized mapping function

Having analyzed the symmetric encircled topology in the previous chapter, we proposed a novel spatial transformation based predictive modeling system to assist cardiologist take advanced therapeutic Interventions. In this method, we further optimize the topology and spatial geometry of clusters in the feature space by reducing the within cluster variance after spatial transformation. As shown in 4.1, through this improvement, the predicting accuracy of personal classifier can be further improved.

4.2 Hyper-Spherical Coordinates

In the precedent chapter, the spatial transformation module was implemented with kernel function and heuristic optimization algorithm. The system performance was proved to be promising. However it's difficult to select the appropriate kernel as spatial mapping function due to the high variety of kernel functions and the limitation of dimensionality.

In order to address this problem, a novel deterministic spatial mapping function is proposed based on hyper-spherical coordinates in this work. Since hyper-spherical coordinates con-

sists of angles and radius, this information together with the cluster topology are used to determine the mapping function parameters.

Hyper-spherical coordinates (n-dimensional spherical coordinates) and its mapping to Cartesian coordinates are first elaborated in [52]. If $\mathbf{x}$ is a sample vector in n-dimensional feature space, where its Cartesian coordinates is $(\xi_1,\ \xi_2\ ...\xi_n)$, then its corresponding hyper-spherical coordinates can be obtained through Eq.4.1 which is originally derived through its reverse mapping (Eq.4.2) using $\sin(\arccos(x)) = \sqrt{1-x^2}$

$$
\begin{aligned}
r &= \sqrt{{\xi_n}^2 + {\xi_{n-1}}^2 + \cdots + {\xi_2}^2 + {\xi_1}^2} \\
\theta_1 &= \arccos \frac{\xi_1}{\sqrt{{\xi_n}^2 + {\xi_{n-1}}^2 + \cdots + {\xi_1}^2}} \\
\theta_2 &= \arccos \frac{\xi_2}{\sqrt{{\xi_n}^2 + {\xi_{n-1}}^2 + \cdots + {\xi_2}^2}} \\
&\vdots \\
\theta_{n-2} &= \arccos \frac{\xi_{n-2}}{\sqrt{{\xi_n}^2 + {\xi_{n-1}}^2 + {\xi_{n-2}}^2}} \\
\theta_{n-1} &= \begin{cases} \arccos \frac{\xi_{n-1}}{\sqrt{{\xi_n}^2 + {\xi_{n-1}}^2}} & \xi_n \geq 0 \\ -\arccos \frac{\xi_{n-1}}{\sqrt{{\xi_n}^2 + {\xi_{n-1}}^2}} & \xi_n < 0 \end{cases}
\end{aligned}
\tag{4.1}
$$

$$
\begin{aligned}
\xi_1 &= r\cos(\theta_1) \\
\xi_2 &= r\sin(\theta_1)\cos(\theta_2) \\
\xi_3 &= r\sin(\theta_1)\sin(\theta_2)\cos(\theta_3) \\
&\vdots \\
\xi_{n-1} &= r\sin(\theta_1)\cdots\sin(\theta_{n-2})\cos(\theta_{n-1}) \\
\xi_n &= r\sin(\theta_1)\cdots\sin(\theta_{n-2})\sin(\theta_{n-1}).
\end{aligned}
\tag{4.2}
$$

where $0 \leq \theta_j \leq \pi,\ j = 1, \ldots, n-2;\ 0 \leq \theta_{n-1} \leq 2\pi;\ 0 \leq r < \infty$

4.3 Orthogonalization

To simplify algorithms, in this paper the topology of clusters in feature space is approximated by relative topology of cluster centroids: $\mathbf{c}_N^k, \mathbf{c}_\mathcal{V}, \mathbf{c}_\mathcal{S}, \mathbf{c}_\mathcal{F}$. Furthermore, as we assume that samples deviates from normality to abnormality and spatial topology stays unchanged if simple translation is applied to data, the cluster topology in original feature space can be simply represented by the following matrix:

$$
\mathcal{C} = \begin{bmatrix} \mathbf{c}_\mathcal{V} - \mathbf{c}_N^k \\ \mathbf{c}_\mathcal{S} - \mathbf{c}_N^k \\ \mathbf{c}_\mathcal{F} - \mathbf{c}_N^k \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{\mathcal{VN}} \\ \mathbf{V}_{\mathcal{SN}} \\ \mathbf{V}_{\mathcal{FN}} \end{bmatrix} \tag{4.3}
$$

As shown in Fig.4.1, in order to improve Personal Classifier and avoid ambiguity, a topology with maximum separation between three vectors in $\mathcal{C}$ is ideal. In order to lower computing

complexity in high dimension, the algorithm aims at transforming vectors in $\mathcal{C}$ to orthogonal vectors with deterministic functions. Therefore, the method proposed in [53] based on the well known orthogonalization method Gram-Schmidt in [54] is deployed. Hence, in the first step of the function a vector of the centroids is fed to the orthogonalization process as follows:

$$\mathcal{C}^{\perp} = \text{Gram-Schmidt}(\mathcal{C}) = \begin{bmatrix} \mathbf{V}^{\perp}_{\mathcal{VN}} \\ \mathbf{V}^{\perp}_{\mathcal{SN}} \\ \mathbf{V}^{\perp}_{\mathcal{FN}} \end{bmatrix} \tag{4.4}$$

where $\mathcal{C}^{\perp}$ is the matrix of Cartesian coordinates orthogonalized vectors. The hyper-spherical coordinates $\mathcal{C}^{\perp}_{*}$ of same orthogonalized vectors, which reveals the angular change from original vectors to the ideal orthogonal vectors, are calculated subsequently using Eq.4.1.

