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, computational methods and networking platforms aiming at facilitate technology-based health service at lower costs. Biomedical signals can directly reflect the information regarding patient health and has therefore been widely investigated by healthcare providers and the research community. The essence of most signal analysis systems is to process a large training dataset and build a reference model to asses the health status of new patients. While the majority of these methods focus on improving classification performance on pooled dataset, the predictive modeling of biomedical signals 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 deviation analysis, two spatial transformation methods are proposed using a controlled nonlinear transformation to reshape the feature space into a more symmetric geometry. 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 a 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 methods [4]. However, these methods require complex manual analysis by trained physicians. Taking this issue into consideration, developing a cost-effective automatic analysis for CVD prevention based on computer is a critical need. 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 experienced physicians, automated ECG classification systems have 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 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 wide range of causing factors and difficulty of recognizing some 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 individuals and due to different physical and environmental conditions including but not limited to gender, age, body-mass index, elevation and air pressure, humidity, temperature, and 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 [12–17]. Some algorithms combine cardiologist manual annotations with automatically generated labels and train personal classifier with updated labels for each individual [12–14]. This design still requires experts' assistance in order to accurately classify ECG signal. Another design in to train a patient-specific classifier using only the patient's 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
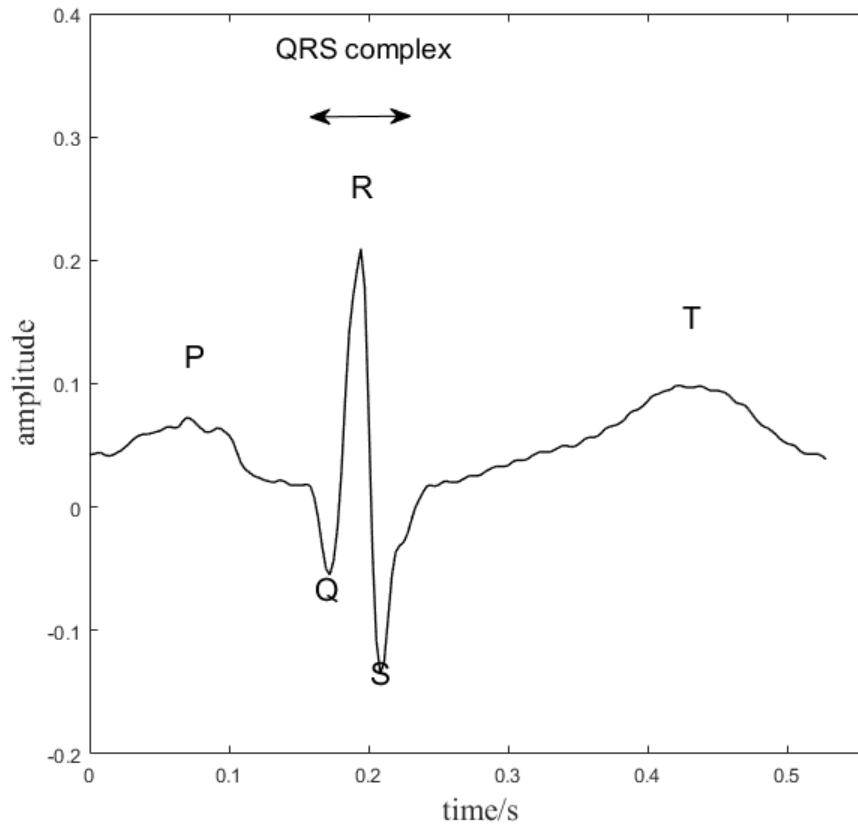
Figure 1.1: A typical cardiac cycle in ECG signal with five characteristic waves.

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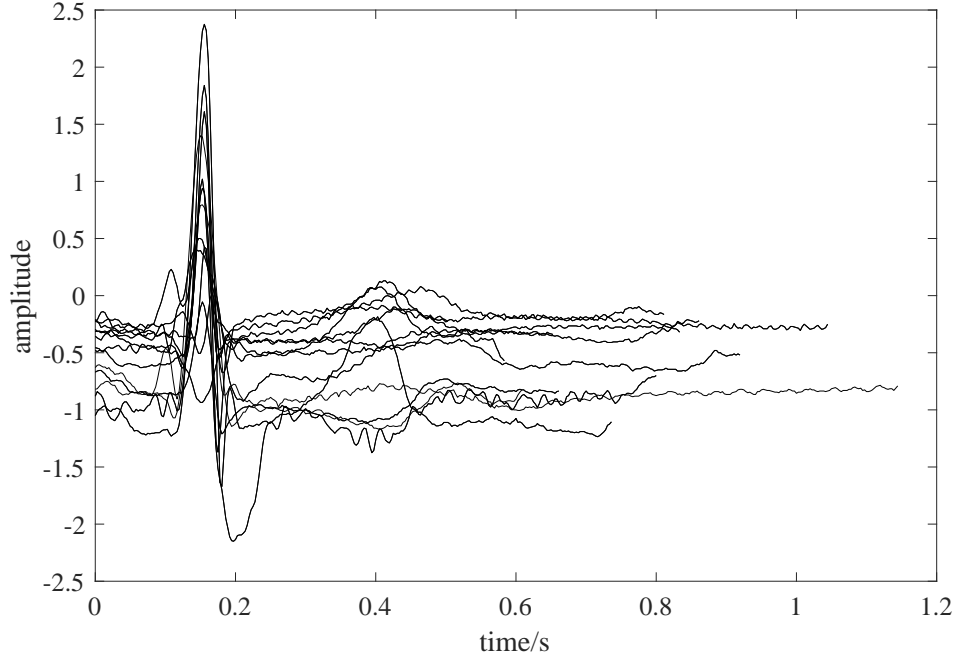
Figure 1.2: ECG signals of normal heartbeat from 15 different records in MIT-BIH reflect the inter-patient variability of ECG signal.

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.

8

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.
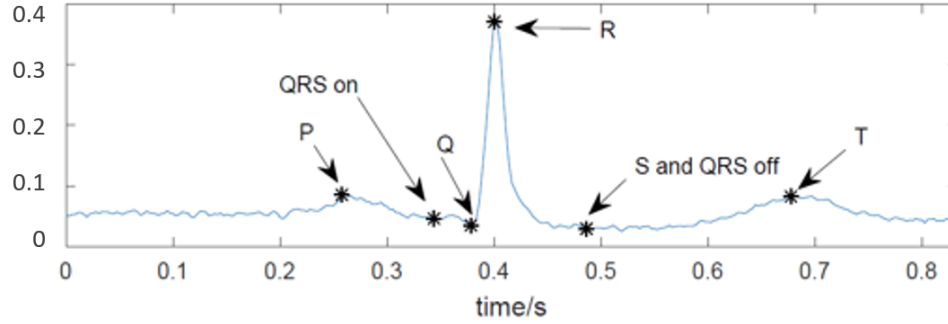
Figure 1.5: Fiducial peaks within one cardiac cycle.

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 a 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

16

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-Adaptive ECG Classification Framework

2.1   Introduction

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

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

sample. Continuous ECG signal is firstly segmented into individual intervals, each of which represents a heartbeat cycle, to be processed by the further stages of the proposed algorithm. Section 2.2 and Section 2.3 in this chapter focus on the data preparation stage, which includes four steps: signal preprocessing, delineation, segmentation, and feature extraction. Following the data preparation, Section 2.4 elaborates on the details of the proposed two-stage hierarchical classifier. The proposed classification system introduces a novel method for patient-adaptation by gradually capturing the normal range for each individual. More specifically, in Section 2.4, the dynamic normal cluster shaping method to achieve rationalization property is discussed. One feature of this method is that the cluster can track a patient's ECG waveform changes and dynamically adapt to it. In many applications, the physicians need to monitor a long-term real-time heart activity. This dynamic adaptation is able to address the issue of intra-patient temporal signal variation as well.

2.2 Utilized Dataset

Since the quality of ECG signals provided by most of the portable ECG measuring instruments is very unstable and may include transient noises. Signals transmitted through wireless communication systems will exhibit even more unstable waveforms. These transient effects appear in the resulting feature vector and consequently they negatively affect the prediction accuracy of the subsequent ML method. In order to eliminate and smooth out these transient terms, we use the concept of *segmentation* here, where each segment includes multiple cardiac intervals, as shown in Fig. 2.1. Note that the number of cardiac cycles in each segment is a modeling parameter shown by s_w. Also, we can arbitrarily slide segments equivalent to n_s cycles to generate the new feature vector. The parameters s_w and n_s can be tuned to improve the overall performance of the method. We choose $s_w = 3$ and $n_s = 3$,
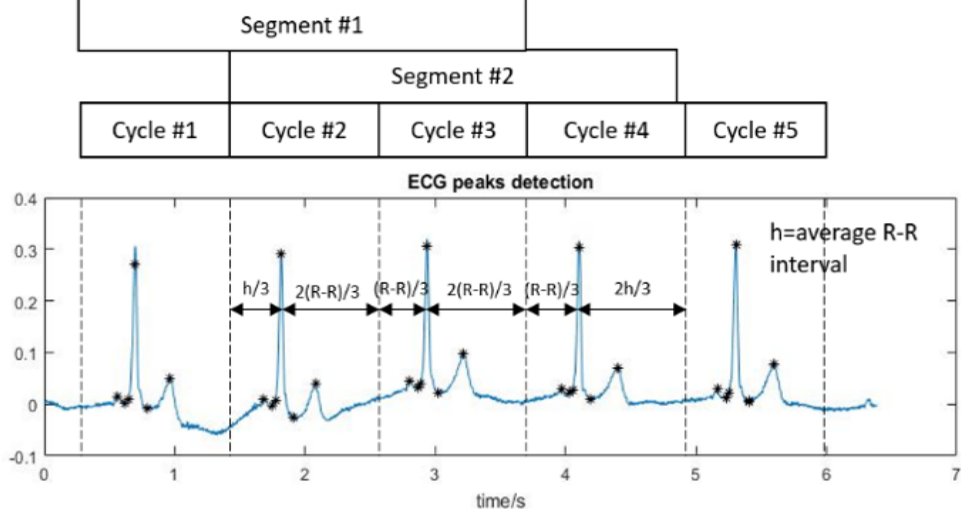
Figure 2.1: Illustration of the segmentation stage. Each segment consists of $s_w = 3$ cardiac cycles. We slide segments with $n_s = 1$ to obtain the next segment. Each segment is converted to a feature vector representing a data sample.

which yield the best classification accuracy, base on intensive simulations. In Fig. 2.1, we have $s_{window} = 3$ and $n_{steps} = 1$.

For the purpose of training and evaluating classifier, MITDB is split into test (DS2) and training (DS1) set by balancing the four classes according to [9].

The ECG signals in MITDB dataset are annotated and labels are provided for each cardiac cycle. However, we define a segment, which may include more than one cycle as a sample. Therefore, we need to translate per-cycle labels into per-segment labels. In this regard, a new label for each segment is generated by integrating all annotations of the beats within the segment. The segment is labeled as *normal*, if all member beats are annotated as N; otherwise, this segment will be labeled as the abnormality type of its member cycles if there is only one abnormality type. If more than one abnormality types are present within the same segment, the segment is excluded. For instance, segments with member cardiac cycles labeled "NNN", "SSS", "NVV", "VVV" are respectively mapped to "N","S", "V", and "V", and a segment with member label "NVS" is discarded. After segmentation and

20

Table 2.1: Training and test datasets in MITDB.

