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

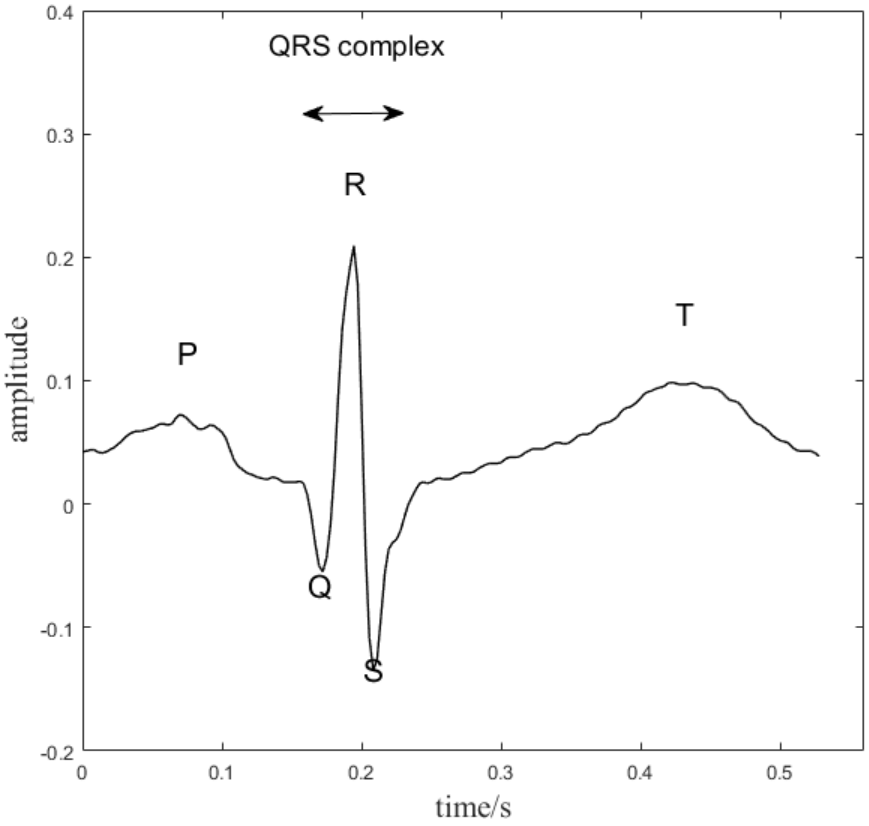

Figure 1.1: A typical cardiac cycle in ECG signal with five characteristic waves

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

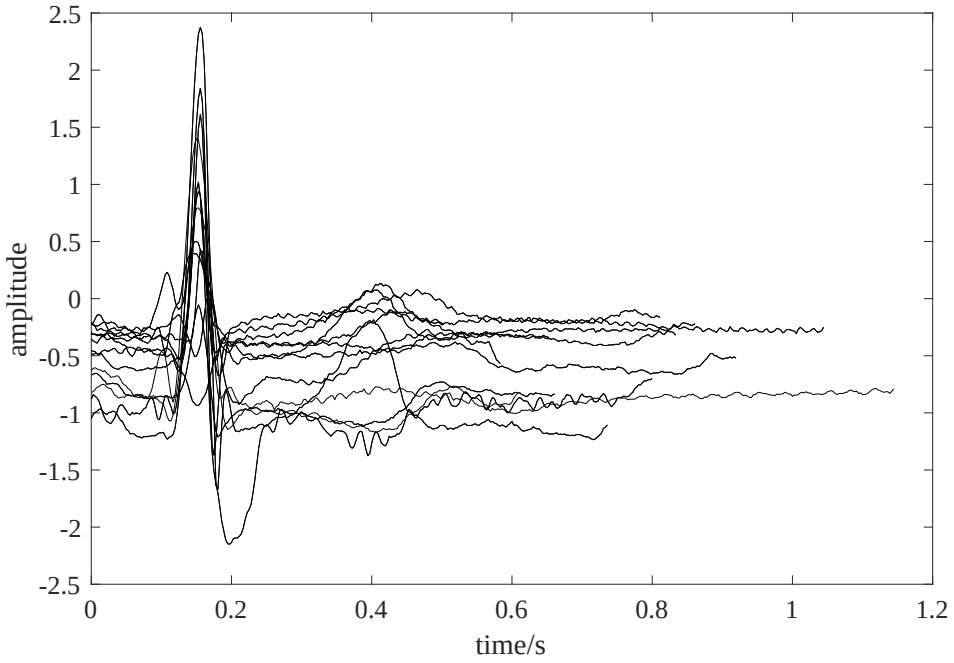

Figure 1.2: ECG signals of normal heartbeat from 15 different records in MIT-BIH reflect the inter-patient variability of ECG signal

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.

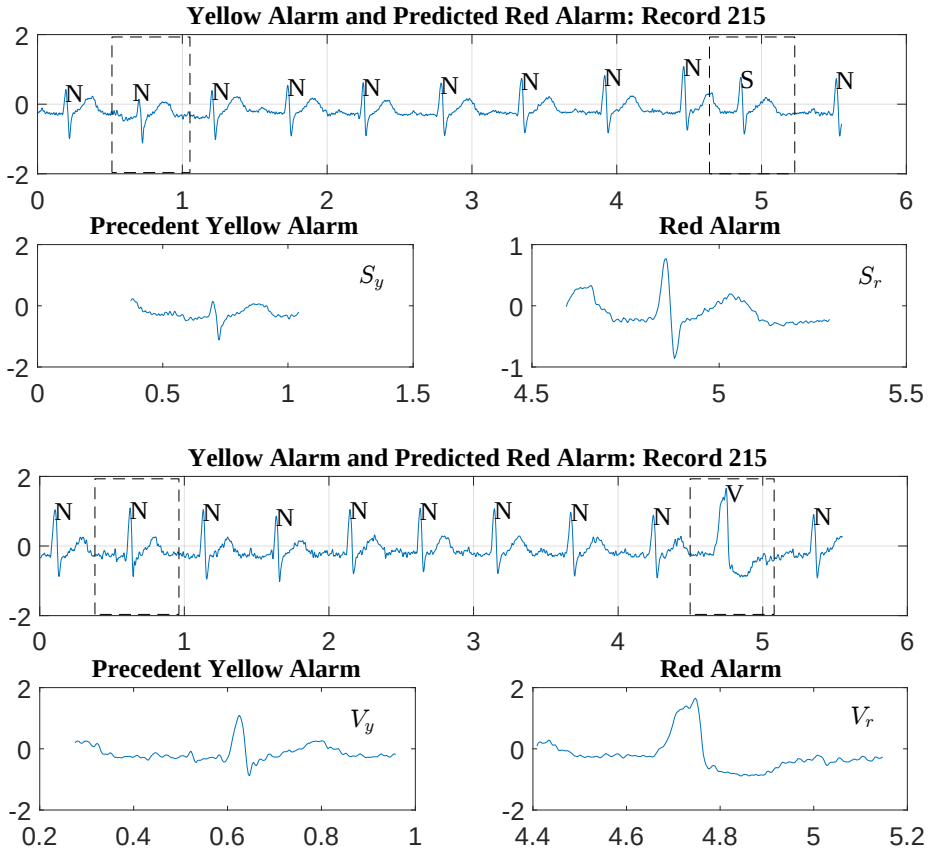

Figure 1.3: A potential latent abnormal status(yellow alarm) predicts an upcoming abnormality of the same type.

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.

