A Predictive System for Patient-Specific ECG Classification Using Nonlinear Transformation

Jiaming Chen[1] and Abolfazl Razi[2]

Abstract— **In this paper, a novel predictive modeling is proposed to predict upcoming heart abnormailities by processing representative features extracted from electrocardiogram (ECG) signal. The core idea behind the proposed method is to use a controlled non-linear transformation to project signal features into a higher-dimensional space with desired geometric properties. In particular, we reshape the clusters through penalizing the clustering non-symmetry such that the projected abnormal clusters of different types symmetrically encircle the projected normal cluster.**

An immediate utility of the proposed method is to characterize the deviation of ECG signal samples towards different patient-specific abnormality classes. Moreover, this method can be used to enhance our prediction about potential heart problems before their occurrences. We applied this method to publicly available MIT-BIH dataset with 3 abnormality classes and the results suggest, respectively, 8%, 9% and 12% improvement in predicting the three abnormality classes. The proposed framework is general and applicable to a broad range of biomedical signals.

I. INTRODUCTION

In today's health care industry, informatics and computer science techniques are frequently used to assist physicians with better diagnosis practice. As an important application of biological signal processing, automated classification of electrocardiogram (ECG) has been investigated by researchers in the past decades [1]–[3]. Since abnormal cardiac cycles can indicate potential life-threatening diseases (e.g. arrhythmia), such systems are often used in intensive care unit, wearable monitor kit, etc. to provide continuous and real-time monitoring.

The common spirit of heart monitoring methods is to use large datasets of annotated ECG signals to model the impact of various heart disorders on the ECG signal morphology and use the developed models to recognize heart disorders for test signals. Several conventional classification algorithms, such as Fuzzy C-Means Clustering (FCM) [4], Support Vector Machine (SVM) [5], Artificial Neural Network (ANN) [6], are reported for ECG signal classification.

Despite a long list of proposed methods, the classification of ECG signals still remains challenging. One of the critical issue involved here is due to the morphological variability of ECG waveforms even among healthy individuals [1], [7]. This large variations are related to body habits, gender, age, and etc., which can influence the ECG recordings significantly [8], [9]. Another source of variation is the temporal variations of the ECG signal due to the physical activity of the patient, since the heart functions is largely influenced by the physical activity of patient, which may generate a large number of false alarms [1]. In order to improve the classification performance when presented with ECG waveforms from different patients, several patient-adaptable classification systems have been proposed, such as Mixture of Expert [7], generic neural network structure [6], Block-Based neural networks [10], expert assistance approach [11], and convolutional neural networks [12]. Therefore, to standardize the evaluation of ECG classifiers, the Association for the Advancement of Medical Instrumentation (AAMI) [13] has proposed standards for database, cardiac patterns and performance reporting for arrhythmia detection algorithms.

One important remaining issue for the approached mentioned above is the equal penalization of normal and abnormal cardiac classes. Since misclassification of abnormalities deserve a heavier penalty, heartbeats with deviations towards abnormal types should be distinguished from normal beats. In [14], a hierarchical classifier for personalized ECG diagnosis is proposed, which provides two level alarms (yellow and red) based on the severity of deviations from the normal class centric. This method, can be generalized to provide multi-level alarms based on the level of deviation. However, the non-symmetry of abnormal classes prevents a robust characterization of tendencies towards different abnormality classes.

In this paper, we propose a novel prediction framework which facilitates a robust deviation analysis. More specifically, in the proposed method, the accumulated ECG signals of different abnormalities in training set are used to recognize the geometry of normal and abnormal classes. Subsequently, a controlled nonlinear transformation is proposed in order to project data into higher order dimension with desired geometric properties. In particular, we reshape the clusters in the projected space by controlling the mapping coefficients, such that different abnormality classes symmetrically encircle the normal class. The patient-adaptable coefficients for this nonlinear transformation are tuned using a multiobjective particle swarm optimization (MOPSO) [15], where two objective functions balance between the clustering *cohesion* and *symmetry*.