Evaluation Dataset	Number of segments per AAMI class				
	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

new annotation, the total number of samples in the training and test set are obtained as summarized in Table 2.1

2.3   ECG Signal Processing

2.3.1   Preprocessing

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

In this work, wavelet analysis is applied to ECG signals in MITDB with a sampling frequency of 360 Hz. Daubechies wavelet of order 8 ($db8$) is selected as mother wavelet for denoising

stage in this work for its optimal performance [31]. The decomposition coefficients and their corresponding frequency components are presented in Fig. 2.2.

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

2.3.2   Segmentation

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

Most existing methods use wavelet analysis to detect the highest peak (R wave) in a cardiac cycle as a reference point and then use signal morphology and typical properties of other waves to determine boundaries between consecutive intervals [20, 44–46]. In this study, we use this common approach. As was mentioned earlier in Section 1.4.2, a cardiac cycle consists of five basic characteristic peaks: P, Q, R, S, and T. Among them, the QRS complex is the most significant peak in one cycle. The energy of ECG signal in one cardiac cycle is mainly concentrated within the QRS complex. The QRS complex also conveys important informa-

22
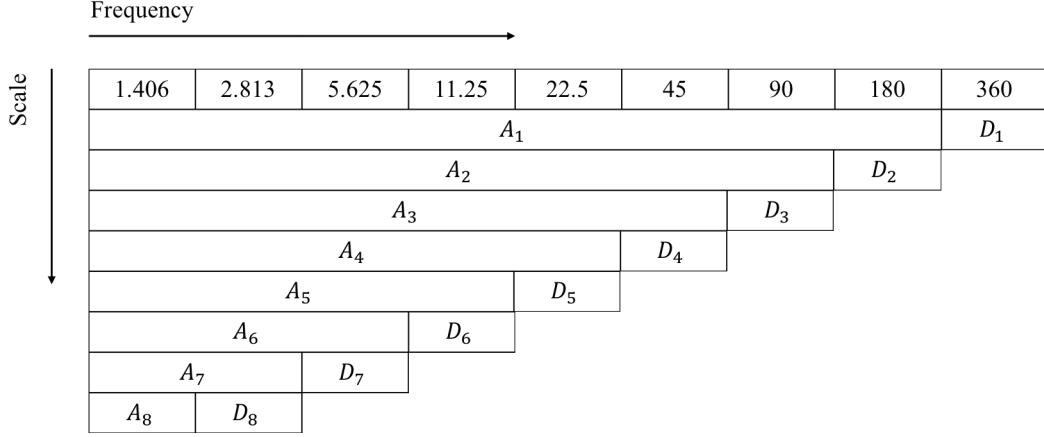
Figure 2.2: Frequency band of wavelet decomposition coefficients for MITDB signals.

tion that reflects the arrhythmia category [45]. Accurate detection of the QRS complex is of crucial importance for subsequent analysis. The energy of the QRS complex is generally within the range of 5-25 Hz. For ECG signals with sampling frequency of 360Hz, the QRS complex can be extracted from the detail coefficients of level 5 (D_5) and level 6 (D_6)

The mother wavelet db4 is utilized at this stage due to its morphological similarity to QRS complexes [43]. By superimposing D_5 and D_6, the QRS complex information in the ECG signal can be characterized in a one-dimensional recombined signal ($QRS\_DET = D_5 + D_6$).Other Fiducial peaks (P, QRS onset, Q, S, QRS offset and T waves for each cardiac circle) are localized according to an algorithm proposed in [45]. The accuracy of peak detection and its coincidence with the true signal are shown in Fig. 2.3.

With the empirical values described in [45], we use 15% of the highest amplitude within the signal as the detection threshold. Since the width of most of the QRS complexes does not exceed 160ms, here we use a sliding window with a width of 160ms to detect the peaks in the QRS_DET. The window step size is set to 200ms, given that the time lag between the two adjacent heartbeat cycles does not exceed 200ms. Fig. 2.4 shows a typical QRS_DET waveform along with the corresponding window of width 160ms. The false peaks are eliminated
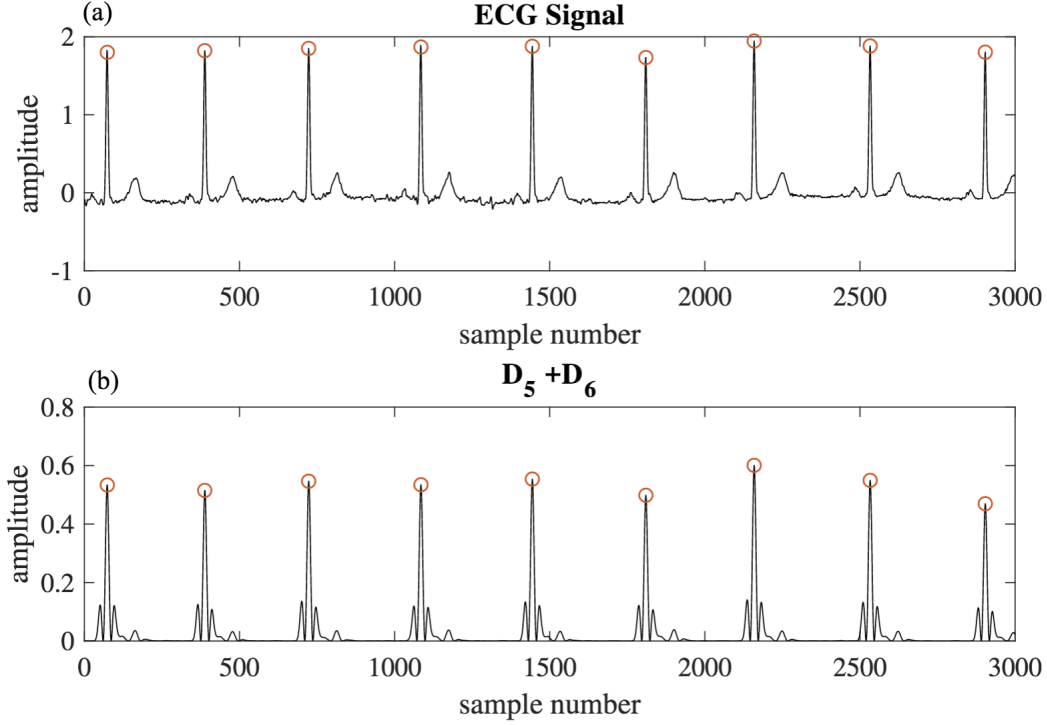
Figure 2.3: (a): the detected R peak locations within the original ECG signal; (b): the corresponding R peak locations within the signal obtained by adding up the level 5 and level 6 detail coefficients of the wavelet decomposition of the signal ($QRS\_DET = D_5 + D_6$).

using a 160ms time window, as seen in Figure.2.4.

The T and P waves are outside the QRS window. Through scanning the region spanning from the end of the precedent QRS window to the beginning of the current QRS window, the P-wave is located as the highest positive peak in this region. In a similar way, the position of the T wave is obtained by finding the maximum of the signal in the region between the end of the QRS in the current cardiac interval to the beginning of the next cardiac interval. In an alternative way, the two peaks between the two consecutive QRS windows, respectively represent the T and P waves. In short, an ECG signal in one heartbeat segment can be fully described by 7 feature points: P, QRS on, Q, R, S, QRS off and T.

After localizing the fiducial peaks, we can use the location of R peaks to determine the
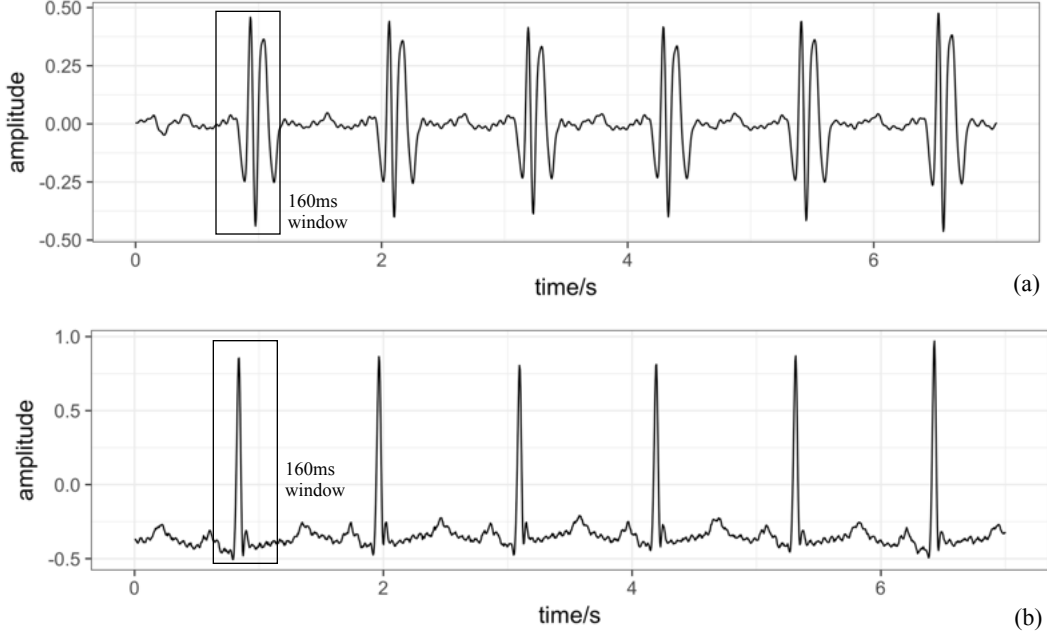
Figure 2.4: Window for detecting R peaks within QRS complexes.

boundaries of a cardiac cycle. By processing a large number of ECG signals, we realize that R peaks approximately split cardiac cycles into two pieces with 1/3 and 2/3 of the entire signal [45]. Based on this observation in the morphology of a typical ECG signal, we define the starting point of a cardiac cycle as the point which divides the distance between the two consecutive R waves into two sections, where the length of the first section is half the length of the second section, as depicted in Fig. 2.1. In this figure, the average interval (RR) between the two adjacent R waves is denoted by h, therefore R wave is located in distance $h/3$ with respect to the beginning of the cycle. The advantage of this method is its low computational complexity and easy generalization to different patients, noting that a slight change in cardiac boundaries do not significantly alter the properties of a cardiac cycle as long as T and P waves remain in the correct interval.

2.4 Feature Extraction

The feature extraction step plays a crucial role in the diagnosis of heart disease and has a great influence on the performance of the subsequent automated classification systems. De Chazal et al. have studied the impact of using waveform morphological features on classification results [9]. As discussed in [20], the combination of three types of characteristics (i.e. temporal, morphology, and frequency domain features) can provide a better discriminative power for the classification algorithm to distinguish between different types of arrhythmia.

According to several studies different types of ECG signals mainly differ in the power level of the frequency band ranging from 5 Hz to 15 Hz [43]. Likewise, some other temporal features (such as the duration between the Q wave and the T wave, the P wave to the R wave, etc.) present different levels of correlation with a specific signal classes [9]. Therefore, we choose to use the combination of temporal, morphological and spectral features as detailed in Table 2.2. Moreover, to account for segment-level characteristics as well as cycle-level properties, the extracted features include both cycle-based features (SET 1) and segment-based features (SET 2). 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 a segment, so it is calculated only once per segment. Therefore, we have a total of $8 \times 2 + 6 = 22$ features per segment as shown in Table 2.2. Therefore, each feature vector is a 221 vector with zero-mean unit-variance elements after a proper normalization. In Table 2.2, mean $(R_{i+1} - R_i)$ refers to the mean of the time lag between two adjacent R waves, while $(R_i - R_{avg})$ is the length of each cardiac cycle and the average cardiac cycle duration of the patient.