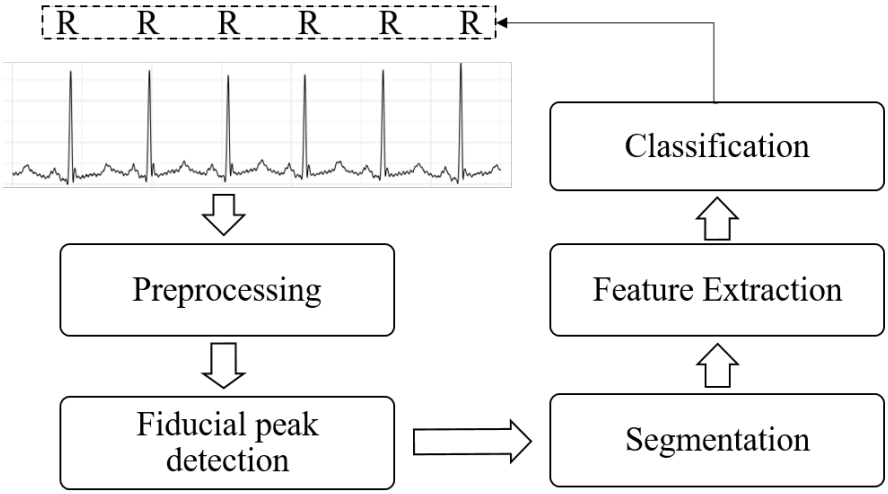

Figure 1.4: General structure of ECG analysis system

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.

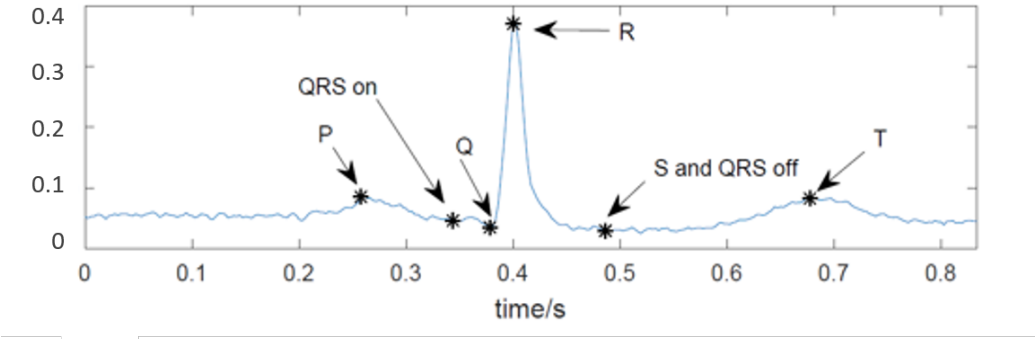

Figure 1.5: Fiducial peaks within one cardiac cycle

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 processing and analysis have been a controversial research topic. Studies carried out by scholars have proved that the development of automatic ECG analysis is conducive to the timely detection and therapeutic intervention of heart disease. However some major challenges need to be resolved before apply automatic ECG analysis as a highly reliable and fully automatic electrocardiogram processing system in clinical diagnosis. One of the most typical challenge is the inter-patient variation of ECG waveforms, which leads to inconsistent performance of ECG classification system. In this chapter, the basic framework of the proposed Patient-Adaptable ECG Classification will be discussed.

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 which represent

heartbeat and processed by designed algorithms. The first sections in this chapter focus on the data preparation stage, which includes four steps: signal preprocessing, delineation and segmentation, and feature extraction. Following the data preparation Section 2.2 elaborate on the framework of a two staged hierarchical classifier. The classification system is patient adaptable by capturing the normal range for each individual. More specifically, in Section 2.3, the personal dynamic normal cluster method is discussed. One feature of this method is that the cluster can dynamically adapt to patient's ECG waveform change. In many application scenarios, the physicians need to monitor long-term real-time heart activity. The dynamic adjusting system is able to address the issue of intra-patient signal variation as well

2.2 ECG Signal Processing

2.2.1 Preprocessing

Biomedical signal, such as ECG signal, is often present accompanied with statistical features which change over time. Therefore, the traditional Fourier transform is not suitable for this type of non-stationary signals since it's unable to capture time-varying statistics in the signal. Wavelet decomposition solves this problem by scaling and translating mother wavelet to constitute its basis functions. Given a time series, wavelet decomposition decompose the signal into linear combinations of the aforementioned basis functions. Thus, the basis functions with large scale is wider than those with small scale and consequently correspond to low frequency component of the signal. Similarly, the coefficients of decomposition correspond to high frequency when scale is small. As such, the algorithm can extract time and frequency features through wavelet decomposition.

Hence, in this work, wavelet analysis is applied to signals in MITDB with a sampling fre-

quency of 360 Hz. With this sampling frequency, the decomposition coefficients and its frequency components are deduced in Fig.2.1.

Low frequency noise or baseline wander between 0.15 to 1 Hz, cause by respiration and body movement, can be removed by deducting approximation coefficient of level 8 (A_8) from the signal. Since the power of ECG signal is mainly located in the frequency band from 1 to 40 Hz, some high frequency noise including electromyogram induced noise and mechanical forces acting on the electrodes can be removed by discarding the detail coefficient of level 1 (D_1). Daubechies wavelet of order 8 ($db8$) is selected as mother wavelet for denoising stage in this work for its optimal performance [31].

2.2.2 Segmentation

Among a continuous ECG signal, the characteristics of a single cardiac cycle in the signal is generally considered as a sample in most machine learning applications. Therefore, it is necessary to segment ECG signal before feature extraction. Most of the methods in literature used wavelet analysis to detect the highest peak R wave in a cardiac cycle and then detect and segment it by the amplitude-frequency characteristics of other waves. In this study, the R feature peak segmentation method used in most of the literature is adopted. In Section 1.4.2 it was mentioned that 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 in the QRS complex. The QRS complex also contains important information that reflects the arrhythmia category [45]. Accurate detection of QRS complexes 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 information can be extracted from the detail

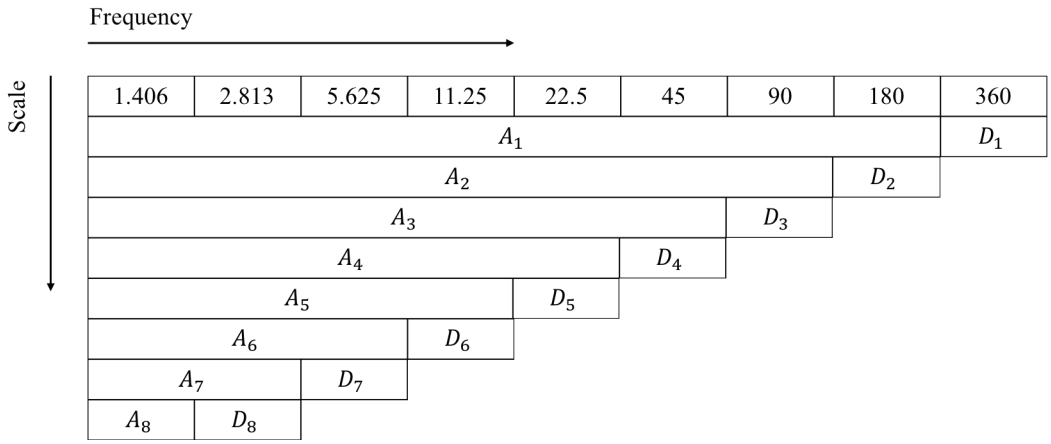

Figure 2.1: Frequency band of wavelet decomposition coefficient for MITDB signals

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 recombination 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 the algorithm suggested by [45]. As 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, this work used a sliding window with a width of 160ms to detect the peaks in the QRS_DET. The window's step size was set to 200ms, given that the time lag between two adjacent heartbeat cycles does not exceed 200ms. Figure.2.3shows the waveform of QRS_DET and the corresponding window width of 160ms.