$$\mathcal{C}^{\perp}_{*} = \begin{bmatrix} r^{\perp}_{\mathcal{VN}} & \theta^{\perp}_{1_{\mathcal{VN}}} & \dots & \theta^{\perp}_{n-1_{\mathcal{VN}}} \\ r^{\perp}_{\mathcal{SN}} & \theta^{\perp}_{1_{\mathcal{SN}}} & \dots & \theta^{\perp}_{n-1_{\mathcal{SN}}} \\ r^{\perp}_{\mathcal{FN}} & \theta 1_{\mathcal{FN}}{}^{\perp} & \dots & \theta^{\perp}_{n-1_{\mathcal{FN}}} \end{bmatrix} \tag{4.5}$$

4.4 Spatial Mapping Function

After obtaining the original spherical coordinates ($\mathbf{V}_{\mathcal{VN}},\ \mathbf{V}_{\mathcal{SN}},\ \mathbf{V}_{\mathcal{FN}}$) and orthogonalized spherical coordinates ($\mathbf{V}^{\perp}_{\mathcal{VN}},\ \mathbf{V}^{\perp}_{\mathcal{SN}},\ \mathbf{V}^{\perp}_{\mathcal{FN}}$), the goal is to design a mapping function $\mathbf{F} : \mathbf{R}^n \rightarrow \mathbf{R}^n$ from original coordinates to the orthogonal ones which is equivalent to the ideal cluster topology.

In Gram-Schmidt algorithm, the very first vector in the input vector array serves as a reference vector and remains unchanged in the orthogonal vector array. As a result, $\mathbf{V}_{\mathcal{VN}} =$

$\mathbf{V}^{\perp}_{\mathcal{VN}}$ and equivalently $\mathbf{F}$ can be defined by the following equations:

$$\begin{aligned} \mathbf{F}(\mathbf{V}_{\mathcal{SN}} - \mathbf{V}_{\mathcal{VN}}) = \mathbf{V}^{\perp}_{\mathcal{SN}} - \mathbf{V}^{\perp}_{\mathcal{VN}} \quad &= \mathbf{V}^{\perp}_{\mathcal{SN}} - \mathbf{V}_{\mathcal{VN}} \\ \mathbf{F}(\mathbf{V}_{\mathcal{FN}} - \mathbf{V}_{\mathcal{VN}}) = \mathbf{V}^{\perp}_{\mathcal{FN}} - \mathbf{V}^{\perp}_{\mathcal{VN}} \quad &= \mathbf{V}^{\perp}_{\mathcal{FN}} - \mathbf{V}_{\mathcal{VN}} \end{aligned} \tag{4.6}$$

Furthermore, since orthogonality is Independent to radius r, $\mathbf{F}$ only needs to apply to $n-1$ dimensions which includes all angles $(\theta_1, \ldots, \theta_{n-1})$ and coordinate r remains unchanged before and after mapping. Consequently, determination of $\mathbf{F}$ can be decomposed to the determination of $n-1$ functions: $f_i : \mathbf{R} \to \mathbf{R},\ i = 1 \ldots n-1$ with constraints in Eq.4.6 and boundary constraints. Note the $\mathbf{V}_{\mathcal{SN}} - \mathbf{V}_{\mathcal{VN}}$ as $\Delta_{\mathcal{SV}}$ and its ith angular dimension of as $\delta_{i_{\mathcal{SV}}}$. Same notation is applied on $\mathbf{V}_{\mathcal{FN}} - \mathbf{V}_{\mathcal{VN}}$. Hence for each angular dimension i, f_i is determined by $(\delta_{i_{\mathcal{SV}}}, \delta^{\perp}_{i_{\mathcal{SV}}})$ and $(\delta_{i_{\mathcal{FV}}}, \delta^{\perp}_{i_{\mathcal{FV}}})$, together with two boundary constraints In order to maintain the simplicity and linearity of the mapping function, f_i needs to be continuous and monotonic. For this purpose, periodicity of angular dimension is used to determine the boundary constrains and the problem is transformed into a curve fitting one. For example, if linking $(\delta_{i_{\mathcal{SV}}}, \delta^{\perp}_{i_{\mathcal{SV}}})$ and $(\delta_{i_{\mathcal{FV}}}, \delta^{\perp}_{i_{\mathcal{FV}}})$ results in a monotonically decreasing line, the boundary constrains would be $(\pi, 0)$ and $(0, \pi)$. Conversely, if it results in a monotonically increasing line, the boundary constrains would be $(0, 0)$ and (π, π). Same rules applies on the last angular dimension where period is 2π instead of pi.