From Table 2.2, one may notice that these 22 features are not completely independent of each other. Also, some of the features may not be as relevant as the others. Therefore, we

Table 2.2: Features extracted from ECG signal

Feature Type	SET1	SET2
Temporal Features	- QRS duration - QT duration - PR duration	$\text{mean}(R_{i+1} - R_i)$ $\text{mean}(R_i - R_{avg})$
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	

reduce the number of features to obtain a more robust predictive modeling [9, 14]. We use Principal Component Analysis (PCA) as a commonly used the common dimension reduction method instead of explicit feature selection for its improved performance in biomedical signal processing [47]. We keep the 8 dominant directions of the signal after PCA as the most informative 8 features.

2.5 Classification Framework

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

2.5.1 Global Classifier

The flowchart in Figure 2.5 presents the overall operation of the proposed system. The *global classifier* is trained using the whole training dataset as the first classification step. It facilitates the subsequent analysis in the system by identifying samples with severe morbidity. Depending on the application (properties of signals, the utilized labels, and the choice of features), different classification algorithms can be utilized [12, 14]. Two important considerations include the classification accuracy and the computation complexity of the method [14]. Any abnormal label generated by the global classifier is considered as a *red alarm* and does not require further processing. However, samples labeled as *normal* go through the subsequent personalized classification step.

2.5.2 Deviation Detection

Since, one objective of this study is to identify the fuzzy states between normality and abnormalities, the subsequent analysis is focused on processing samples classified as $normal$(N) to check whether or not they show considerable deviations towards one of the abnormality classes. For this purpose, a one-layer classifier is not sufficient due to multiple reasons: i) the numbers of samples in the normal and different morbid classes are not balanced in the training set DS1 as shown in Table 2.1, ii) the patient-specific normal trend is unknown, iii) detecting yellow alarms require a new set of decision rules to determine whether or not the deviations are worthy of calling a yellow alarm. Therefore, a deviation detection module is added after the Global Classifier to identify fuzzy state between normality and abnormalities using patient-specific normal cluster. In order to develop a ground for patient-specific normal functionality and adapt the classifier accordingly, the first 20% of total Normal samples of each patient are selected as the initialization of *personalized dynamic normal cluster* $\mathcal{N}_0^k$. As
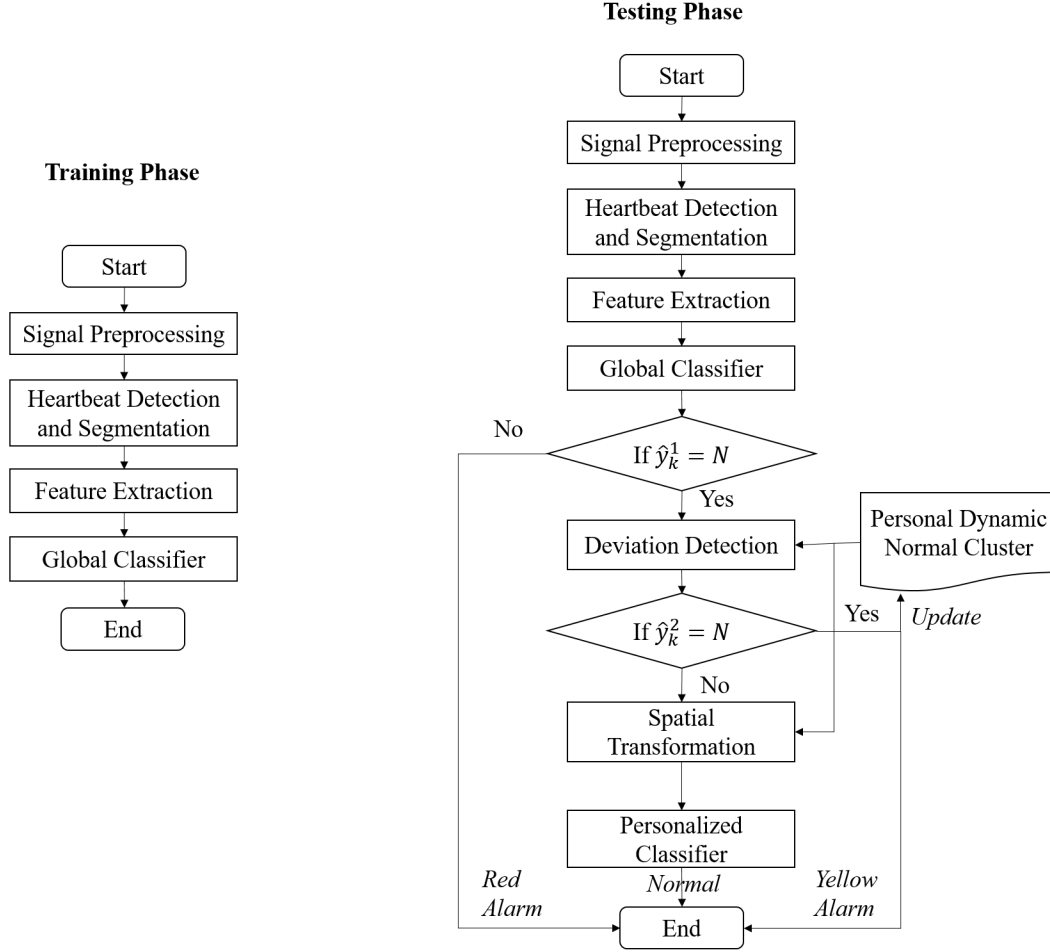
Figure 2.5: The general flowchart of proposed framework.

we collect more samples from the patient, the samples which are newly classified as normal (N) will be used to update the personalized dynamic normal cluster. Therefore, if a sample $\mathbf{x}_k(i)$ is classified as N after deviation detection and the *mahalanobis distance* between $\mathbf{x}_k(i)$ and the centroid of the personalized dynamic normal cluster $\mathbf{c}_N^k$ is less than 2, then $\mathbf{x}_k(i)$ is included in the updated $\mathcal{N}_i^k$. mahalanobis distance quantifies how many the standard deviation away $\mathbf{x}_k(i)$ is from $\mathbf{c}_N^k$, it can be formulated as follows:

$$d_{\mathbf{mahal}} = \sqrt{(\mathbf{x}_k(i) - \mathbf{c}_N^k)'\mathbf{S}^{-1}(\mathbf{x}_k(i) - \mathbf{c}_N^k)}, \qquad (2.1)$$

where $\mathbf{S}^{-1}$ stands for the covariance matrix of $\mathcal{N}_i^k$.

To distinguish between normal state and fuzzy state, we use a binary classification of N versus non-N, where the second includes all abnormal classes. We firstly calculate the following distance metrics:

$$R_i^{\mathrm{max}} = \max_{\mathbf{x}_j \in \mathcal{N}_i^k, \mathbf{x}_k \in \mathcal{N}_i^k} \{\sqrt{(\mathbf{x}_j - \mathbf{x}_k)^2}\}, \qquad (2.2)$$

$$D_{\mathcal{X}}(\mathbf{x}_k(i)) = \underset{\mathbf{x} \in \mathcal{X}}{\mathrm{median}}\{\sqrt{(\mathbf{x}_k(i) - \mathbf{x})^2}\}, \qquad (2.3)$$

$$D_{\mathcal{N}}^{\mathrm{max}}(\mathbf{x}_k(i)) = \max_{\mathbf{x} \in \mathcal{N}_i^k}\{\sqrt{(\mathbf{x}_k(i) - \mathbf{x})^2}\}. \qquad (2.4)$$

The following conditions in Eq. 2.5 are then examined to verify if the deviation of a sample is within the range defined by α. Since some rare abnormalities are unlikely to be observed within the limited initialization period, therefore abnormal clusters: $\mathcal{S}, \mathcal{V}, \mathcal{F}$, which include abnormal beats for all patients in DS1, are deployed as the training dataset when calculating $D_{\mathcal{X}}(\mathbf{x}_k(i))$ in Eq. 2.3

$$\begin{cases} D_{\mathcal{N}}^{\mathrm{max}}(\mathbf{x}_k(i)) \leq \alpha R_i^{\mathrm{max}}, \\ D_{\mathcal{N}}(\mathbf{x}_k(i)) < D_{\mathcal{X}}(\mathbf{x}_k(i)) \quad \text{for } \mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\} \end{cases} \qquad (2.5)$$
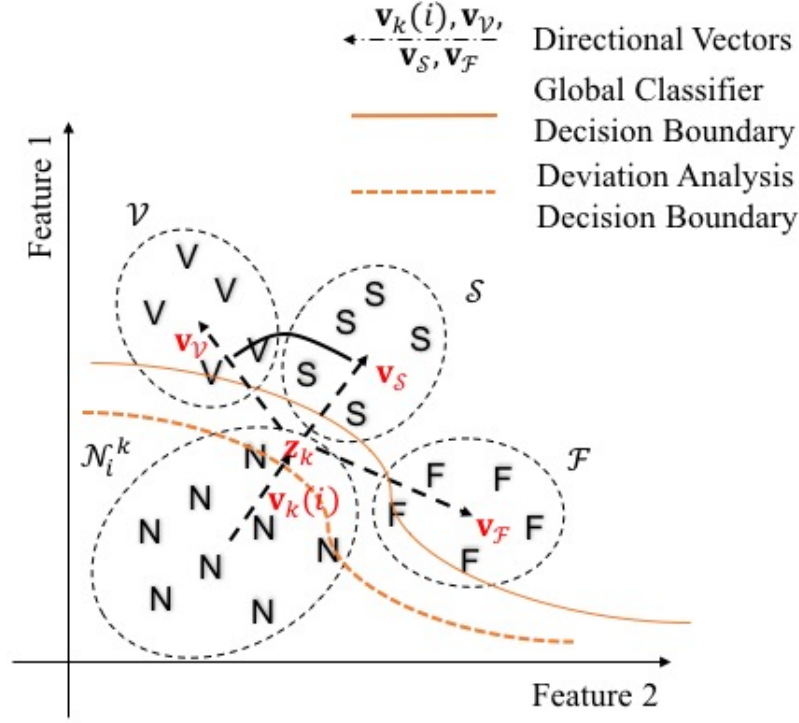
Figure 2.6: The deviation analysis boundary restrict on latent status between normal and abnormal samples compared with the Global Classifier boundary.

If a sample already labeled normal by the global classifier, is again confirmed as N in this module, it will be used to update the $\mathcal{N}_i^k$ as mentioned above. Otherwise, the system assumes that the sample has a considerable deviation towards one of the abnormal clusters and hence will pass it to the subsequent *personalized classifier*. The *personalized classifier* uses controlled transformation with optimized parameters to discern the deviation to different morbid types regardless of the cluster topology within the original feature space, as detailed in Chapters 3 and 4.