The false detection peaks are eliminated within a 160ms time window Fig.2.3

The T and P waves are outside the QRS window. By searching in a region from the start of a QRS to the end of its neighboring QRS, the P-wave is located in the highest positive peak in this region. In the same way, the position of the T wave is the position of the largest

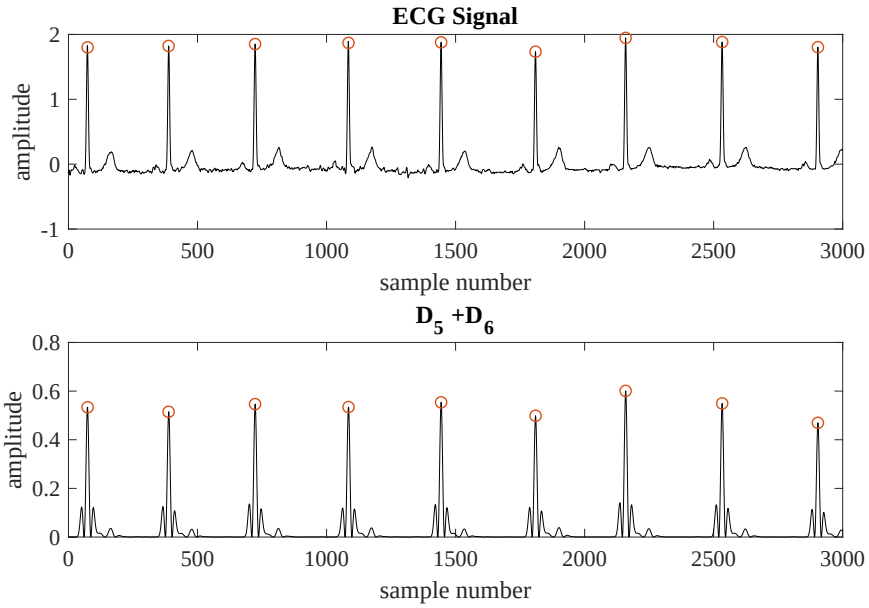

Figure 2.2: QRS detection with wavelet coefficients of level 5 and level 6

positive wave peak in the area from the end of the QRS to the next start of the next QRS. To sum up, the ECG signal in one heartbeat segment can be described by 7 feature points: P, QRS on, Q, R, S, QRS off and T.

Finally, using the location of the R wave, a segment of the ECG signal was divided into cardiac cycles. The starting position of a cardiac cycle is defined as the position 2/3 between the R wave of the cycle and the R wave of the previous cycle. The end position is defined as 2/3 of the R wave between the cycle R and the R wave of the following cycle position. The distance (RR) between each two adjacent R waves is h. The advantage of this method is that the computational complexity is low, and the scope of application is wide and it is independent of individual patient differences.

However the quality of ECG signals provided by most of the portable ECG measuring instruments is very unstable. Signals transmitted through wireless communication systems will exhibit various unstable waveforms. Moreover, the objective of this work is predictive mod-

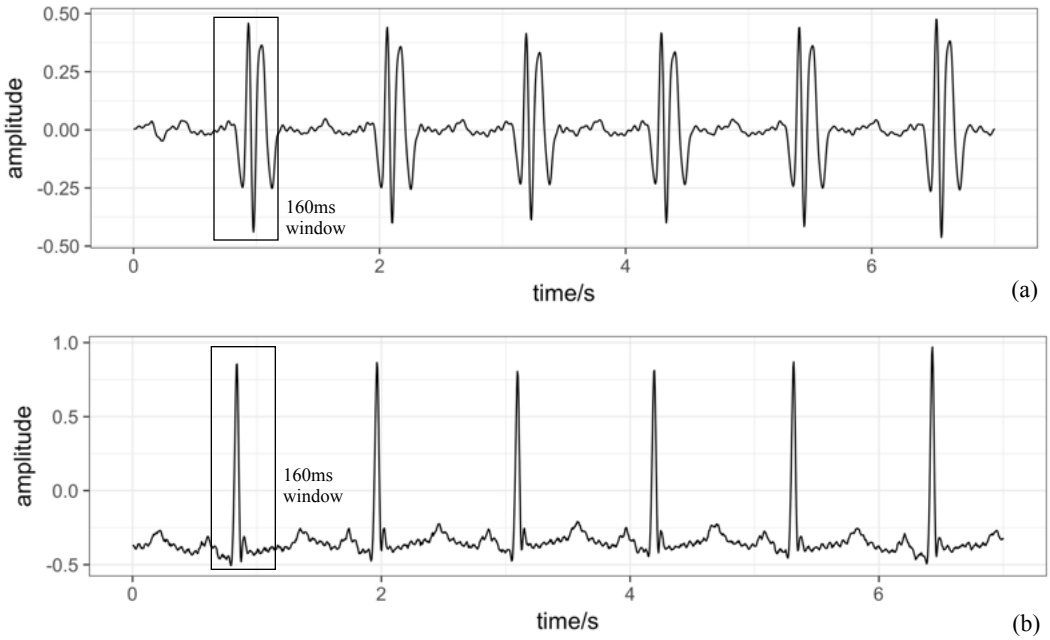

Figure 2.3: Window for detecting R peaks within QRS complexes

eling, thus by including statistical features capturing change between consecutive beats, the system obtains information regarding the following beats. Taking these facts into account, this work performed subsequent feature extraction and analysis combined three consecutive cardiac cycles (i.e. one representative segment). In the subsequent analysis, an segment is considered as a sample of data as shown in Fig.2.4.

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].

For each segment, a new annotation is generated by integrating all annotations of beats within the segment. Unless all beats are annotated as N within the segment, this segment will be annotated as the unique abnormality type within it. If two different abnormality types present within the same segment, this segment would be considered as transient and discarded. After segmentation and annotation generation, the total number of samples in training and test set is summarized in Table.2.1

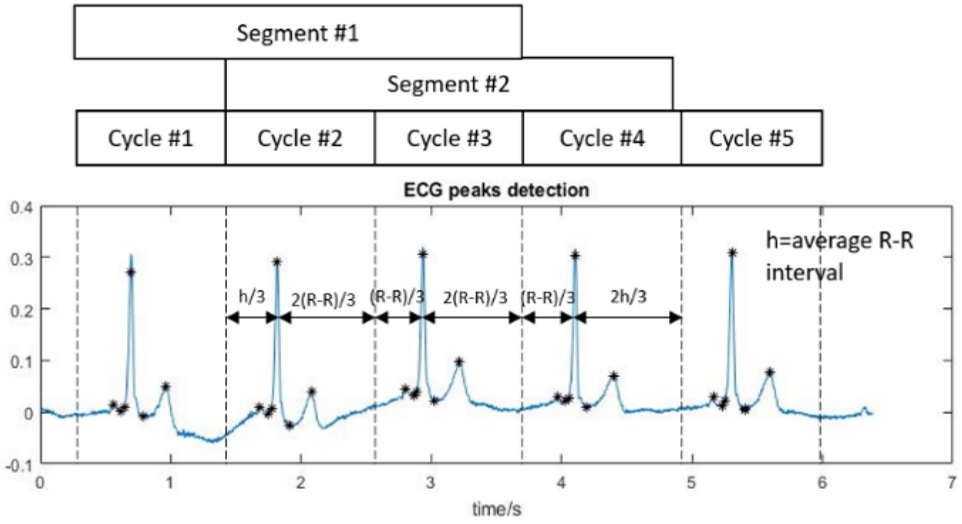

Figure 2.4: Segment samples correspond to three consecutive cardiac cycles

2.3 Feature Extraction

The feature extraction of ECG signals plays a crucial role in the diagnosis of heart disease and has a great influence on the performance of subsequent automated classification systems. De Chazal et al. discussed in detail the classification results using waveform morphological features in [9]. As discussed in [20], the combination of three types of characteristics (ie, temporal, morphology, frequency domain) can distinguish between types of arrhythmias.

According to the related methods in literature, the different types of ECG signals mainly differ in the power level of the frequency band of 5 to 15 Hz in the segmented signal, and some other morphological features (such as the duration between the Q wave and the T wave, P The distance to the R wave, etc.) also shows a correlation with the signal class.