The simplest candidate function for f_i, which connect two boundary points and two target points in the 2-D plane would be a linear spline as shown in Fig.4.2

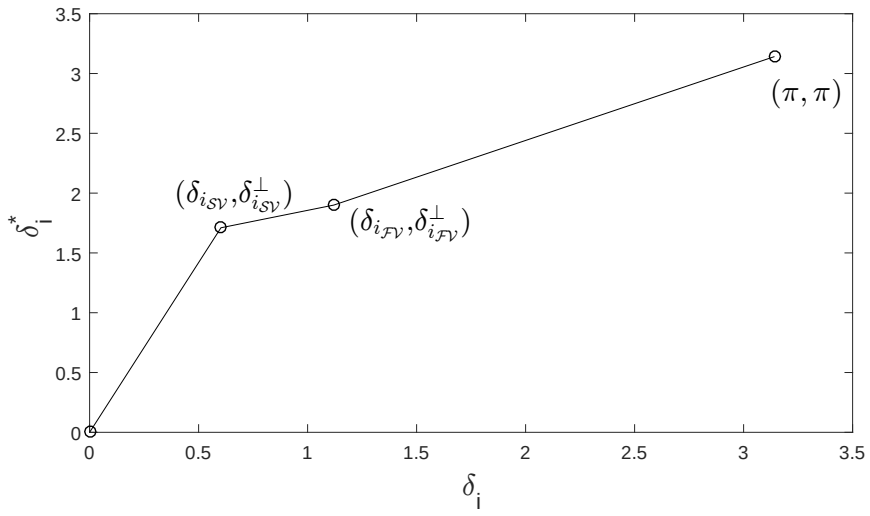

Figure 4.2: The simple mapping function

4.5 Optimized Mapping Function

The mapping function in Fig.4.2 demonstrated monotonicity and continuity, which is a part of targeted characteristics as an ideal mapping function for f_i. However there exists some drawbacks in the simple mapping function. First of all, the function is not differentiable at two target points $(\delta_{i_{\mathcal{SV}}}, \delta^{\perp}_{i_{\mathcal{SV}}})$ and $(\delta_{i_{\mathcal{FV}}}, \delta^{\perp}_{i_{\mathcal{FV}}})$, which will lead to cluster deformation. Secondly, since the mapping function is applied on angular dimension, linearity of each spline in the function will bring on non-convex clusters in Cartesian coordinates after mapping. In order to avoid deformations and preserve within cluster geometry, it is required to have a region for each cluster centroids within which the function maps the data to a equal or even smaller range in Cartesian coordinates. In the other word, the samples close to the centroids has to be close to the same corresponding centroids in the mapped space. In order to avoid deformations, an optimized mapping function is proposed.

To compromise three constrains, a function which satisfies the following mathematical conditions is studies:

- mapping function derivatives around centroids 0 and $\delta_{i_{\mathcal{SV}}}$ and $\delta_{i_{\mathcal{FV}}}$ are small
- mapping function derivatives between two centroids large
- quasi differentiable at all points

Therefore, we proposed the basis function p satisfies the aforementioned constrains. The function p is composed of two parts: $h(x)$ and $g(x)$. The boundaries that each function activates are defined as target points ($\delta_{i_{\mathcal{SV}}}$,$\delta^{\perp}_{i_{\mathcal{SV}}}$) and ($\delta_{i_{\mathcal{FV}}}$,$\delta^{\perp}_{i_{\mathcal{FV}}}$) along with the mid-points between target points. The lower boundary mid points are noted as (γ,$\gamma^{\perp}$). The upper boundary mid points are noted as (ϵ,$\epsilon^{\perp}$). To ensure the continuity of mapping function f, we have

$$(\epsilon_{i_{\mathcal{SV}}}, \epsilon^{\perp}_{i_{\mathcal{SV}}}) = (\gamma_{i_{\mathcal{FV}}}, \gamma^{\perp}_{i_{\mathcal{FV}}}) \quad = (\frac{\delta_{i_{\mathcal{SV}}} + \delta_{i_{\mathcal{FV}}}}{2}, \frac{\delta^{\perp}_{i_{\mathcal{SV}}} + \delta^{\perp}_{i_{\mathcal{FV}}}}{2}) \tag{4.7}$$

Therefore, the two piecewise functions $h(x)$ and $g(x)$ are defined as follows:

$$\begin{aligned} K_h &= \frac{\epsilon^{\perp} - \delta^{\perp}}{e^{\alpha(\epsilon-\delta,0)^+} - 1} \\ h(x) &= K_h[e^{\alpha(x-\delta,0)^+} - 1] + \delta^{\perp} \end{aligned} \tag{4.8}$$

$$\begin{aligned} K_g &= \frac{\gamma^{\perp} - \delta^{\perp}}{e^{\alpha(-\gamma+\delta,0)^+} - 1} \\ g(x) &= K_g[e^{\alpha(\delta-x,0)^+} - 1] + \delta^{\perp} \end{aligned} \tag{4.9}$$