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

2.6 Personalized Classifier

Here, we provide the core idea behind the design of the personalized classifier, and the details of implementation will be discussed in Chapters 3 and 4. Since the proposed system aims at predicting subsequent abnormalities by analyzing a signal's deviation from the patient's normal functionality of the sample signal, it is vital to quantify the deviations using topological characteristics of the training dataset. For most of the ECG applications, ECG signal is analyzed by its representative feature vectors. A natural choice for deviation analysis in the high-dimensional feature space, is *cosine distance* (as defined in Eq. 2.6) to quantify the distance between two vectors $\mathbf{v}$ and $\mathbf{w}$.

$$d(\mathbf{v}, \mathbf{w}) = 1 - \frac{\mathbf{v}^T \mathbf{w}}{|\mathbf{v}||\mathbf{w}|} = 1 - \frac{\mathbf{v}^T \mathbf{w}}{\sqrt{(\mathbf{v}^T \mathbf{v})(\mathbf{w}^T \mathbf{w})}} \tag{2.6}$$

Consequently, relative deviations of a sample from normal cluster(N) to other abnormal clusters (V, S, F) are defined by the cosine distance between the vector $\mathbf{v}_k(i)$ (defined in Eq.2.7) and the three vectors $\mathbf{v}_{\mathcal{X}}(i) = \mathbf{c}_{\mathcal{X}} - \mathbf{x}_k(i)$ where $\mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\}$. In this case, a smaller cosine distance represents a higher alignment between the vector from the normal cluster centroid $\mathbf{v}_k(i)$ to the current sample $\mathbf{x}_k(i)$ and the reference vector from the normal cluster centroid to abnormal centroids $\mathbf{c}_{\mathcal{X}}$.

$$\mathbf{v}_k(i) = \mathbf{x}_k(i) - \mathbf{c}_N^k(i) = \mathbf{x}_k(i) - \sum_{\mathbf{x} \in \mathcal{N}_i^k} \mathbf{x}/|\mathcal{N}_i^k|, \tag{2.7}$$

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

$$\hat{y}_k^2(i) = \underset{\mathcal{X} \in \{\mathcal{S}, \mathcal{V}, \mathcal{F}\}}{\operatorname{argmin}} \{d(\mathbf{v}_k(i), \mathbf{v}_\mathcal{X}(i))\} \tag{2.8}$$

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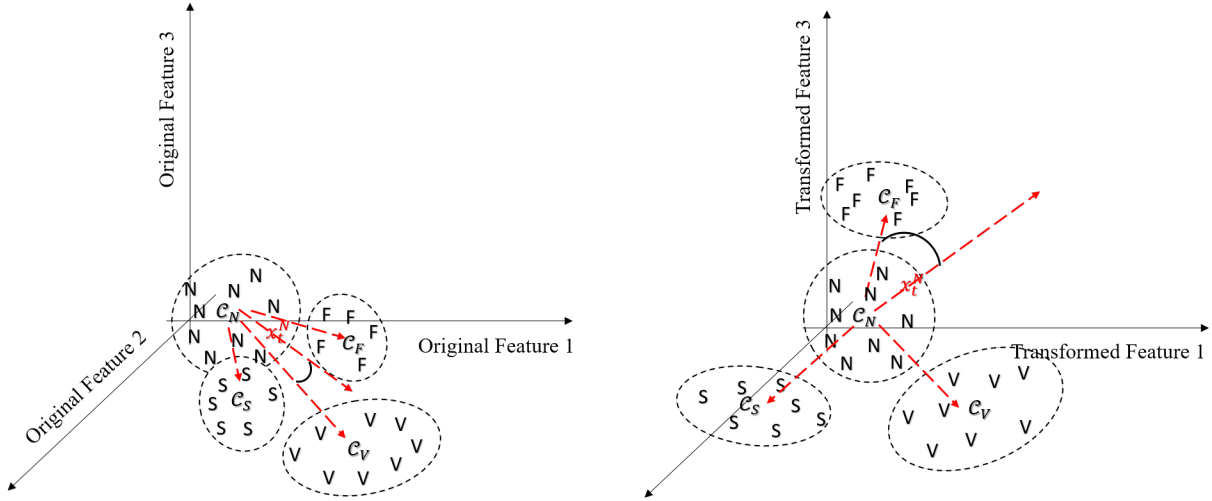
Figure 2.7: Left: illustration of the cluster topology in the original feature space; Right: illustration of the cluster topology in the transformed feature space using a nonlinear mapping function.

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 discussed in Chapter 2, the main objective of the personalized classifier is to reassess the normal samples to identify deviation of seemingly normal samples into any of abnormality types. The original geometry of clusters in the feature space Ω^d depends on the choice of features implied by the feature extraction and feature selection stages $g()$. We noticed that with the resulting features (in this work), the clustering geometry does no exhibit the necessary symmetric property and thus leads to a poor performance and even failure in predicting subsequent abnormalities. Therefore, an optimization method based on spatial transformation is proposed to solve this issue. More specifically, we propose a method to reshape the clusters such that

- The abnormal classes surround the normal class;

- A maximal separation among the abnormal classes are achieved;

- The angles between the vectors connecting the centroid of normal cluster to different abnormal clusters are equalized.

These properties can be achieved through imposing the following conditions:

- Vectors pointing from the normal centroid to different abnormal clusters centroids present maximum mutual cosine distances.

- The overlapping parts among all clusters are minimized.

Given that the clusters in original feature space do not meet these symmetric properties, developing a spatial transformation is unavoidable. In this chapter, a *kernel*-based nonlinear spatial transformation is proposed to reshape the feature space to reach the above-mentioned required symmetric properties. This reshaping process is part of the *personalized classification* stage (as shown in Fig.2.5) of the ECG classification system as described in chapter 2. The nonlinear mapping projects the corresponding feature vector of each sample $\mathbf{x}_k$ in the original space Ω^d onto a new vector $\mathbf{z}_k$ in a higher dimensional space denoted by $\Omega^{d'}$. This is achieved using a nonlinear mapping function $\Psi^{d'} : \Omega^d \rightarrow \Omega^{d'}$. The resulting vectors are used by the personalized classifier to identify the minor yellow alarm type out of $\{\mathcal{S}, \mathcal{V}, \mathcal{F}\}$.

3.2   *Kernel* Method

Kernel method has been widely used in machine learning algorithms. For instance, it is the integral part of nonlinear Support Vector Machine (SVM), which has been utilized in numerous applications recently [48]. Nonlinear kernel methods can efficiently improve the classification performance when there exists a nonlinear relationship between the input and

output variables. Because of the complexity and diversity of feature vectors used in ECG analysis, the assumption of nonlinear relationship is considered valid in this work. Therefore, incorporating nonlinear kernel methods in the ECG analysis system, can be beneficial.

In kernel SVM, the nonlinearities are introduced to the model through a kernel function, which implicitly maps data points $\mathbf{x_i} \in \mathcal{X}$ in the input space Ω into a Hilbert space Φ via a nonlinear function $\Psi()$ [49]. Then, the algorithm minimizes the expected error $E[L(y, f(x)]$ between the true labels y and the predicted values $f(x)$ for samples in a training dataset, by finding an optimal classification function f, which also depend on the choice of $\Psi()$. Here, $L()$ is an arbitrary loss function, and a popular choice is the least squared errors $\sum_{x_i \in \mathcal{X}} (y_i - f(x_i))^2$ [50]. Other choices for loss function include Hinge loss, absolute loss, hit and miss loss, etc [51, 52].

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. For instance, if f is a polynomial function of order p, (i.e. $f(v) = \alpha_0 + \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_i v_1^2 + \alpha_j v_1 v_2 + ... \alpha_k v_N^p$), the norm of f is defined as:

$$||f||_2 = \sqrt{\sum_{j=0}^{k} \alpha_j^2}. \tag{3.2}$$

The positive constant γ, also known as the regularization parameter, controls the balance between training error and the model complexity (smoothness).

When optimizing the above objective function, SVM only requires the inner products of the transformed features $\Psi(\mathbf{x})$ in the 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 the inner product calculation and induces the necessary nonlinearities into the model [53].

Different *kernels* represent different nonlinear mapping functions. For machine learning models, the selection of kernel plays a crucial role. Therefore is no straightforward method to choose the best kernel and it is typically chosen by try and error and other heuristic model selection methods. An effective kernel function generally needs to satisfy Mercer conditions, so that the inner products can be replaced by kernel functions, as used in SVM [54]. An exhaustive search for all possible kernels is a computationally expensive and unrealistic task [55]. A more efficient way to resolve this issue would be to search for an optimally weighted combination of a set of base kernels, such as polynomial kernel function and Gaussian kernel function [56]. This method has been proven to be robust and efficient since the base kernels satisfy Mercers condition individually and it can be consistent with different datasets [57].

Polynomial kernel is usually applied on normalized data for its explicit expression and steady performance. Polynomial kernels with lower degrees are most commonly used, since higher degrees tend to overfit the training data [58].

Gaussian kernel function denoted by $k(v, w) = exp(-\frac{||v-w||^2}{\sigma})$ for vectors v and w is a very classic robust radial function, which has shown a string robustness in the case of noisy datasets [59]. However, it is equivalent to the inner product of samples after projecting into an infinite dimensional space; therefore it is difficult to visualize the projected observations $\Psi(\mathbf{x})$ and interpret the results.

Considering the above-mentioned facts, in this work, the polynomial kernel is selected for the purpose of validating the proposed method and interpreting the effect of an optimized

37

nonlinear kernel method on feature space reshaping. However, the proposed methodology is general and applicable to other nonlinear kernels.

The mapping function, which is a weighted combination of polynomial kernels can be explicitly written in the following format:

$$
\mathbf{z}_k = \mathbf{\Psi}_{\mathbf{w}}(\mathbf{x}_k) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d'} \end{bmatrix} \circ \begin{bmatrix} \psi_1(\mathbf{x}_k) \\ \psi_2(\mathbf{x}_k) \\ \vdots \\ \psi_{d'}(\mathbf{x}_k) \end{bmatrix}, \tag{3.3}
$$

where $\mathbf{w}$ is the vector of normalized coefficient.

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 free parameters increases exponentially with the order of polynomial function, an exhaustive search is not practical for parameter optimization. Therefore, it is necessary to implement a heuristic optimization algorithm, in which parameters are obtained by maximizing or minimizing an objective function. More specifically, the nonlinear reshaping module in this chapter aims to adjust mapping coefficients $\mathbf{w} = [w_1, w_2, \ldots w_d]^T$ to achieve the ideal symmetric geometry in the reshaped feature space while maintaining the maximal separation between clusters.