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

Table 2.2: Features extracted from ECG signal

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	

Based on the sample segmentation method used in this paper, the feature extraction stage divided the three major types of features into periodic-based features (SET 1) and segment-based features (SET 2). Here, 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 once in segment intervals. 22 feature quantities are extracted for each segment, which is then normalized to obtain a feature of zero-mean unit variance, and a 221 vector is used to represent the k-th segment. Summarized details are described in Table.2.2. The 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, it is explicit that these 22 feature are not completely independent of each other. Therefore, in this work, Principal Component Analysis (PCA) is used to reduce the dimensions of features. Finally, 8 feature quantities were obtained after dimension reduction.

2.4 Classification Framework

In this section, a system is designed to address the classification and prediction tasks based on the processed ECG data. Based on our previous study and similar works which aimed to solve patient-specific classification problem, a two-staged structured which contains a global classifier and personalized classifier is retained as outline in this paper [12–14, 29]. Moreover, the proposed algorithm incorporates a novel deviation quantification module depicted in details in the following section.

The framework of system is summarized by the flowchart in Figure 2.5. A pre-trained Global Classifier serves as preliminary classifier, it facilitates the following analysis in the system for it identifies samples with severe morbidity. Depending on applications, different conventional classification algorithms with low complexity can be considered as Global Classifier. Any abnormal labels generated by Global Classifier are considered as Red Alarms and do not require further processing.

However, as the objective of this study is to identify latent status between normality and abnormalities, the design focus on processing samples classified as N while deviating towards abnormal classes. For this purpose, a simple structured classifier is insufficient since sample numbers of N and other morbid classes are unbalanced in training set DS1 as shown in Table.2.1, which results in missed abnormalities. Thus a deviation detection module is added after the Global Classifier to specifically identify latent status using patient-specific normal cluster.

In order to extract patient-specific information and adapt the classifier accordingly, the first 20% of total Normal samples of each patient are selected as initialization of Personal Dynamic Normal Cluster $\mathcal{N}_0^k$. The binary classification of N versus non-N which includes all abnormal classes is completed by first calculating the following distance metrics:

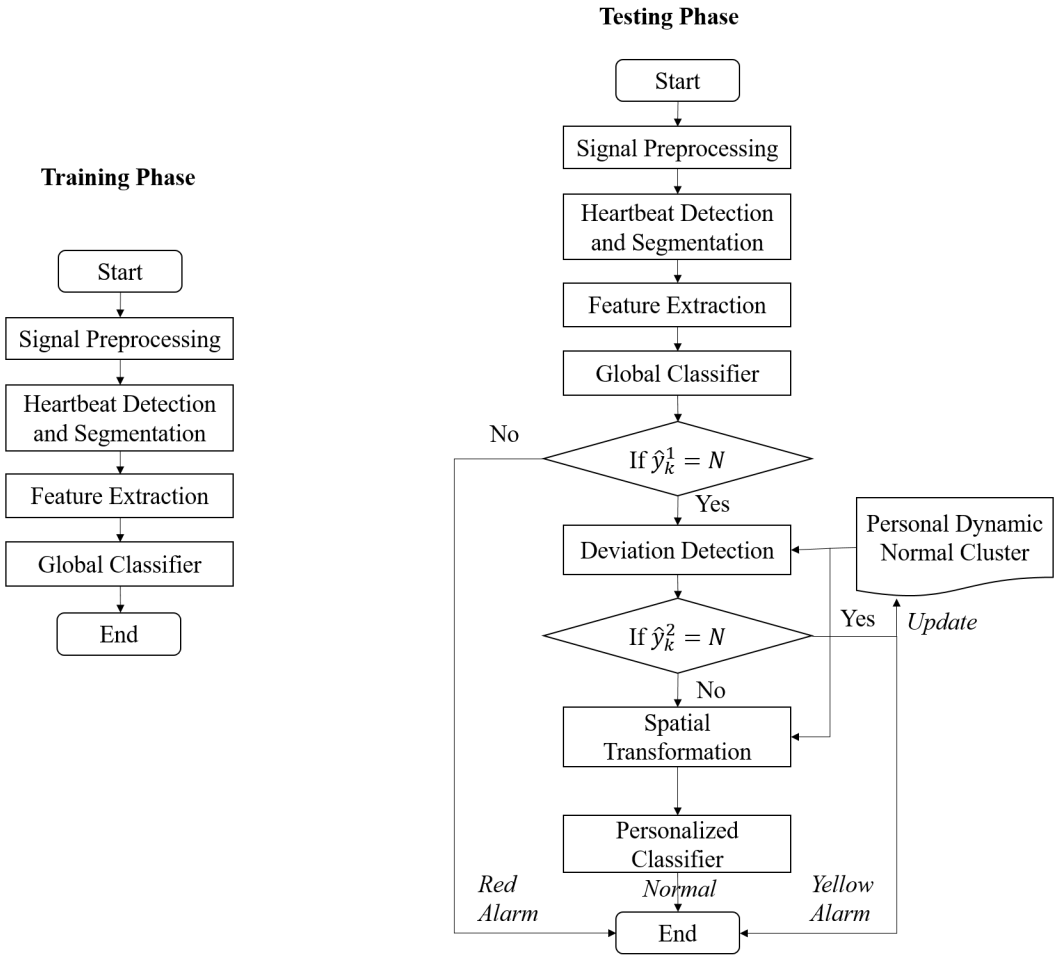

Figure 2.5: The general flowchart of proposed framework

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

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

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

The following conditions are thus examined to verify if deviation of a sample is within the range defined by α. Considering the fact that some rare abnormalities are unlikely to be

observed within the limit initialization time range, abnormal clusters: $\mathcal{S}, \mathcal{V}, \mathcal{F}$,which are composed of abnormal beats in DS1, are deployed as training data together with Personal Dynamic Normal Cluster $\mathcal{N}_i^k$ in the following steps.

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

If a sample is confirmed as N in this module, it will be used to update the $\mathcal{N}_i^k$. Otherwise the system assumes that the sample have demonstrated deviation towards abnormal clusters and will pass it to the subsequent Personal Classifier with controlled nonlinear transformation. Using transformation with optimized parameters, Personal Classifier is able to discern the deviation to different morbid types regardless of the cluster topology within the original feature space. Details of the module will be presented in the next Section.

Overall, if given a sample x_k at time k, the proposed framework maps it to label $\hat{y}_k \in \{N, Y_V, Y_S, Y_F, R_V, R_S, R_F\}$, where N stands for normal, Y_V stands for ventricular yellow alarm, R_V stands for ventricular red alarm.

2.5 Personal Classifier

As the proposed system aims at predicting subsequent abnormalities by analyzing the deviation of sample signal, it's vital to quantify the deviations using topological characteristics of training data. For most of the ECG applications analysis is conducted within high-dimensional feature space, we therefore select Cosine Distance(defined in Eq.2.5) to quantify deviation.