Applying p on two target points ($\delta_{i_{\mathcal{SV}}}$,$\delta^{\perp}_{i_{\mathcal{SV}}}$) and ($\delta_{i_{\mathcal{FV}}}$,$\delta^{\perp}_{i_{\mathcal{FV}}}$) will result in a smooth curve as

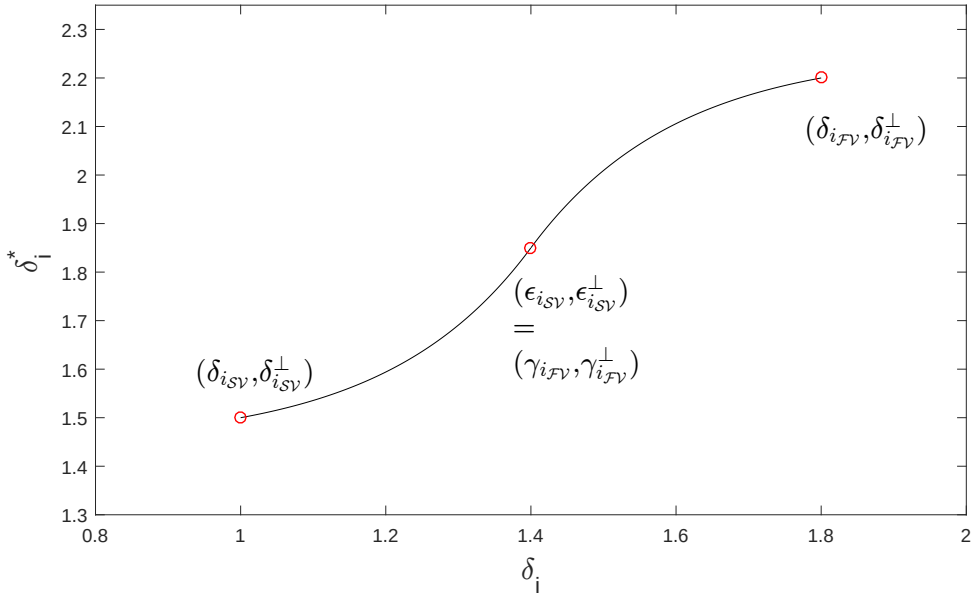

Figure 4.3: Optimized Piecewise Interpolate Function p

shown in Fig.4.3

And applying the piecewise interpolate function p at all four points: $(\delta_{i_{\mathcal{SV}}},\delta^{\perp}_{i_{\mathcal{SV}}})$, $(\delta_{i_{\mathcal{FV}}},\delta^{\perp}_{i_{\mathcal{FV}}})$ and two boundary points, the final result is shown in Fig.4.4

4.6 Experimental Results

In this section, the performance of proposed method is evaluated according to two aspects. We will first analyze the classification performance of the system and demonstrate the comparison with other representative ECG classifiers. Furthermore, the classification results are partitioned into two sets: red alarms generated by Global Classifier and final results by combining yellow and red alarms. Impacts of Personal Classifier on final results will be studied. Finally the prediction performance which is typical in the proposed system is evaluated.