3.3 Multiobjective Optimization

3.3.1 Objective Functions

To elucidate the details of the optimization problem, here we consider an illustrative example, where the original feature space is a 2-dimensional space Ω^2. We also assume for simplicity that the order of the polynomial function is 2. Therefore, we have:

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

$$\psi_1(\mathbf{x}) = x_1, \psi_2(\mathbf{x}) = x_2, \psi_3(\mathbf{x}) = x_1^2, \psi_4(\mathbf{x}) = x_2^2, \psi_5(\mathbf{x}) = x_1 x_2. \tag{3.4}$$

The coefficient vector $\mathbf{w}$ is a 5×1 vector. Therefore, the algorithm needs to search in a 5-dimensional space to find an optimal vector $\mathbf{w}$ so that the geometry of feature space after transformation has the desired properties described in Section 3.1. For this purpose, we define two objective functions to impose the symmetry and separation of different abnormality classes. The concepts of *Fisher discriminant* function and cosine distance are used to quantify these properties. The objective functions are formulated as follows:

$$o_1(\mathbf{w}) = \frac{1}{\displaystyle \min_{c,d=2,\ldots,p \text{ and } c \neq d} \{d(\mathbf{v}_{\mathcal{X}_c}, \mathbf{v}_{\mathcal{X}_d})\}} \tag{3.5}$$

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

The maximization of pairwise cosine distance between the vectors $\mathbf{v}_{\mathcal{X}_{c,d}}$ connecting the centroid of the normal cluster to the centroids of abnormal clusters $\mathcal{X}_c$ is achieved by minimizing

39

$o_1(\mathbf{w})$. In fact, this objective functions is deduced from discrimination function of personal classifier in Eq.2.8. Cosine distance is defined by Eq.2.6 and the calculation of $\mathbf{v}_{\mathcal{X}_{c,d}}$ can be written as follows:

$$\mathbf{v}_{\mathcal{X}_i} = \mathbf{c}_N^k - \mathbf{c}_{\mathcal{X}_i} \tag{3.6}$$

Since for some patients, the total number of a certain type of abnormal samples are very limited, the abnormal samples in training set DS1 are utilized in calculating the two objective functions. In Eq.3.5, the abnormal cluster centroids are calculated using the abnormal samples in training dataset DS1, while the centroid of the normal cluster is defined by the preceding normal samples for the same person.

On the other hand, $o_2(\mathbf{w})$ represents the ratio of the within-cluster variance to the between-cluster variance and consequently controls the separation between the clusters. By minimizing $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ jointly, the algorithm eliminates the ambiguity of classification while improving the predictive power of the personalized classifier due to the symmetric geometry of clusters.

3.3.2 Multi-objective Particle Swarm Optimization

We notice that $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$ are not necessarily independent of 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: $|w|_2 = 1$. This constraint is necessary since the first objective function $o_1(\mathbf{w})$ is inversely proposal to $|w|_2$, whereas the second objective function $o_2(\mathbf{w})$ is scale-invariant.

This problem is a non-convex multi-objective optimization problem. Therefore, neither closed form solutions nor the optimization methods proposed for convex problems are applicable to this case. In this work, we utilize *multi-objective particle swarm optimization* (MOPSO) algorithm to solve this optimization problem and obtain the optimal coefficients [60].

Particle swarm optimization (PSO) is based on heuristic search and has the advantage of fast convergence, and easy implementation [60, 61]. PSO is defined to solve problems with a single objective function, where closed form solutions are not tractable. Several research works are devoted in the past decade to extend this method to multi-objective optimization problems [62, 63]. In the MOPSO framework, the goal is to solve the typical *Pareto optimization problem* based on the evolutionary algorithm used in PSO. In other words, it aims at solving an optimization problem with two or more conflicting objective functions by approximating the *Pareto front*.

In order to compare different set of coefficient in this optimization 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 [60].
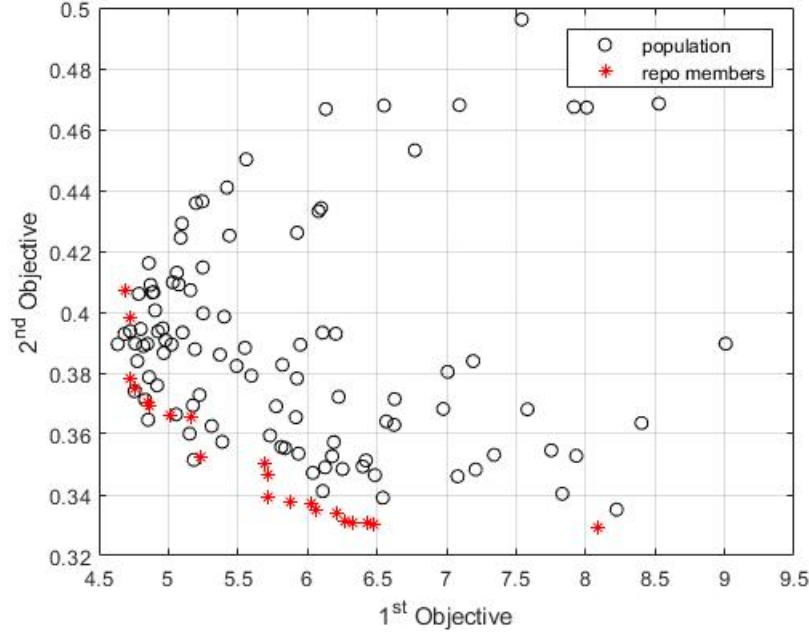
Figure 3.1: Particles stored in external repository approximate the Pareto front.

Implementation Details of MOPSO

MOPSO is a popular multiobjective optimization algorithm, which uses Pareto dominance to determine the direction of a particle in each iteration by combining the optimal direction found by the particle itself and the global optimal direction. Among different implementations of MOPSO, the algorithm proposed by Coello Coello and Lechug presents a better performance and lower computational complexity in most applications [60]. Therefore, this algorithm is implemented and utilized in this work to solve the multi-objective optimization problem. One special property of this algorithm is the use of 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 [60].

Fig.3.1 presents the results of joint minimization of objective functions $o_1(\mathbf{w})$ and $o_2(\mathbf{w})$.

This figure demonstrates that the repository members are Pareto optimal compared to the other particles. This figure also confirms that the repository members converge to a uniform Pareto front.

Using the concept of Pareto optimality, we demonstrate the impact of applying kernel functions in this spatial reshaping problem by comparing the Pareto front of the optimization problem obtained by using MOPSO for two different scenarios including i) the optimized coefficients for the original data sample, (i.e. a linear identity function) and ii) the transformed samples under polynomial kernel whose coefficients are optimized using MOPSO. Therefore, we first optimize the coefficients of the third-order polynomial kernel function, as formulated in Eq.3.4 and then optimize the coefficients of linear features in the origin feature space. The purpose of this comparison is to investigate whether or not the resulting objective functions are fundamentally improved by incorporating nonlinear terms into the feature vectors through the proposed polynomial function.

As shown in Fig. 3.2, the estimated Pareto front of the nonlinear model using the polynomial kernel dominates the Pareto front of the original linear model. This result is expected since the transformed samples exhibit a higher degree of freedom by adding new dimension to the data in the feature space through the nonlinear mapping. A higher degree of freedom enables the MOPSO algorithm to tune the optimization parameters and find better solutions than the best solutions achievable by the original data samples. In other words, the kernel method combined with multi-objective particle swarm optimization algorithm can improve the spatial structure of the clusters quantified by the two objective functions.
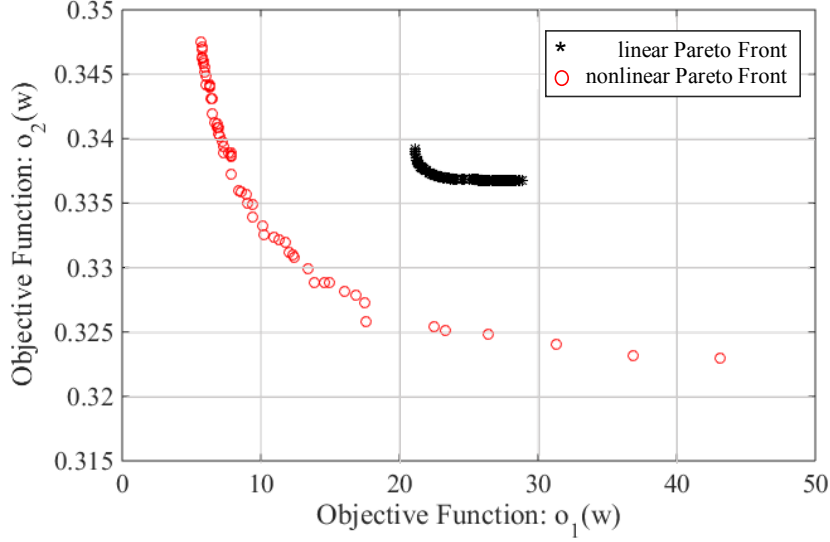
Figure 3.2: The *Pareto front* of the results of MOPSO is significantly shifted when using the transformed feature vectors. This improvement is due to the increase in the degree of freedom provided by additional non-linear dimensions added to the samples.

3.4 Experimental Results

As mentioned in Section 2.3, a cardiac segment is represented by an 8-dimensional vector $\mathbf{x} = [x_1, x_2, \ldots, x_8]$ after the feature extraction and the PCA-based dimension reduction stages. To specify the nonlinear transformation in (3.3), a polynomial function of order 3 is applied to the feature vectors. The resulting transformed vectors as follows:

$$[x_1, x_2, \ldots, x_8, x_1^2, x_2^2, \ldots, x_8^2, x_1^3, x_2^3, \ldots, x_8^3, x_1 x_2, \ldots x_6 x_7 x_8], \tag{3.7}$$

which include 165 terms, 8 of which are the original features. This high dimension may cause the classifier to trap into the overfitting problem. It also significantly increases the computational complexity of the algorithm. To solve these issues, we discard some of the induced terms and include only 8 square terms x_i^2, 8 cubic terms x_i^3, and 8 cross terms of power two $x_i x_j$ and 8 cross terms of power three $x_i x_j^2$. We randomly choose these terms after

44

discarding the redundant cross terms. Therefore, the mapped vectors $\mathbf{z}_{32\times 1}$ include a total of 32 terms as follows:

$$\mathbf{z} = \{x_i^2 | i = 1, 2 \ldots 8\} \cup \{x_i^3 | i = 1, 2 \ldots 8\} \cup \tag{3.8}$$
$$\{x_i x_j | i, j = 1, 2 \ldots 8, i \neq j\} \cup \{x_i^2 x_j | i, j = 1, 2 \ldots 8, i \neq j\}$$

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 as well as the 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. The results are promising and the median of the classification accuracy for all classes are in the range of $88\% - 99\%$. Sensitivity and specificity of the proposed method exhibit 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, since it is unique feature provided by the proposed system. In order to quantify the posterior probability of observing an abnormal signal after a preceding yellow alarm of similar type in (2.8), the number of predicted samples are counted as formulated in Eq 3.9:

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

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

$$P(\hat{y}_{k+i} = X_r) = \frac{\# \text{ of true alarm of type } X \ (y_k = X)}{\# \text{ of all true alarms}} \tag{3.9}$$

The summary of results for all 21 test records is presented in Table. 3.2. The values provided under the column *Probability of next abnormality (%)* in Table. 3.2 present 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 regardless of the preceding yellow alarm in the very last column. These results confirm the predictive capability of yellow alarms as well as the scientific fact that yellow alarm are indicative of upcoming red alarms. This conjecture supported by the fact that at least some of the heart problems develop over time, although the symptoms may appear suddenly.

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

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

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 $75\% - 67\% = 8\%$ 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. We conclude this section by stating that a new methodology provided in Chapter 4 to optimize the nonlinear transformation using an analytical approach, which significantly reduces the computational cost.

3.5 Summary of Contributions

In this chapter, we proposed a novel method which combines kernel-based nonlinear transformation with MOPSO optimization method. With the concept of kernel method and the loss function, we implement the method using 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.