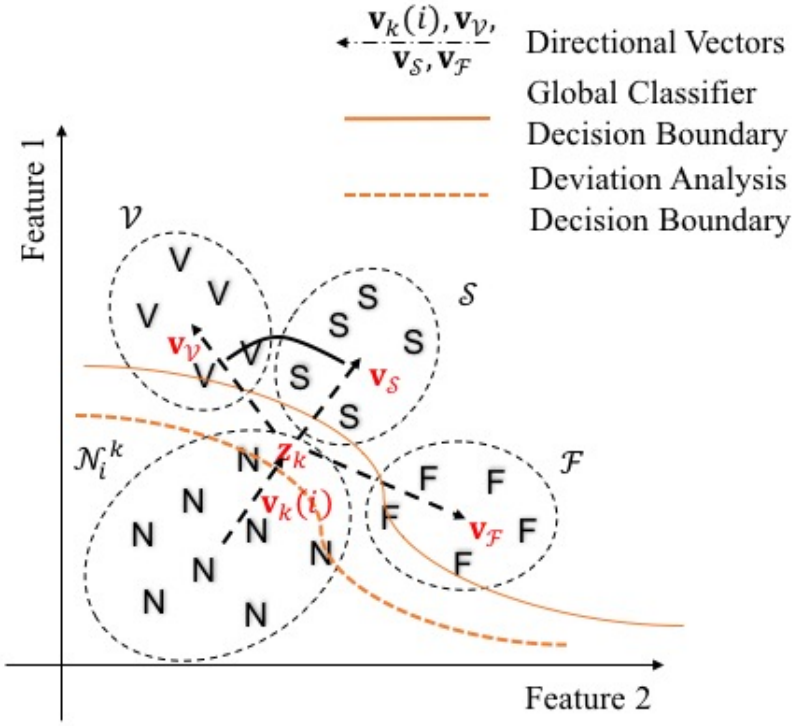

Figure 2.6: The deviation analysis boundary restrict on latent status between normal and abnormal samples compared with the Global Classifier boundary

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

Consequently, relative deviations of sample from normal cluster(N) to other abnormal clusters (V, S, F) are defined by cosine distance between the vector $\mathbf{v}_k(i)$ (defined in Eq.2.6) to 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, smaller cosine distance represents higher alignments between normal cluster centroid $\mathbf{v}_k(i)$, current sample $\mathbf{x}_k(i)$ and the abnormal centroid $\mathbf{c}_\mathcal{X}$.

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

Therefore the classification results of Personal Classifier $\hat{y}_k^2(i)$ is defined 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 measurement accuracy of cosine distance varies according to topology of clusters in feature space. Whereas the topology in feature space relies on feature extraction and feature selection. 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 avoid the ambiguity, a topology where abnormal clusters encircle normal cluster is required.

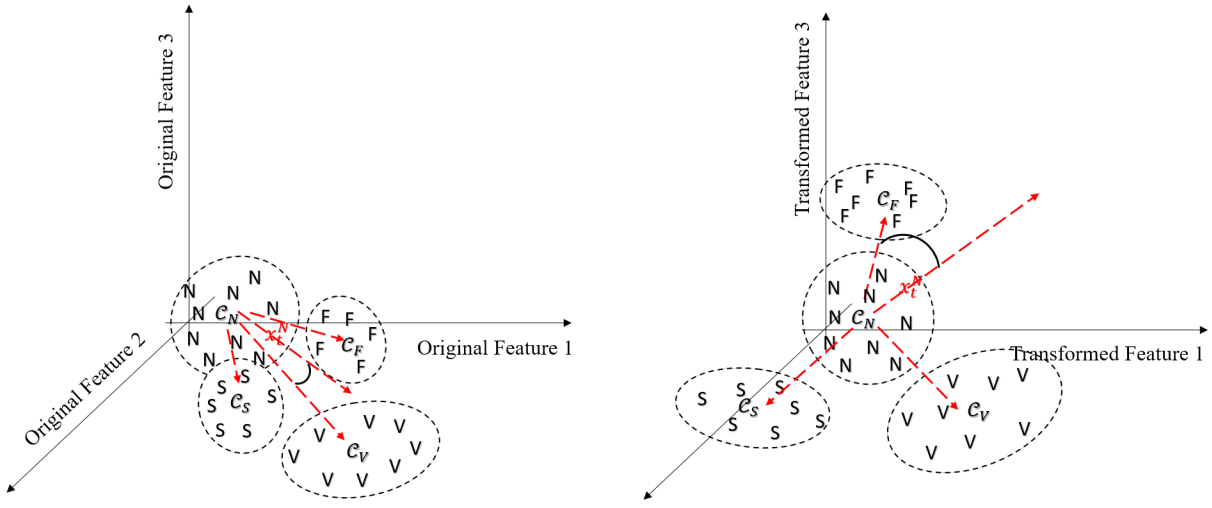

Figure 2.7: Left: illustration of cluster topology in original feature space; Right: illustration of cluster topology in feature space transformed with simple mapping function

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

Chapter 3

Kernel-Based Nonlinear Spatial Transformation

3.1 Introduction

As proposed in last chapter, the objective of the proposed system is to predict subsequent abnormalities by quantifying the deviation of a sample. Whereas the original cluster geometry in feature space Ω^d depends on the feature extraction stage $g()$. With the extracted features, the cluster geometry may not present a necessary symmetric property and will thus lead to poor performance or even failure in predicting subsequent abnormalities. Therefore, 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 reach the required symmetric property. The geometry of reshaped feature space should possess the following features:

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

- The overlapping part of all clusters is minimized.

Given that clusters in original feature space do not meet the symmetric property mentioned above and a spatial transformation is required, we assume in this chapter that new feature space obtained by non-linear reshaping is noted as $\Phi^{d'}$. This reshaping process is included in the 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 dimensional space, which can be considered as the union of subspaces for each type: $\Phi^{d'} = \{\mathcal{N}, \mathcal{S}, \mathcal{V}, \mathcal{F}\}$, where $d' > d$

3.2 Kernel Method

Kernel method has been widely applied in the machine learning algorithms, among which nonlinear Support Vector Machine (SVM) has been utilized in numerous applications recently [46]. Nonlinear kernel method can efficiently improve classification performance when there exists a nonlinear relationship between input and output variables. Because of the complexity and variety of feature vectors used in ECG analysis, the nonlinear relationship assumption is considered valid in this work. Therefore, it is practical to incorporate the nonlinear kernel method in the ECG analysis system.

In kernel SVM, the nonlinearities are introduced to the model by a kernel, which implicitly maps data $\mathbf{x}$ in input space Ω into a Hilbert space Φ via a function $\Psi()$ [47]. And the algorithm minimize the expected error $E[L(y, f(x)]$ so as to obtain the final classification mapping function f defined by $\Psi()$, where L is the designated loss function, such as square

error $(y - f(x))^2$ [48]. If there are m observations in the input space, we use notation $\mathbb{N}$ for index set $1 : m$. Based on the input space $\mathbf{x_i} \in \Omega(i \in \mathbb{N})$ and classification mapping function f, the optimization problem can be written as:

$$\text{minimize } \frac{1}{m} \sum_{i \in \mathbb{N}} L(y_i, f(\mathbf{x_i})) + \gamma ||f||^2 \tag{3.1}$$

where, $||f||^2$ is the squared norm of f; positive constant γ, also known as the regularization parameter, controls the balance between training error and the model complexity (smoothness).

When optimizing the objective function above, SVM only requires the inner products of the mapped features $\Psi(\mathbf{x})$ in Hilbert Space Φ. Therefore, a kernel defined as $k(\mathbf{x_i}, \mathbf{x_j}) = \Psi(\mathbf{x_i})^T \Psi(\mathbf{x_j})$ can efficiently substitute inner product calculation and introduce necessary nonlinearities in the model [49].

Different kernels represent distinctive nonlinear mapping functions. For machine learning models, the selection of kernel plays a decisive role. An effective kernel function generally needs to satisfy Mercer theorem [50]. Whereas, exhaustively searching for all possible kernel is computationally expensive and unrealistic [51]. A more efficient way to resolve this problem is to search for the optimal weighted combination of a set of base kernels, such as polynomial kernel function and Gaussian kernel function [52]. This method has been proven to be robust and efficient since the base kernels satisfy Mercers condition individually and is consistent with different dataset [53].

Polynomial kernel is usually applied on normalized data for its explicit expression and steady performance. Whereas the degree of freedom of polynomial kernel is comparatively high,

which leads to a large number of parameters. Gaussian kernel function, a very classic robust radial function, has high robustness when strong noise presents in the dataset. Nevertheless, since Gaussian kernel actually projects samples to an infinite dimensional space, it's difficult to visualize the projected observations $\Psi(\mathbf{x})$ and interpret the result.