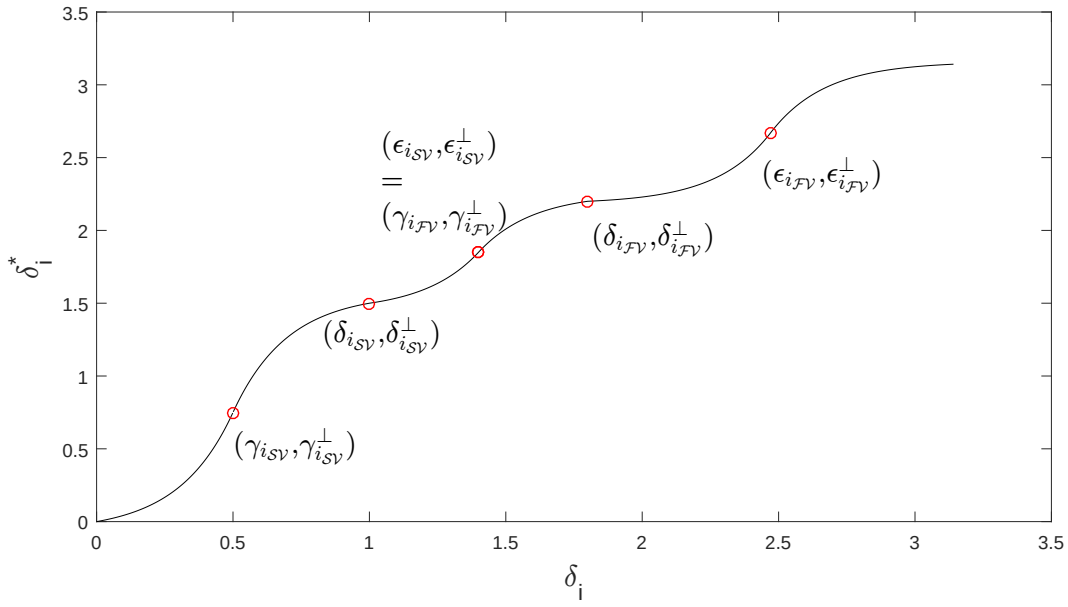

Figure 4.4: Optimized Mapping Function f

4.6.1 Classification Performance

The experimental results are evaluated with the classification performance of 4 AAMI ECG classes using the test subset of MITBIH Arrhythmia DS2. Originally DS2 contains 15357 samples after feature extraction. While training Personal Classifier, the first 20% of total normal samples serve as initialization samples for Personal Dynamic Normal Cluster. Therefore, all samples before the last initialization normal sample should be excluded from test set for each records. Consequently the actual test set contains 12414 samples in total consisting of 10105 type-N, 1702 type-V, 508 type-S and 99 type-F samples.

To present the result, we select weighted k-Nearest Neighbors where $k = 10$ as Global Classifier as it's comparatively simple and representative among low complexity models. Parameter α in the deviation detection module is set to 1 for test purpose.

Table 4.1 summarized the cumulated confusion matrix for all records in the test set. In order to compare the result of Global Classifier and combined result, the sample numbers

Table 4.1: Cumulated Confusion Matrix for All Records in DS2

		Ground Truth			
Result		N	V	S	F
	N	9255(10076)	21(38)	72(90)	1(5)
	V	657(22)	1678(1663)	8(2)	9(7)
	S	71(6)	3(1)	417(416)	0(0)
	F	122(1)	0(0)	11(0)	89(87)

are presented in the following format: *combined*(*globaled*). In order to measure classification performance, we adopted four metrics proposed in [12]: accuracy(Ac), sensitivity(Se), specificity(Sp) and precision(Pr). All four metrics are calculated based on true positive TP, false positive FP, false negative FN and true negative TN in a binary confusion matrix. Therefore all four metrics are calculated for each class by converting the 4x4 matrix to a 2x2 matrix.

While cumulated classification results are demonstrated in Table 4.1, the robustness of the proposed method should be evaluated based on the performance variation over 22 test records in DS2. Hence medians and IQRs(interquartile range) for each metric and each class are included in Table 4.2 to represent the robustness of proposed methods. The lower variation between performances measured on different tapes, the more robust the system is. In Table 4.2, we observe that among all abnormal classes, the proposed algorithms demonstrated stable performance on class V and high performance but less stable on class V and F.

As MITDB is widely used to verify ECG classifier performance, we compared the proposed system with five significant methods proposed in literature. According to AAMI standards,

Table 4.2: Classification Performance and Within-Set Variation of Proposed System

statistics	N			V			S			F		
	Ac	*Se*	*Sp*	*Ac*	*Se*	*Sp*	*Ac*	*Se*	*Sp*	*Ac*	*Se*	*Sp*
cumulated	92.4	91.59	95.93	94.38	98.59	93.71	98.67	82.09	99.38	98.85	89.9	98.92
median	94.45	92.21	95.42	96.17	99.55	95.71	99.38	80.65	99.84	99.11	90.91	99.11
IQR	6.33	10.08	11.91	5.17	1.64	8.62	1.76	19.35	0.61	1.58	23.33	1.49

Table 4.3: V and S classification performance compared with five algorithms in literature using 11 common records in MITDB

Methods	VEB			SVEB		
	Ac	Se	Sp	Ac	Se	Sp
Proposed	96.6	98.2	92.4	98.63	88.89	99.41
Hu *et al.* [12]	94.8	78.9	96.8	N/A	N/A	N/A
de Chazal *et al.* [9]	96.4	77.5	N/A	N/A	N/A	N/A
Jiang and Kong [15]	98.8	78.9	96.8	97.5	74.9	98.8
Ince *et al.* [16]	97.9	90.3	98.8	96.1	81.8	98.5
Kiranyaz *et al.* [17]	98.9	95.9	99.4	96.4	68.8	99.5

ECG classifier performance should be evaluated over the binary classification performance of Ventricular (VEB) versus non-VEB types and Supraventricular(SVEB) versus non-SVEB types. For methods proposed in literature, same evaluation metrics are deployed on records from MITDB. To standardize the metrics, we select 11 common ECG records from all 5 methods and compared the median of each classification metrics over these 11 records. The comparison results are demonstrated in Table.4.3. Generally speaking, the proposed method shows higher sensitivity for both VEB and SVEB. Especially for SVEB, the proposed method has advantage on all three metrics over other five methods in the literature.