By studying kernel method as well as heuristic optimization methods based on the spatial

topology optimization concept, the outcome of this chapter can be concluded as follow:

- translate the ideal symmetric geometry of feature space into two objective functions;

- design a kernel-based transformation using weighted combination of base kernels;

- validate that nonlinear transformations are more capable of improving the spatial topology than linear transformations.

In order to verify the efficiency of this method, we applied the method to test ECG samples in the MIT-BIH database [28]. The experimental results show that the proposed method has a classification accuracy in the range of $88\% - 99\%$ for different ECG records.

Above all, the proposed algorithm demonstrates the potential of assessing signals deviation from its patient-specific normal trend in order to indicate the upcoming abnormality classes. The predictive capacities of the system is tested with ECG signals, but the proposed method is general and applicable to similar problems beyond the ECG signal processing, studied in this work. If a biomedical signal has one base class (i.e. normal state) and several abnormal states, we can deploy the proposed method to trigger yellow alarms and predict upcoming abnormality classes.

Chapter 4

Controlled Spatial Transformation With Deterministic Mapping Function

4.1   Introduction

In Chapter 2, we elaborated on the details of the patient-adaptive ECG classification framework. An important part of this methodology is to design a *personalized classifier* with a *deviation analysis module*. The performance of the *deviation analysis module* depends on the geometry of different classes of data samples in the feature space. In Chapter 3, we proposed a novel spatial transformation method based on MOPSO to achieve a desired symmetry in the transformed feature space to boost the performance of the deviation analysis module. This methodology enables us to further process the normal samples and identify fuzzy states between the normal and abnormal states. However, the proposed method uses an iterative optimization which may have high computational complexity. In this chapter, we introduce another deterministic spatial transformation method to model the fuzzy state of samples with a more tractable and analytical solution.
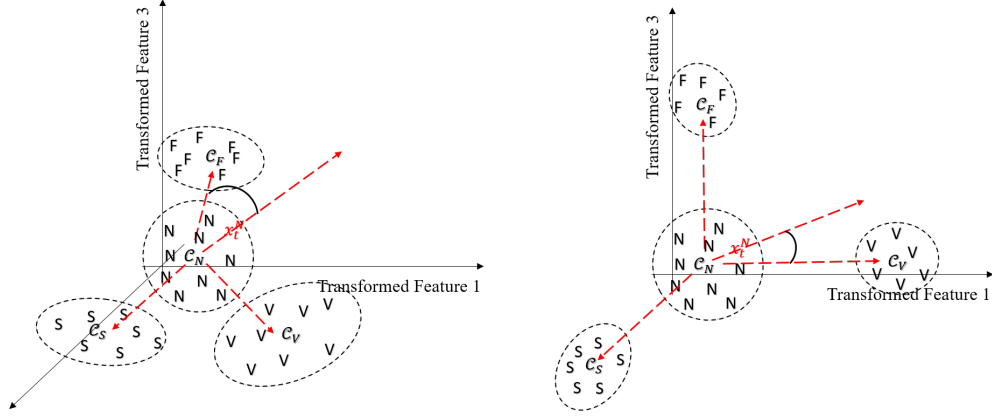
Figure 4.1: Left: illustration of the clustering topology in the transformed feature space without reducing the within-cluster variance; Right: illustration of the clustering topology in the feature space transformed with the optimized mapping function, which reduces the within-cluster variance.

While the aforementioned method in Chapter 3 yields the desired capacity of predicting upcoming abnormalities, the interpretation of the mechanisms of the system is not straightforward and easily tractable, thus it is not easily generalizable to a broader range of applications in the biomedical signal processing. The main objective of this chapter is to develop a deterministic transformation with a stronger prediction power based on a controlled spatial topology studied in Chapter 3. In this method, we further optimize the topology and spatial geometry of the clusters in the feature space by reducing the within-cluster variance after the spatial transformation. As shown in Fig. 4.1, through this improvement, the performance of the personalized classifier in terms of prediction accuracy can be further enhanced with the proposed method.

4.2   Hyper-Spherical Coordinates

In the previous chapter, the spatial transformation module is implemented with a polynomial kernel function and a heuristic optimization method. The system performance has been

proven to be promising. In addition to the general drawbacks of heuristic method, it is not straightforward to select an appropriate base kernel as the core of spatial mapping function due to the high variety of kernel functions.

In order to address this issue, a novel deterministic spatial mapping function is proposed in this chapter based on *hyper-spherical* coordinates [64]. Since hyper-spherical coordinates consist of angles and radius, these parameters in both original feature space and the desired target space are used to determine the mapping function.

The hyper-spherical coordinate system (n-dimensional spherical coordinate system) and its mapping to the *Cartesian* coordinate system are first introduced in [64]. If $\mathbf{x}$ is a sample vector in a n-dimensional feature space, with its Cartesian coordinates $(\xi_1,\ \xi_2\ ...\xi_n)$, 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 equation: $\sin(\arccos(x)) = \sqrt{1 - x^2}$.

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

$$\xi_1 = r\cos(\theta_1)$$

$$\xi_2 = r\sin(\theta_1)\cos(\theta_2)$$

$$\xi_3 = r\sin(\theta_1)\sin(\theta_2)\cos(\theta_3)$$

$$\vdots$$

$$\xi_{n-1} = r\sin(\theta_1)\cdots\sin(\theta_{n-2})\cos(\theta_{n-1})$$

$$\xi_n = r\sin(\theta_1)\cdots\sin(\theta_{n-2})\sin(\theta_{n-1}). \tag{4.2}$$

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

4.3 Orthogonalization

To simplify the algorithm, in this chapter, the topology of clusters in the feature space is approximated by their centroid locations, denoted by $\mathbf{c}_N^k, \mathbf{c}_\mathcal{V}, \mathbf{c}_\mathcal{S}, \mathbf{c}_\mathcal{F}$, respectively for the normal cluster and the three abnormality clusters of type V, S, and F. Furthermore, as we assume that the samples with fuzzy states deviate from the normal cluster to a specific abnormality cluster, the spatial topology stays unchanged if the normal centroid is simply translated to the origin of the Cartesian coordinate system. The clustering topology in the original feature space can be equivalently represented by the following matrix with three row vectors:

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

As shown in Fig. 4.1, in order to improve the performance of the personalized classifier and to avoid ambiguity in quantifying the deviations to different abnormality classes, a topology with a maximum separation between the three vectors of $\mathcal{C}$ is preferred. To achieve a lower computational complexity in higher order dimensions, the algorithm aims at transforming vectors in $\mathcal{C}$ to orthogonal vectors using a deterministic function. This approach not only simplifies the transformation derivations, it also ensures a full symmetry among abnormality clusters. As such, the problem reduces to finding a transformation that maps a set of vectors to a set of orthogonal vectors in the same space. Therefore, we can use the popular method of *Gram-Schmidt* orthogonalization method explained in [65, 66]. Hence, in the first step of this stage, the three row vectors of $\mathcal{C}$ representing the centroids of the three abnormality clusters are fed to the orthogonalization process as follows:

$$
\mathcal{C}^{\perp} = \text{Gram-Schmidt}(\mathcal{C}) = \begin{bmatrix} \mathbf{v}_{\mathcal{V}\mathcal{N}}^{\perp} \\ \mathbf{v}_{\mathcal{S}\mathcal{N}}^{\perp} \\ \mathbf{v}_{\mathcal{F}\mathcal{N}}^{\perp} \end{bmatrix} \tag{4.4}
$$

where $\mathcal{C}^{\perp}$ is the matrix of orthogonalized vectors in the Cartesian coordinate system. The hyper-spherical coordinates $\mathcal{C}_*^{\perp}$ of these orthogonalized vectors are calculated subsequently using Eq.4.1. After this step, the orthogonalized vectors in the hyper-spherical coordinate system are obtained as formulated in Eq. 4.5. The jth angular dimension denoted as $\theta_j^{\perp}$ and the radius is noted as $r^{\perp}$.

53

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

4.4 Spatial Mapping Function

After obtaining the original hyper-spherical coordinates of $[\mathbf{v}_{\mathcal{VN}}, \ \mathbf{v}_{\mathcal{SN}}, \ \mathbf{v}_{\mathcal{FN}}]^T$ and the orthogonalized hyper-spherical coordinates $[\mathbf{v}^{\perp}_{\mathcal{VN}}, \ \mathbf{v}^{\perp}_{\mathcal{SN}}, \ \mathbf{v}^{\perp}_{\mathcal{FN}}]^T$, the goal is to design a mapping function $\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^n$ from the original coordinates to the orthogonal ones which exhibit the desired clustering topology, such that Eq. 4.6 holds.

In the Gram-Schmidt algorithm, the very first vector serves as a reference vector and remains unchanged in the orthogonalization process, namely $\mathbf{v}_{\mathcal{VN}} = \mathbf{v}^{\perp}_{\mathcal{VN}}$. Therefore the mapping function $\mathbf{F}$ shall satisfy the following equations:

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

Furthermore, since the orthogonality of vectors is independent to their radii r, we only need to design $\mathbf{F}$ for the $n - 1$ angular dimensions $(\theta_1, \ldots, \theta_{n-1})$, and the coordinate r remains unchanged after the mapping. Consequently, the function $\mathbf{F}$ can be decomposed into $n - 1$ functions: $f_i : \mathbf{R} \to \mathbf{R}$, $i = 1 \ldots n - 1$ with constraints in Eq.4.6 as well as two additional extreme points defining the range of input and output values. We use the notation $\mathbf{v}_{\mathcal{SN}} - \mathbf{v}_{\mathcal{VN}} = \Delta_{\mathcal{SV}}$ and denote the the ith angular dimension of $\Delta_{\mathcal{SV}}$ as $\delta_{i_{\mathcal{SV}}}$. We use similar

notations for the other dimensions, (e.g. $\mathbf{v}_{\mathcal{FN}} - \mathbf{v}_{\mathcal{VN}}$). Hence for each angular dimension i, f_i is determined by $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$, along with the two *extreme boundary points* defining the domain and range of the functions.

In order to maintain the simplicity and the linearity of the mapping function, the functions f_i should be continuous and monotonic. For this purpose, the valid range of angular dimensions (after considering the periodicity of these functions) is used to determine the boundary constrains; therefore the problem is transformed into a curve fitting problem.

For example, if linking $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$ results in a monotonically increasing function, noting the fact that the range of $n-2$ first angular dimensions is the interval $[0, \pi]$, then the two extreme boundary points would be $(0, 0)$ and (π, π), as depicted in Fig. 4.2. Conversely, for a monotonically decreasing function, the extreme boundary points would be $(\pi, 0)$ and $(0, \pi)$. These rules apply for the functions defined for the $n-2$ first angular dimensions. Similar rules apply to the last angular dimension with only one consideration that the period is 2π instead of π. Since $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$ represent the orthogonalization process, they are decisive for this process and will be called as *orthogonalization points*. So we call these two points as well as the two extreme boundary points as *target points* of the angular dimensions in the following sections.