In this work, the representative polynomial kernel is selected for the purpose of validating the proposed method and interpreting the effect of optimized nonlinear kernel method on feature space reshaping. The mapping functions, which is a weighted combination of polynomial kernel can be explicitly written in the following format:

$$\mathbf{z}_k = \mathbf{\Psi}_{\mathbf{w}}(\mathbf{x}_k) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d'} \end{bmatrix} \circ \begin{bmatrix} \psi_1(\mathbf{x}_k) \\ \psi_2(\mathbf{x}_k) \\ \vdots \\ \psi_{d'}(\mathbf{x}_k) \end{bmatrix} \tag{3.2}$$

where $\mathbf{w}$ is the coefficient of weight. Instead of selecting kernel, the process of spatial geometry optimization is accomplished by adjusting the coefficients of fixed polynomial basis functions $\psi()$. Since the number of parameters to adjust will increase exponentially when the order of polynomial functions increase, exhaustive searching are not practical for parameter optimization. Therefore, it's necessary to implement an heuristic optimization algorithm, in which parameters are obtained by maximizing or minimizing objective functions. More specifically, the nonlinear reshaping module in this chapter aims to adjust mapping coefficient $\mathbf{w} = [w_1, w_2, \dots w_d]^T$ to achieve the ideal symmetric geometry in the reshaped feature space while maintaining the separation between clusters.

3.3 Multiobjective Optimization

3.3.1 Objective Functions

To illustrate the optimization problem, we assume, in this example, that the original feature space is a 2-dimensional space Ω^2, then the mapping base kernel may adopt second-order polynomial function as follows:

$$\begin{aligned} &\mathbf{x} = [x_1\ x_2]^T, \quad \mathbf{w} = [w_1\ w_2\ \ldots\ w_5]^T, \quad d = 2, \quad d' = 5 \\ &\psi_1(\mathbf{x}) = x_1, \psi_2(\mathbf{x}) = x_2, \psi_3(\mathbf{x}) = x_1^2, \psi_4(\mathbf{x}) = x_2^2, \psi_5(\mathbf{x}) = x_1 x_2 \end{aligned} \tag{3.3}$$

Using similar concept to the loss function in the standard kernel SVM, the following objective functions can be used to represent the symmetric structure to be obtained:

$$\begin{aligned} o_1(\mathbf{w}) &= \frac{1}{\min\limits_{c,d=2,\ldots,p \text{ and } c\neq d} \{d(\mathbf{v}_{\mathcal{X}_c}, \mathbf{v}_{\mathcal{X}_d})\}} \\ o_2(\mathbf{w}) &= \frac{SW}{SB} = \frac{\sum_{c=1}^{C} \sum_{\mathbf{z}\in\mathcal{X}_c} (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})^T (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})}{\sum_{c=1}^{C} \sum_{d=1, d\neq c}^{C} (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})^T (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})} \end{aligned} \tag{3.4}$$

The maximization of pairwise cosine distance between vectors $\mathbf{v}_{\mathcal{X}_{c,d}}$, which connect the centroid of the normal cluster to centroids of abnormal clusters $\mathcal{X}_c$, can be obtained by minimizing $o_1(\mathbf{w})$. In fact, this objective functions is deduced from discrimination function of personal classifier in Eq.2.7. Cosine distance is defined by Eq.2.5 and the calculation of $\mathbf{v}_{\mathcal{X}_{c,d}}$ is as follows:

$$\mathbf{v}_{\mathcal{X}_i} = \mathbf{c}_N^k - \mathbf{c}_{\mathcal{X}_i} \tag{3.5}$$

Since, for some patients, the total number of a certain type of abnormal samples are very limited, abnormal samples in trainign set DS1 are utilized in calculating the two objective functions. In Eq.3.4, the abnormal cluster centroids are calculated using abnormal sample sets in training set DS1, while the normal cluster centroid is defined by the personal normal cluster.

On the other hand, $o_2(\mathbf{w})$ represents the ratio of within-cluster variance to between-cluster variance and consequently controls the separation between clusters. By minimizing $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ jointly, the algorithm eliminates the ambiguity while using Eq.2.7 to discriminate latent abnormal state and improves the predictive capability.

3.3.2 Multiobjective Particle Swarm Optimization

We noticed that $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ are not necessarily independent to each other. Thus, the optimization problem defined above is equivalent to joint minimization of $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ subject to a constraint condition: $|w|_2 = 1$. Since this is a non-convex multiobjective optimization problem, closed form solution as well as the optimization methods for convex function are not adopted in this method. In this work, we utilize Multiobjective Particle Swarm Optimization (MOPSO) algorithm to solve this optimization problem and obtain the optimal coefficients.

Particle Swarm Optimization (PSO) has the advantage of fast-converging, heuristic searching and easy implementation [54,55]. Therefore, researchers have been investigating in extending

PSO to multiobjective optimization problems. In the framework of MOPSO, the goal is to solve the typical Pareto optimization problem based on the algorithm in PSO. In other words, it aims at solving an optimization problem with two or more conflicting objective functions by approximating the Pareto front.

Pareto Front

In order to compare different set of coefficient in this opmization problem, the concept of Pareto front is briefly introduced in this section. For a multiobjective optimization problem with two objective function, if a solution $\mathbf{w}^1$ is said to *dominate* another $\mathbf{w}^2$ when the following two conditions are satisfied:

1. $o_1(\mathbf{w}^1) \leq o_1(\mathbf{w}^2)$ and $o_2(\mathbf{w}^1) \leq o_2(\mathbf{w}^2)$
2. $o_1(\mathbf{w}^1) < o_1(\mathbf{w}^2)$ or $o_2(\mathbf{w}^1) < o_2(\mathbf{w}^2)$

If a solution is not dominated by any other solutions in the searching space, then this solution is an *optimal* solution for this problem. A Pareto front is defined by the set of Pareto optimal solutions. However, in non-convex optimization, the Pareto front can not be represented explicitly by a deterministic function. Therefore, the majority of algorithms use heuristic searching algorithms to approximate the Pareto front [54].

MOPSO

Among all the MOPSO algorithms in the literature, the algorithm proposed by Coello Coello and Lechug presents a better performance and lower computational complexity than most of the other MOPSO algortihms [54]. Therefore, this algorithm implemented to solve the optimization problem in this work. One characteristic property of this algorithm is the

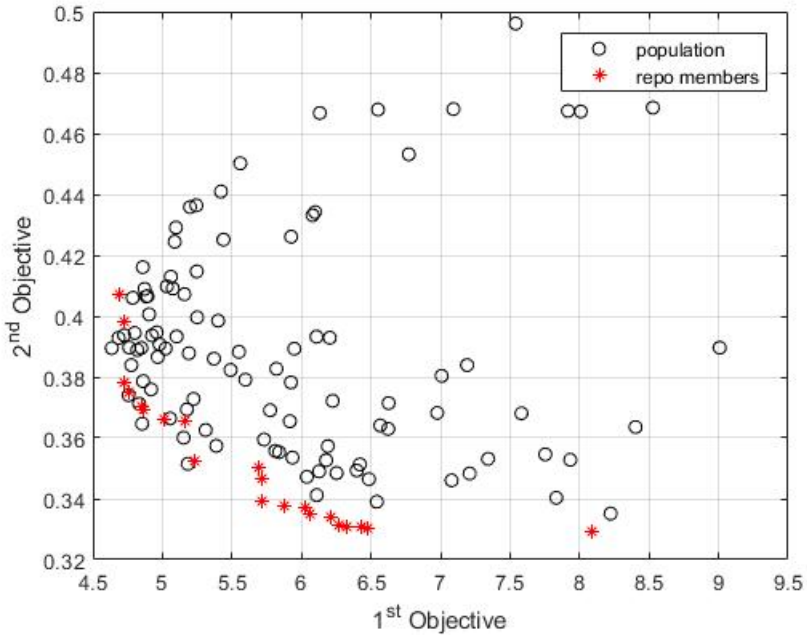

Figure 3.1: Particles stored in external repository approximate the Pareto front

external repository, in which all Pareto optimal particles for every swarm is recorded for each iteration. The solution represented by repository members are stored and used as an optimal approximation of the Pareto front because they converge to the actual Pareto front as proved in [54]. Fig.3.1, which is generated by jointly minimizing $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$, demonstrates the repository members are Pareto optimal than other particles and they converge to a uniform Pareto front.