4.6.2 Prediction Performance

As an important characteristic of the proposed methods, yellow alarms triggered by Personal Classifier after feature space reshaping indicate a higher probability of observing subsequent abnormalities. In order to verify this functionality, all beats following a yellow alarm is investigated for each yellow alarm. Abnormality type which occurs the earliest within the window is recorded. Similar to confusion matrix for classification evaluation, the performance of prediction can be summarized by a confusion matrix with the 3 abnormal types. Probabilities of observing an certain type of abnormal beats after a yellow alarm is calculated using the

Table 4.4: predictive probability versus prior probability without windowing

		# of predicted ground truth			% of predicted ground truth		
		V	S	F	V	S	F
yellow alarm	V	467	122	14	**77.45**	20.23	2.32
	S	36	15	0	70.59	**28.41**	0
	F	40	60	5	38.10	57.14	**4.76**
total		543	197	19	**71.54**	**25.96**	**2.50**

prediction confusion matrix and compared to the prior probability of observing the same type of abnormality. This process is formulated in the following two equations:

$$P(\hat{y}_{k+i} = X_r | \hat{y}_k = X_y) = \frac{\#\text{ of } y_{k+i} = X \text{ after } \hat{y}_k = X_y}{\#\text{ of true alarms after } \hat{y}_k = X_y}$$
$$P(\hat{y}_{k+i} = X_r) = \frac{\#\text{ of true alarm of type } X \ (y_k = X)}{\#\text{ of all true alarms}} \tag{4.10}$$

The capacity of predicting each type of abnormalities is evaluated by comparing $P(\hat{y}_{k+i} = X_r | \hat{y}_k = X_y)$ and $P(\hat{y}_{k+i} = X_r)$. A shown in Table.4.4, the probability of observing a certain type of abnormalities after a yellow alarm are higher than its prior for each type of abnormalities. For example, without knowing the type of a yellow alarm, the probability of observing a type V sample is 71.54% while the probability of observing a type V sample given a type V yellow alarm was triggered is 77.45%. The improvement are consistent among all three types of abnormalities but the system has stronger predicting capacity for type S.

In order to study the time window in which real abnormality occurs after yellow alarms, we also studied a window of 10 consecutive samples following a yellow alarm. Similarly, prior and posterior probabilities are compared to evaluate the performance as shown in Table.4.5.

Compared with the result without windowing, the predicting performance within 10 beats window shows that the proposed algorithm can better predict the occurrence of abnormalities

Table 4.5: predictive probability versus prior probability within 10 beats' window

		# of predicted ground truth			% of predicted ground truth		
		V	S	F	V	S	F
yellow alarm	V	290	85	12	**74.94**	21.96	3.10
	S	22	13	0	62.86	**37.14**	0
	F	29	37	6	40.28	51.39	**8.33**
total		341	135	18	**69.03**	**27.32**	**3.64**

in a centrain time window. Especially for type S, the probability of observing a type S sample within 10 beats after a yellow alarm is 27.32% while given that the yellow alarm is type S, the posterior probability is raise to 37.14%. With almost 10% increase, it's proved that the yellow alarm types are informative. The results shows that same improvements are made within the 10-sample window as well. In general, the predicting performance are promising, indicating the efficiency of personal classifier and deviation analysis.

Chapter 5

Conclusions And Future works

5.1 Conclusions

In this thesis, we a patient-adaptable ECG classification framework. The system has a two-staged hierarchical classifier structure including Global Classifier and Personal Classifier. While Global Classifier is designed to filter the signal with severe distortion and abnormal waveforms by triggering red alarms and pass other samples to the deviation detection stage. In this stage, the personal dynamic normal cluster is constructed and used to specify the normal range for each patient. By comparing the current sample and personalized normal range, this module decides if a yellow alarm will be triggered to provide predictive information about upcoming abnormalities. If a sample is detected with deviation towards abnormal clusters, it will be passed to the Personal Classifier and labeled as one of the three abnormal types. Whereas samples without deviation are further feed back to personal dynamic normal cluster to update the classification system about the newest personal normal range.

In Chapter 3, a kernel based nonlinear transformation is proposed to address the problem

of cluster topology in original feature space. Inspired by Support Vector Machine, kernel functions are deployed in this method as a spatial transformation function. The target topology is formulated as two objective functions so that by tuning the parameters in kernel function, the system is able to select the best transformation for the following predicting stage. This non-convex multi-objective optimization is solved with Multi-Objective Particle Swarm Optimization. In order to validate improvement by using high order kernel function, we compared the Pareto front generated with linear combination of original features and mapped high order features with polynomial kernel. The result verifies that applying high order kernel function allows more degree of freedom so that the topology can be further optimized according to objective functions. Having this concept proved, we applied this method on MITDB test data and obtained similar sensitivity and specificity as proposed in the literatures. More importantly, the predicting capacity of yellow alarms are analyzed. The performance is quantified by comparing prior probability and posterior probability giving the types of yellows alarm. The comparison result shows that a promising improvement has been made by applying the nonlinear transformation.