The simplest candidate function for f_i, which connects all the four target points, including orthogonalization points and extreme boundary points, in the 2-D plane would be a linear spline as shown in Fig.4.2.
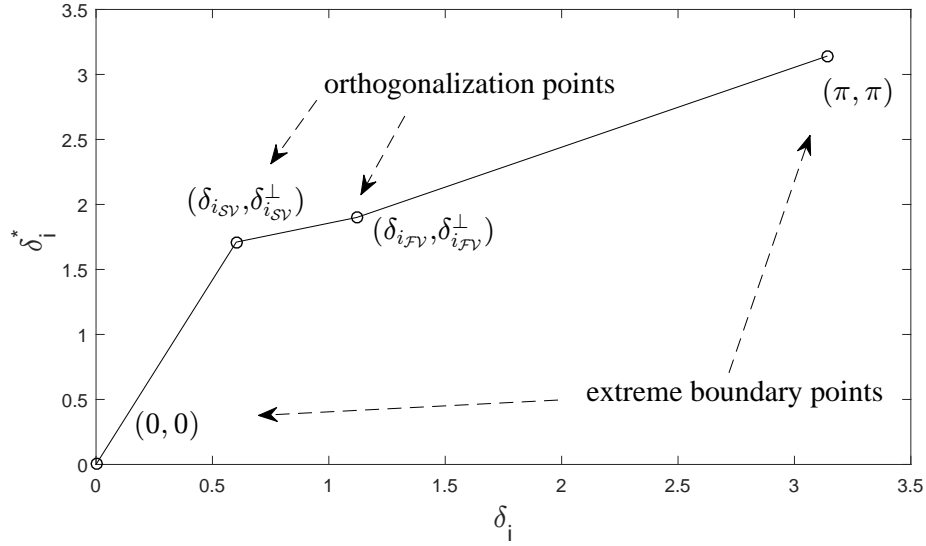
Figure 4.2: The simple mapping function for one angular dimension which maps the target points in the original space to the desired target points. Target points include the two extreme boundary points $(0, 0)$, and (π, π) as well as $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$ to yield the desired mapping.

4.5   Optimized Mapping Function

The mapping function in Fig.4.2 exhibits the two desired properties of monotonicity and continuity for an ideal mapping function f_i. However, this simple piecewise linear mapping function suffers from some drawbacks. Firstly, the function is not differentiable at the orthogonalization points $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$, which can lead to severe cluster deformation. Secondly, the mapping function is applied to the angular dimensions, the discontinuity of the derivative of the function can cause the convex-clusters to be mapped into non-convex clusters in the Cartesian coordinate system. In order to avoid deformations and to preserve maximal similarity with the original clustering geometry, it is desired to use a more linear functions with continuous derivatives. Also, in order to provide maximal separation among clusters and to concentrate clusters into disjoint non-overlapping clusters, it is beneficial to map the regions between two target points into a region with equal or even smaller range in

the Cartesian coordinate system. In other words, it is desired to keep the samples close to the centroids in the mapped space as much as possible. This property is achieved by using nonlinear functions. However, there is a trade-off between the level of concentration and the linearity, which needs to be carefully addressed. In order to avoid deformation while providing maximal separation, an optimized mapping function is proposed in this section.

In order to accommodate satisfy the above-mentioned requirements, it is desired to find a function, which satisfies the following mathematical conditions:

- the function is differentiable everywhere (continuous first derivative);

- the derivative of the function is small at the target points, which correspond to the centroids of clusters in the original and transformed space, i.e. $\delta_{i_{\mathcal{SV}}}$ and $\delta_{i_{\mathcal{FV}}}$ in Fig. 4.3;

- the derivative of the function is large at the boundaries of two regions (point $(\epsilon_{i_{\mathcal{SV}}}, \epsilon_{i_{\mathcal{SV}}}^{\perp})$ in Fig. 4.3);

Therefore, we propose to use the basic function p with adjustable parameters, which satisfies the aforementioned conditions. Each target point is associated with a basic function. The function p is composed of two constitutive functions: $h(x)$ and $g(x)$, respectively defined in two regions: i) from the target point to the upper boundary point, ii) from the lower boundary point to the next target point. Before defining the two functions $h(x)$ and $g(x)$, we need to define the boundaries between two consecutive target points $(\delta_{i_{\mathcal{SV}}}, \delta_{i_{\mathcal{SV}}}^{\perp})$ and $(\delta_{i_{\mathcal{FV}}}, \delta_{i_{\mathcal{FV}}}^{\perp})$, in which these functions are defined.We simply choose the midpoint as the boundary points. For instance, we have $\epsilon = (\delta_1 + \delta_2)/2$ as stated in Eq. 4.7. The lower and upper boundary points are noted respectively as $(\gamma, \gamma^{\perp})$ and $(\epsilon, \epsilon^{\perp})$. To ensure the continuity of the mapping function f, the boundaries and target points should satisfy:

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

The two piece-wise functions $h(x; \alpha, \epsilon, \epsilon^{\perp}, \delta, \delta^{\perp})$ and $g(x; \alpha, \gamma, \gamma^{\perp}, \delta, \delta^{\perp})$ are the inverse of logit function defined as follows:

$$K_h = \frac{\epsilon^{\perp} - \delta^{\perp}}{e^{\alpha(\epsilon - \delta, 0)^+} - 1} \tag{4.8}$$

$$h(x; \alpha, \epsilon, \epsilon^{\perp}, \delta, \delta^{\perp}) = K_h[e^{\alpha(x - \delta, 0)^+} - 1] + \delta^{\perp}$$

$$K_g = \frac{\gamma^{\perp} - \delta^{\perp}}{e^{\alpha(-\gamma + \delta, 0)^+} - 1} \tag{4.9}$$

$$g(x; \alpha, \gamma, \gamma^{\perp}, \delta, \delta^{\perp}) = K_g[e^{\alpha(\delta - x, 0)^+} - 1] + \delta^{\perp}$$

In order to investigate the satisfaction of the above-mentioned conditions, we first verify that these functions pass through the target points. More specifically, $g()$ should pass through points $(\gamma_i, \gamma_i^{\perp})$, $(\delta_i, \delta_i^{\perp})$ and $h()$ pass through $(\epsilon_i, \epsilon_i^{\perp})$, $(\delta_i, \delta_i^{\perp})$

Once, we confirm that that the basic function $h(x)$ has the desired properties, we fit this function with appropriate parameters to all regions. The final function is a smooth and continuous function, which passes through all target points $(\delta_{i_{SV}}, \delta_{i_{SV}}^{\perp})$, $(\delta_{i_{FV}}, \delta_{i_{FV}}^{\perp})$ as well as the two extreme boundary points as shown in Fig.4.4.

In the training process, data samples with abnormal labels belonging to all patients in DS1 and the personalized normal cluster for one patient are used to determine the mapping function for one corresponding patient. In the predicting process, the hyper-the the feature
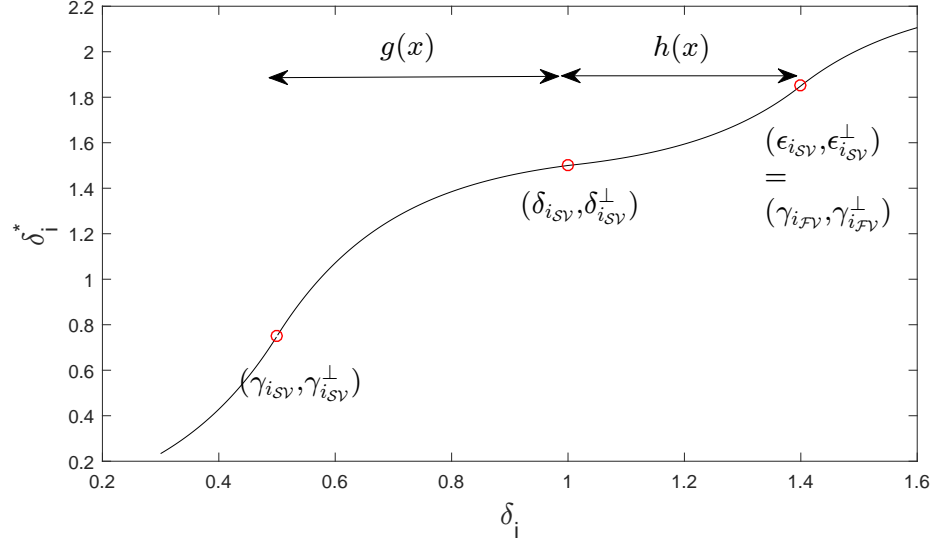
Figure 4.3: Optimized Piece-wise Interpolated Function p.

vector of new ECG sample in the spherical coordinate system is calculated and the mapping function is applied on its hyper-spherical coordinates to yield the transformed feature vector. After this step, we calculate the transformed data samples in the Cartesian coordinate system, which further fed into the personalized classification stage defined by Eq. 2.8 to generate the corresponding type of yellow alarm.

4.6   Experimental Results

In this section, the performance of the proposed method is evaluated in terms of two aspects. We first analyze the classification performance of the system and then present the comparative results with respect to other representative ECG classifiers. Furthermore, the classification results are partitioned into two sets: red alarms generated by the global classifier and the final labels by combining the yellow and red alarms. In this way, the impact of personalized classifier on the final labels can be revealed. Finally, the prediction power
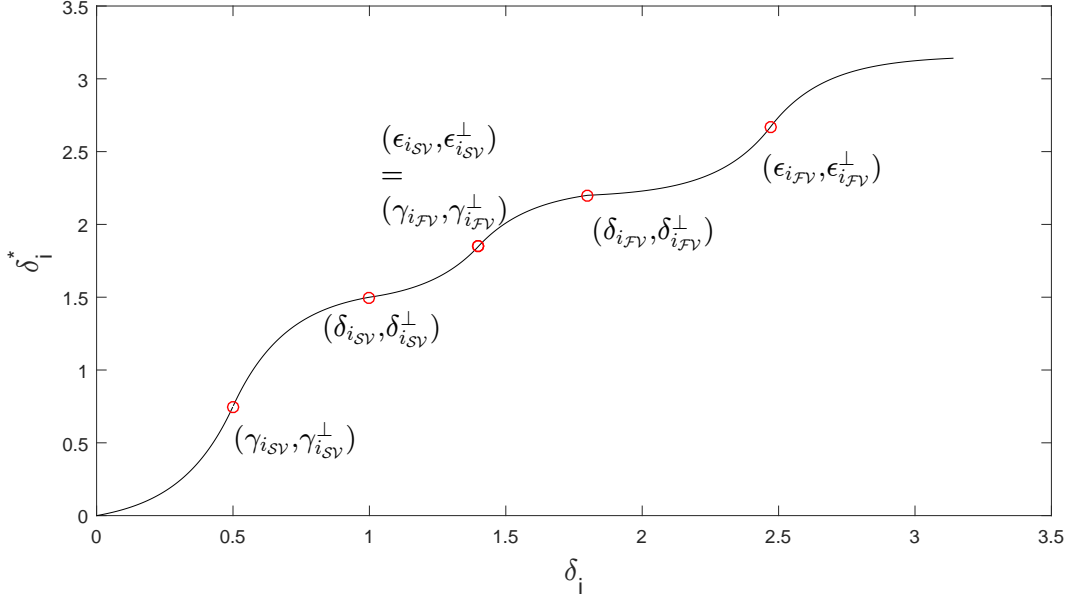
Figure 4.4: Optimized Mapping Function f.

of the proposed method in terms of providing precaution hints for upcoming red alarms is evaluated.

4.6.1 Classification Performance

The experimental results are assessed in terms of 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 the personalized classifier, the first 20% of the normal samples serve as initialization set for the personalized dynamic normal cluster. Therefore, we exclude the first 20% of normal samples from each record since they are already included in the training process. 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 the weighted k-Nearest Neighbors method with $k = 10$ as our choice of global classifier because it is one of the low-complexity methods with a relatively

Table 4.1: Cumulated Confusion Matrix for All Records in DS2. The numbers are *final label(primary label by global classifier)*.