With the concept of Pareto optimal, we further demonstrate the impact of applying kernel functions in this spatial reshaping problem by comparing the optimal solutions for coefficients of linear function combination and those for coefficients of polynomial kernel functions. Therefore, we first optimize the coefficients of third-order polynomial kernel functions, as formulated in Eq.3.3 and then optimize coefficients of linear features in origin feature space. The purpose is to verify if the objective functions can be fundamentally better optimized by introducing nonlinearities into the model.

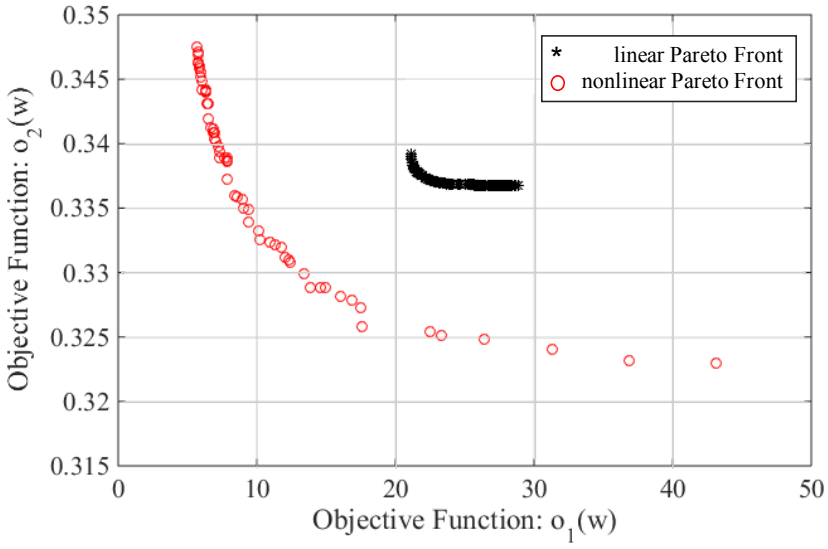

Figure 3.2: Increase of degree of freedom in optimization is proved by comparing the Pareto fronts generated by linear and nonlinear basis function

As shown in Fig. 3.2, the estimated Pareto front of the nonlinear model using polynomial kernel obviously dominates the Pareto front of linear model consist. The result shows that optimization with nonlinear kernel functions possess a higher degree of freedom and is able to find optimal solutions than the best solutions by linear combination. In other words, kernel method combined with multiobjective particle swarm optimization algorithm can improve the spatial structure of clusters according to the two objective functions.

3.4 Experimental Results

As mentioned in Section. 2.3, a cardiac segment is represented by an 8-dimensional vectors after feature extraction and PCA. To specify the nonlinear transformation in (3.2), the polynomial functions of order 3 is applied on the feature vectors. Moreover, taking the overfitting problem in high-dimensional space and computational cost into account, only 32 terms, which include 8 square terms x_i^2, 8 cubic terms x_i^3, 8 cross terms $x_i x_j$ of second order

and 8 cross terms $x_i x_j^2$ of third order, are retained by randomly discard the redundant cross terms. Therefore, the mapped vectors $\mathbf{z}_{32\times 1}$ include a total of 32 terms as follows.

$$\mathbf{z} = \{x_i^2 | i = 1, 2 \dots 8\} \cup \{x_i^3 | i = 1, 2 \dots 8\} \cup \tag{3.6}$$
$$\{x_i x_j | i, j = 1, 2 \dots 8, i \neq j\} \cup \{x_i^2 x_j | i, j = 1, 2 \dots 8, i \neq j\}$$

The performance of the aforementioned kernel-based method is tested on 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 classifying ECG signal segments. In order to evaluate the consistency and general classification classification results over all recordings, the median, interquartile range (IQR), mean and standard deviation of accuracy (AC), sensitivity (SE) and specificity (SP) are presented as evaluation metrics. 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 not likely to miss an important alarm or to report false alarms.

More importantly, the predictive capability of the proposed method is worthy of evaluating exclusively for it is the featuring specialty of the proposed system. In order to quantify the posterior probability of observing an abnormal signal after a preceding yellow alarm of similar type in (2.7), the number of predicted samples are counted as formulated in Eq 3.7:

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

$$P(\hat{y}_{k+i} = X_r | \hat{y}_k = X_y) = \frac{\#\text{ of } y_{k+i} = X \text{ after } \hat{y}_k = X_y}{\#\text{ of true alarms after } \hat{y}_k = X_y}$$
$$P(\hat{y}_{k+i} = X_r) = \frac{\#\text{ of true alarm of type } X \ (y_k = X)}{\#\text{ of all true alarms}} \tag{3.7}$$

The summary of results for all 21 test records is presented in Table. 4.4. The subsection *Probability of next abnormality (%)* in Table. 4.4 shows the confusion matrix of the probability of having a subsequent true abnormality of all types after observing a yellow alarm of all types along with the prior probability of observing a certain type of abnormal sample in the very last column.

These results validate the predictive capability of yellow alarms. For instance, the prior probability of observing a sample segment with abnormal types 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 in the dataset. However, the corresponding posterior 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 real abnormal segment of type V is $75\% - 67\% =$

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

	Count numbers of subsequent abnormality				Probability of subsequent 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

8%, which is higher than its prior 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. Therefore, we investigate in improving the spatial transformation module of the system and meanwhile reduce the computational cost in Chapter 4.

3.5 Conclusions

This chapter introduces and studies the kernel-based nonlinear transformation using Multiobjective Particle Swarm Optimization (MOPSO) method. Inspired by the concept of kernel method and loss function utilized in SVM, we implement the method with a weighted combination of base nonlinear kernels to reshape the input feature space by mapping it to a high-dimensional space. The coefficients of kernels are optimized according to two conditions, namely, maximum separation between cluster and maximum cosine similarities between abnormal clusters.

Result shows that approximated Pareto front produced by kernel method in the objective function space is apparently optimal to the one which is produced by the linear combinations of original features. The results verify that the proposed method has a classification accuracy

in the range of $88\% - 99\%$ for different ECG records in the test set of MIT-BIH database.

Above all, the proposed algorithm demonstrates the potential of providing detailed information about the sample deviations, which indicate the upcoming abnormal sample types. The predictive capacities of the system is verified with ECG signal, but this method is general and not bound to this application. If a biomedical signal has one base class (i.e. normal state) and several abnormal states, the proposed method can be implemented to predict upcoming abnormal types.

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].