While the method in Chapter 3 demonstrated capacity of predicting upcoming abnormalities, it's challenging to interpret the mechanisms of the systems and thus hindering the generalization of predictive warning to other applications of biomedical signals. Therefore, the main object of Chapter 4 is developing a classification system with abnormality predicting capacity based on spatial topology studied in Chapter 3. In Chapter 4, we proposed a novel spatial transformation specifically designed to reshape the feature space according to angles between cluster center. In this method, between cluster cosine distance are optimized through orthogonalization in spherical coordinate space and within cluster variance is reduced by a mapping function which is fitted piecewise with a basis function. The basis function proposed in the chapter has the feature of saturating at the boundaries, similar

to sigmoid function but more flexible. An advantage of deploying such basis function is that the cluster geometry may be preserved after spatial transformation. With this novel module integrated in the patient-adaptable classification framework, the performance of this system is evaluated through classification and prediction results on the test set data. The classification results show that by triggering yellow alarm through this method, specificity of abnormal types is improved. Especially for Supraventricular types, the proposed system performs better than all 5 methods in the literature. The same conclusion holds for prediction performance. Compared to the method proposed in Chapter 3, this method improves predicting capacity for all abnormal classes and the most significant improvement is for type S. Moreover, we also studied the time delay of real abnormalities following a yellow alarm. It's proved that most of the real abnormalities occurs within 10 beat after a yellow alarm. Generally speaking, the system is proved to be efficient both in classification and prediction.

5.2 Future works

In this research, we focused on two challenges of ECG classification, namely, inter-patient variation and anomaly prediction. The framework of patient-adaptable classier includes both features. The methods of improving prediction accuracy are proposed and studied. While the result shows the efficiency of designed system, further improvements can be made through research. The following tasks can be resolved as a continuation of this research:

- Research on other kernel functions which is potentially a better transformation for spatial topology optimization.
- Investigate on the deterministic solution for the objective functions proposed in Chapter 3.

- Assess the performance of proposed spatial transformation on other biomedical signals with similar characters as ECG signal
- Improve the deterministic mapping function in Chapter 4 by including the variance of individual clusters into function parameters

Bibliography

[1] C. J. Murray and A. D. Lopez, "Measuring the global burden of disease," *New England Journal of Medicine*, vol. 369, no. 5, pp. 448–457, 2013.

[2] D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, T. B. Ferguson, E. Ford, K. Furie, C. Gillespie, *et al.*, "Heart disease and stroke statistics2010 update," *Circulation*, vol. 121, no. 7, pp. e46–e215, 2010.

[3] W. H. Organization, "Cardiovascular diseases (cvds)," 2017.

[4] S. C. Smith, R. Jackson, T. A. Pearson, V. Fuster, S. Yusuf, O. Faergeman, D. A. Wood, M. Alderman, J. Horgan, P. Home, *et al.*, "Principles for national and regional guidelines on cardiovascular disease prevention: a scientific statement from the world heart and stroke forum," *Circulation*, vol. 109, no. 25, pp. 3112–3121, 2004.

[5] E. Besterman and R. Creese, "Waller–pioneer of electrocardiography.," *British Heart Journal*, vol. 42, no. 1, p. 61, 1979.

[6] B. E. Kreger, L. A. Cupples, and W. B. Kannel, "The electrocardiogram in prediction of sudden death: Framingham study experience," *American heart journal*, vol. 113, no. 2, pp. 377–382, 1987.

[7] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ecg complexes using hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, 2000.

[8] G. K. Prasad and J. Sahambi, "Classification of ecg arrhythmias using multi-resolution analysis and neural networks," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 1, pp. 227–231, IEEE, 2003.

[9] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1196–1206, July 2004.

[10] R. Ceylan, Y. Özbay, and B. Karlik, "A novel approach for classification of ecg arrhythmias: Type-2 fuzzy clustering neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6721–6726, 2009.

[11] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE transactions on biomedical engineering*, vol. 51, no. 4, pp. 582–589, 2004.

[12] H. H. Yu, P. S., and J. T. W., "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, 1997.

[13] P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 2535–2543, Dec 2006.

[14] M. Llamedo and J. P. Martínez, "An automatic patient-adapted ecg heartbeat classifier allowing expert assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312–2320, 2012.

[15] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1750–1761, 2007.

[16] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

[17] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[18] P. W. Wilson, R. B. DAgostino, D. Levy, A. M. Belanger, H. Silbershatz, and W. B. Kannel, "Prediction of coronary heart disease using risk factor categories," *Circulation*, vol. 97, no. 18, pp. 1837–1847, 1998.

[19] M. A. Whooley, P. de Jonge, E. Vittinghoff, C. Otte, R. Moos, R. M. Carney, S. Ali, S. Dowray, B. Na, M. D. Feldman, *et al.*, "Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease," *Jama*, vol. 300, no. 20, pp. 2379–2388, 2008.

[20] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ecg signals using machine learning techniques: A survey," in *Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in*, pp. 714–721, IEEE, 2015.

[21] S. Kiranyaz, T. Ince, and M. Gabbouj, "Personalized monitoring and advance warning system for cardiac arrhythmias," *Scientific Reports*, vol. 7, no. 1, p. 9270, 2017.

[22] L. S. Green, R. L. Lux, C. W. Haws, R. R. Williams, S. C. Hunt, and M. J. Burgess, "Effects of age, sex, and body habitus on QRS and ST-T potential maps of 1100 normal subjects.," *Circulation*, vol. 71, no. 2, pp. 244–253, 1985.

[23] R. Hoekema, G. J. H. Uijen, and A. van Oosterom, "Geometrical aspects of the interindividual variability of multilead ecg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 551–559, May 2001.