		Ground Truth			
		N	V	S	F
	N	9255(10076)	21(38)	72(90)	1(5)
Result	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)

good classification accuracy [67]. The parameter α, used in Eq. 4.8 in the deviation detection module is set to 1 for test purpose.

Table 4.1 summarizes the cumulated confusion matrix for all records in the test set. In order to compare the result of global classifier (red alarms) and combined results (final labels including both red and yellow alarms), the sample numbers are presented in the following format: *final label(primary label by global classifier)*. In order to measure the classification performance, we adopt three metrics: accuracy(Ac), sensitivity(Se), specificity(Sp), as proposed in [12,13,16]. All three metrics are calculated based on the true positive TP, false positive FP, false negative FN and true negative TN in a binary confusion matrix, where one class is the specific abnormality class and all other abnormality and normal classes combined into one class. Therefore all four metrics are calculated for each class by converting the 4x4 matrix to a 2x2 matrix.

While cumulative 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. Further, 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 robustness of the system is assesed in terms of the variation between the performance of different types. The system is more robust if this variation is lower. In Table 4.2, we observe that among all abnormality classes, the proposed method demonstrates a stable performance on class V

but a lower stability for classes S and F.

As MITDB is widely used to verify ECG classifier performance, we compared the proposed system with five significant methods proposed in the literature. According to AAMI standards, the performance of ECG classification should be evaluated over the binary classifiers applied to *Ventricular (V)* versus *non-V* types and *Supraventricular (S)* versus *non-S* types. For methods proposed in the literature, the same evaluation metrics are commonly applied to records from MITDB. To standardize the metrics, we select 11 ECG records which are common among all 5 methods and compare the median of each classification metrics over these 11 records. The comparison results are presented in Table 4.3. Generally speaking, the proposed method shows a higher sensitivity for both types V and S. Especially for type S, the proposed method shows an advantage over all three metrics compared to the five reference methods.

4.6.2 Prediction Performance

As an important feature of the proposed method, yellow alarms triggered by the personalized classifier indicate a higher probability of observing subsequent abnormalities. In order to verify this functionality, all beats following a yellow alarm of a specific type is investigated to asses the chance of upcoming red alarms of different types. This process is repeated for yellow alarms of all types. We only account for the first abnormality type which occurs after the yellow alarm. As we used confusion matrix to evaluate the classification accuracy, the

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

performance of prediction can be summarized by a confusion matrix with the 3 abnormal types. Probabilities of observing a certain type of abnormal beat after a yellow alarm is calculated using the prediction confusion matrix and compared to the prior probability of observing the abnormality of the same type. 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 prediction power of each abnormality type 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 is higher than its prior probability and this fact is consistent for abnormality types. 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 after observing a yellow alarm of a type V is 77.45% (5.91% higher than the prior probability). The improvement are consistent among all three types of abnormalities but the system shows a stronger prediction power for type S.

Table 4.4: predictive probability versus prior probability without windowing

		# of predicted ground truth			% of predicted ground truth		
		V	S	F	V	S	F
	V	467	122	14	**77.45**	20.23	2.32
yellow alarm	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**

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
	V	290	85	12	**74.94**	21.96	3.10
yellow alarm	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 the above analysis we consider the first subsequent red alarm regardless of the time passes since the preceding yellow alarm. In order to study the impact of the timing window (the time between the yellow alarm and the subsequent red alarm), we also studied a window of 10 consecutive samples following a yellow alarm. Similarly, the prior and posterior probabilities are compared to evaluate the performance of the prediction capacity as shown in Table.4.5.

Compared with the result without windowing, the prediction performance within a 10-beat window shows that the proposed algorithm can better predict the occurrence of abnormalities if a certain timing window is used. Especially for type S, the probability of observing a sample of type S within 10 beats after a yellow alarm of type S is 27.32%, i.e. the posterior probability rises to 37.14%. With almost 10% increase, it is 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 personalized classifier and deviation analysis. We believe that this concept is worthy of further more in-depth studies for different physiological signals.

4.7   Summary of Contributions

In this chapter, we proposed a novel deterministic spatial transformation. The purpose is to reshape the original feature space so that the clusters in the transformed feature space demonstrate symmetry and low within-cluster variance. This method is more tractable and analytical than the method proposed in Chapter 3.

The method utilizes hyper-spherical coordinate and orthogonalization to determine the parameters in basic functions of mapping function. The mapping function is deterministic. We applied this method on test dataset MITDB DS2 and obtained promising results. Both classification and prediction performance of the proposed method are analyzed in this chapter. The algorithm has been proven to be efficient in classifying samples and predicting upcoming abnormalities as well.

Chapter 5

Conclusions And Future Works

5.1 Conclusions

In this thesis, we consider an overlooked problem in the biomedical signal processing community. Several automated algorithms are designed by the biomedical engineering community to process physiological signals to assist health providers in diagnosing different disorders and making better therapy plans. However, an important concept of predicting severe heart conditions before their occurrence by processing physiological signals is not well studied. In this work, we developed a novel methodology as a primary step towards developing predictive diagnosis tools. We applied the proposed methodology into ECG signals to assess the power of the system in predicting upcoming heart problems. In this regard, we proposed a patient-adaptive ECG classification framework. The system has a two-staged hierarchical structure including a global classifier and a personalized classifier. Global classifier is designed to filter the samples with severe distortions represented by their abnormal waveforms by triggering red alarms. The samples classified as "normal" the by global classifier are delivered to the subsequent deviation detection module. In this stage, the personalized dynamic normal

66

cluster is constructed and used to specify the normal range for each patient's ECG signal. By comparing a sample with its personalized normal range, we use two joint conditions to decide if the sample is in a fuzzy state between the normality and abnormality conditions. If the sample fails to meet both of these conditions, a yellow alarm is triggered to provide predictive information about upcoming abnormalities. The samples that show considerable deviations from their ground normal are passed to the personalized classifier, which will label it as one of the three abnormal types. On the other hand, the samples without considerable deviations are further utilized to update the personalized dynamic normal cluster.

In Chapter 3, a kernel-based nonlinear transformation is proposed to address the problem of cluster topology in the original feature space. More specifically, the weighted combination of kernel functions are deployed in this method as a spatial transformation function. The desired topology is formulated in terms of two objective functions, so that the system is able to find the optimal coefficients of kernels by jointly optimizing these two functions. This non-convex multi-objective optimization is solved with MOPSO. In order to validate the improvement on spatial topology by introducing nonlinearities with kernels, we compared the Pareto front generated with linear combination of the original features and Pareto front produced in the transformed feature space with polynomial kernels. The results confirm that the kernel-based transformation allows more degree of freedom in optimizing the topology in optimizing according to the objective functions. Moreover, we applied this method to MITDB test dataset and obtained similar sensitivity and specificity results as proposed in the literature. More importantly, the predictive capability of yellow alarms is analyzed. The performance is quantified by comparing the prior and posterior probabilities for each type of yellow alarms. The comparison results show that a promising improvement has been achieved by applying the nonlinear transformation.

While the method in Chapter 3 demonstrates a capacity of predicting upcoming abnormali-

ties of ECG signals, it remains challenging to interpret the mechanisms of the proposed system and thus hindering the generalization of the proposed predictive warning methodology to similar biomedical signal processing applications. Therefore, the main objective of Chapter 4 is to develop a deterministic spatial transformation function, which is able to achieve the desired spatial topology with a more tractable analytical approach. Thus, we proposed a novel spatial transformation specifically designed to reshape the feature space according to angles between cluster centroids. In this method, between-cluster cosine distances are optimized through orthogonalization of cluster centroids using spherical coordinate. Meanwhile, within-cluster variance is reduced by a piecewise mapping function composed with designed basis functions. The basic function proposed in Chapter 4 has the property of saturating at the boundaries, which is similar to the inverse of logit function but yet more flexible. An advantage of deploying such basics function is that the cluster geometry is preserved after spatial transformation. We implement this novel transformation in the patient-adaptive classification framework, the performance of this system is evaluated with classification and prediction results on the test dataset. The classification results show that by triggering yellow alarms through this method, the *specificity* of abnormal types is improved. Especially for the type S (supraventricular), the proposed system performs better in identifying class S than all 5 similar benchmark methods reported in the literature. Moreover, compared to the method proposed in Chapter 3, this method improves the predictive capability of the developed system for all abnormal classes but the improvement for type S is most significant.

We also studied the time lag of between a yellow alarm and the subsequent real abnormality in Chapter 4. The result shows that most of the real abnormalities occur within 10 beat after a yellow alarm. Generally speaking, the system has been proven to be efficient both in classification and prediction in this work. In short, this study suggest that predictive modeling of physiological signals can be used as alarming hints for upcoming health conditions, which

can have a wide range of applications in developing wearable biosensors, automated diagnosis tools to assist patents and physicians in predicting health problems, This methodology have a great potential to impact the emerging research fields of smart health and smart cities.

5.2   Future Works

In this research, we focused on improving two main drawbacks of automated ECG analysis in the literature, namely, the failure in capturing the inter-patient variability and the incapability of early detection and prediction. We proposed two methods for improving predictive capability. While the results show the promising performance of the designed system, further investigations can help to improve and generalize the proposed system to other types of biomedical signals. The following tasks can be considered as some future direction enabled by this research:

- Investigate other kernel functions to improve the transformation for spatial topology optimization.

- Seek deterministic solutions and develop more systematic performance analysis for the objective functions proposed in Chapter 3.

- Assess the performance of the proposed spatial transformation on other biomedical signals with similar properties, such as EEG, EMG, and EOG.

- Improve the deterministic mapping function in Chapter 4 by including clusters size in the mapping function.

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] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ecg classification by means of feature selection," *IEEE transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2168–2177, 2011.

[47] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, and J. M. Roig, "Principal component analysis in ecg signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 074580, 2007.

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

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

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

[51] J. Weston, C. Watkins, *et al.*, "Support vector machines for multi-class pattern recognition.," in *Esann*, vol. 99, pp. 219–224, 1999.

[52] H. Masnadi-Shirazi and N. Vasconcelos, "On the design of loss functions for classification: theory, robustness to outliers, and savageboost," in *Advances in neural information processing systems*, pp. 1049–1056, 2009.

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

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

[55] 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.

[56] 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.

[57] 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.

[58] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and testing low-degree polynomial data mappings via linear svm," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1471–1490, 2010.

[59] J. C. Carr, R. K. Beatson, B. C. McCallum, W. R. Fright, T. J. McLennan, and T. J. Mitchell, "Smooth surface reconstruction from noisy range data," in *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 119–ff, ACM, 2003.

[60] 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.

[61] 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.

[62] M. R. Sierra and C. A. C. Coello, "Improving pso-based multi-objective optimization using crowding, mutation and-dominance," in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 505–519, Springer, 2005.

[63] C. A. C. Coello, "Recent trends in evolutionary multiobjective optimization," in *Evolutionary multiobjective optimization*, pp. 7–32, Springer, 2005.

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

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

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

[67] K. Hechenbichler and K. Schliep, "Weighted k-nearest-neighbor techniques and ordinal classification," 2004.