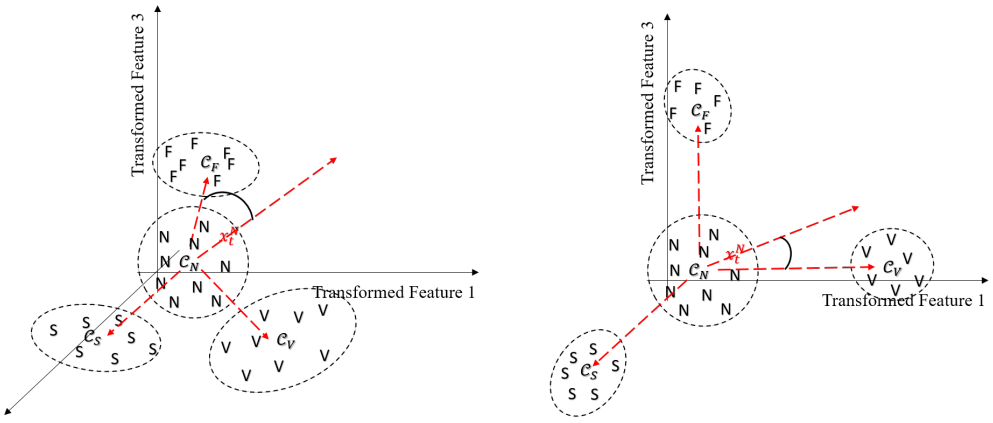

Figure 4.1: Left: illustration of cluster topology in feature space transformed with simple mapping function; Right: illustration of cluster topology in feature space transformed with optimized mapping function

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 [56]. 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 $0 \le \theta_j \le \pi,\ j = 1, \ldots, n-2;\ 0 \le \theta_{n-1} \le 2\pi;\ 0 \le r < \infty$

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 [57] based on the well known orthogonalization method Gram-Schmidt in [58] is deployed. Hence, in the first step of the function a vector of the centroids is fed to the orthogonalization process as follows:

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

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.

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

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 ($\mathbf{V}_{\mathcal{VN}}^{\perp},\ \mathbf{V}_{\mathcal{SN}}^{\perp},\ \mathbf{V}_{\mathcal{FN}}^{\perp}$), 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

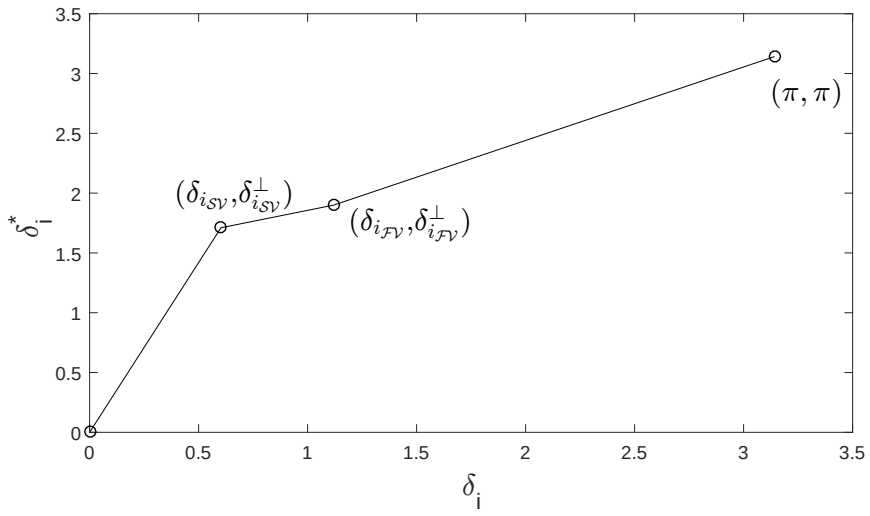

Figure 4.2: The simple mapping function

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

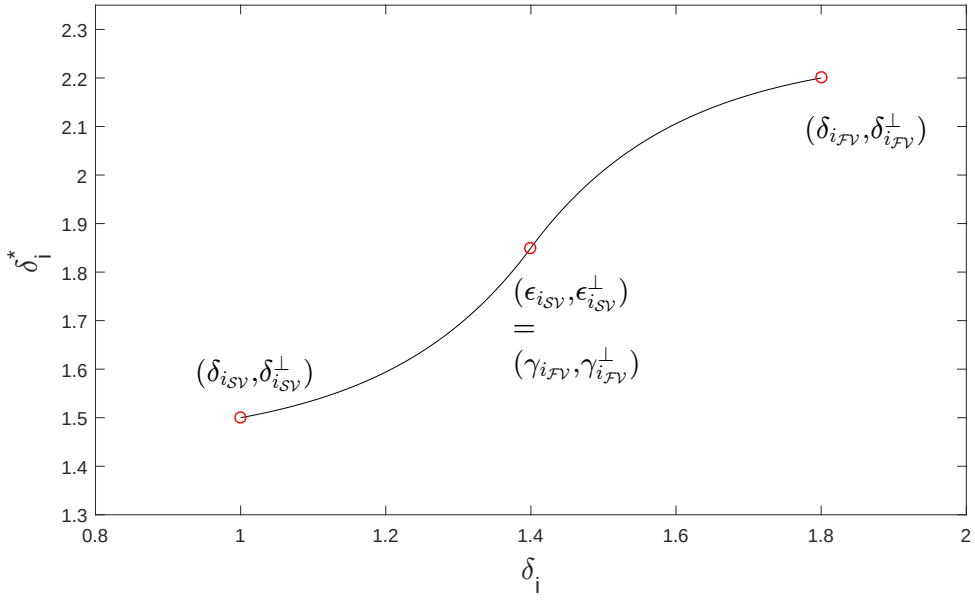

Figure 4.3: Optimized Piecewise Interpolate Function p

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.

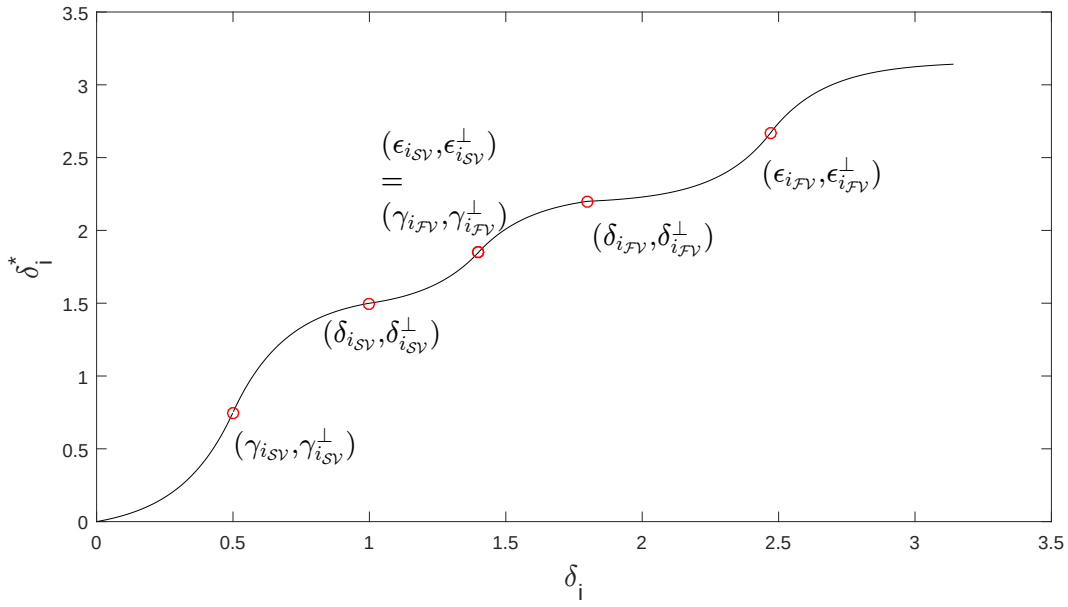

Figure 4.4: Optimized Mapping Function f

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] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American mathematical society*, vol. 68, no. 3, pp. 337–404, 1950.

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

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

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

[51] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.

[52] G. R. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, and M. I. Jordan, "A robust minimax approach to classification," *Journal of Machine Learning Research*, vol. 3, no. Dec, pp. 555–582, 2002.

[53] T. Jebara, "Multi-task feature and kernel selection for svms," in *Proceedings of the twenty-first international conference on Machine learning*, p. 55, ACM, 2004.

[54] 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.

[55] 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.

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

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

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