[24] A. Houghton and D. Gray, *Making sense of the ECG: a hands-on guide*. CRC Press, 2014.

[25] G. A. Ng, "Treating patients with ventricular ectopic beats," *Heart*, vol. 92, no. 11, pp. 1707–1712, 2006.

[26] A.-A. EC57, "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms," *Association for the Advancement of Medical Instrumentation, Arlington, VA*, 1998.

[27] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[28] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[29] J. Chen and A. Razi, "A predictive framework for ecg signal processing using controlled nonlinear transformation," in *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on*, pp. 161–165, IEEE, 2018.

[30] J. Chen, H. Peng, and A. Razi, "Remote ECG monitoring kit to predict patient-specific heart abnormalities," *Journal of Systemics, Cybernetics and Informatics*, vol. 15, no. 4, pp. 82–89, 2017.

[31] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ecg signal denoising," *Digital signal processing*, vol. 16, no. 3, pp. 275–287, 2006.

[32] N. V. Thakor, J. G. Webster, and W. J. Tompkins, "Estimation of qrs complex power spectra for design of a qrs filter," *IEEE Transactions on biomedical engineering*, no. 11, pp. 702–706, 1984.

[33] Y. Lian and P. C. Ho, "Ecg noise reduction using multiplier-free fir digital filters," in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, vol. 3, pp. 2198–2201, IEEE, 2004.

[34] Y.-W. Bai, W.-Y. Chu, C.-Y. Chen, Y.-T. Lee, Y.-C. Tsai, and C.-H. Tsai, "Adjustable 60hz noise reduction by a notch filter for ecg signals," in *Instrumentation and Measurement Technology Conference, 2004. IMTC 04. Proceedings of the 21st IEEE*, vol. 3, pp. 1706–1711, IEEE, 2004.

[35] O. Sayadi* and M. B. Shamsollahi, "Ecg denoising and compression using a modified extended kalman filter structure," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 2240–2248, Sept 2008.

[36] K. Park, K. Lee, and H. Yoon, "Application of a wavelet adaptive filter to minimise distortion of the st-segment," *Medical and Biological Engineering and Computing*, vol. 36, no. 5, pp. 581–586, 1998.

[37] N. Nikolaev, Z. Nikolov, A. Gotchev, and K. Egiazarian, "Wavelet domain wiener filtering for ecg denoising using improved signal estimate," in *Acoustics, Speech, and Signal*

Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on, vol. 6, pp. 3578–3581, IEEE, 2000.

[38] S. Poungponsri and X.-H. Yu, "An adaptive filtering approach for electrocardiogram (ecg) signal noise reduction using neural networks," *Neurocomputing*, vol. 117, pp. 206–213, 2013.

[39] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "Ecg beat detection using filter banks," *IEEE transactions on biomedical engineering*, vol. 46, no. 2, pp. 192–202, 1999.

[40] D. Sadhukhan and M. Mitra, "R-peak detection algorithm for ecg using double difference and rr interval processing," *Procedia Technology*, vol. 4, pp. 873–877, 2012.

[41] S. Mehta and N. Lingayat, "Svm-based algorithm for recognition of qrs complexes in electrocardiogram," *IRBM*, vol. 29, no. 5, pp. 310–317, 2008.

[42] R. V. Andreão, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *IEEE Transactions on Biomedical engineering*, vol. 53, no. 8, pp. 1541–1549, 2006.

[43] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ecg delineator: evaluation on standard databases," *IEEE transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.

[44] S. Banerjee, R. Gupta, and M. Mitra, "Delineation of ecg characteristic features using multiresolution wavelet analysis method," *Measurement*, vol. 45, no. 3, pp. 474–487, 2012.

[45] Z. Zidelmal, A. Amirou, M. Adnane, and A. Belouchrani, "QRS detection based on wavelet coefficients," *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 490–496, 2012.

[46] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[47] B. Schölkopf, C. J. Burges, and A. J. Smola, *Advances in kernel methods: support vector learning*. MIT press, 1999.

[48] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in computational mathematics*, vol. 13, no. 1, p. 1, 2000.

[49] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[50] C. A. Coello Coello, "Mopso: A proposal for multiple objective particle swarm optimization," *Proc. Congr. Evolutionary Computation (CEC'2002), Honolulu, HI, 5*, vol. 1, pp. 1051–1056, 2002.

[51] J. E. Alvarez-Benitez, R. M. Everson, and J. E. Fieldsend, "A mopso algorithm based exclusively on pareto dominance concepts," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 459–473, Springer, 2005.

[52] L. Blumenson, "A derivation of n-dimensional spherical coordinates," *The American Mathematical Monthly*, vol. 67, no. 1, pp. 63–66, 1960.

[53] G. W. Stewart, *Matrix algorithms volume 1: Basic decompositions*, vol. 2. Society for Industrial and Applied Mathematics, 1998.

[54] G. Arfken, "Gram-schmidt orthogonalization," *Mathematical methods for physicists*, vol. 3, pp. 516–520, 1985.