The rest of this paper is organized as follows: Section II presents the signal preprocessing and feature extraction stage; Section IV depicts the general outline of the proposed hierarchical system; Section III elaborates on the proposed nonlinear cluster reshaping approach; SectionV presents the experimental result of this method followed by concluding

[1] School of Informatics, Computing and Cyber Systems,Northern Arizona University, Flagstaff, Arizona 86001 jiaming_chen@nau.edu

[2]School of Informatics, Computing and Cyber Systems,Northern Arizona University, Flagstaff, Arizona 86001 abolfazl.razi@nau.edu

remarks and discussions in section VI.

II. ECG DATA PROCESSING

In this work, the benchmark database, MIT-BIH arrhythmia database [16] (MITDB) has been used to obtain experimental results and evaluate the performance of our approach. This database includes 48 recordings in total, covering various of arrhythmias and also normal ECG waveforms with a uniform sampling frequency of 360Hz. According to the American Association of Medical Instrumentation (AAMI)'s recommended practice and suggestions provided in [17], we keep 44 tapes from MITDB and divide them into two non overlap subsets (DS1 and DS2), each includes 22 tapes and approximately 50 000 cardiac cycles. Furthermore, the expert annotations in MITDB are combined into five classes,namely N class(normal and bundle branch block beats), V class(Ventricular Ectopic beats), S class(Supraventricular Ectopic beats), F class(fusion of N and V beats) and Q class(unknown beats and paced beats). Taking the observation numbers in each class into consideration, we discard Q class in our following analysis for it contains only seven beats.

A. Preprocessing

In this work, the Daubechies wavelet of order 8 is used for ECG signal denoising, for it is proven to be more efficient than other basis functions [18]. If an ECG signal with a different sampling frequency is used, we can find the number of decomposition layers L to obtain the same level of smoothness as follows:

$$L = 1 + \lfloor \log_2 \frac{Fs}{360} \rfloor \tag{1}$$

Where, $\lfloor x \rfloor$ is the floor function and represents the largest integer number lower than x. By discarding detail coefficient d_1 , the electrosurgical and muscle noise which is the 100khz-10MHz range [3] is eliminated. Furthermore, a five ordered fitted polynomial is subtracted from the ECG signal to remove low frequency noise.

B. Segmentation and Feature Extraction

A cardiac cycle is depicted by its five fiducial peaks: P, Q, R, S and T [19]. The accurate detection of QRS complex, the most dominant peak within one cycle, is critical for the subsequent analysis. Cardiac cycles typically exhibit QRS complex in the frequency range of 5 to 22Hz [20]. Applying wavelet decomposition to signals with sampling frequency of 360Hz, the detail coefficients of level 4 (d_4) and level 5 (d_5) include the information of QRS complexes. Based on maximal overlap discrete wavelet transform (MODWT), the algorithm proposed in [16] and [21], is applied to detect R peaks in this work. The mother wavelet $db4$ is selected due to its morphological similarity to QRS complexes. Other fiducial peaks (P, QRS onset, Q, S, QRS offset and T waves for each cardiac circle) are localized according to the algorithm suggested by [20].

TABLE I
THE OVERVIEW OF MITDB AS TRAINING AND TEST DATASET.

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

TABLE II
FEATURES EXTRACTED FROM ECG SIGNAL

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

To delineate ECG waveform and extract representative features while eliminating temporal noise terms, here we combine three consecutive cycles to form a representative segment, following the work in [14]. In the subsequent discussion, one segment is regarded as one sample associated with one class label. Table.I shows the number of samples in each class for DS1 and DS2.

As discussed in [1], [17], [20], [22], a collective analysis of features in three categories, namely temporal, morphological and frequency domain, reflect the main differences between arrhythmia types. Table.II summarizes the extracted features including cycle-based (SET 1) and segment-based (SET 2) features, where we use the mean and standard deviation of cycle-based features. Therefore, for each segment, a set of 22 features are extracted and normalized to yield zero mean unit variance. In short, the mathematical expression of segmentation and feature extraction process is as shown in Eq.**??**, where $s(t)$ is the ECG time series, $\boldsymbol{T}^n$ is the range of time t for segment n, x is a feature vector and $g()$ is the mapping function.

$$x^n = g(s(t)) \quad t \in \boldsymbol{T}^n \tag{2}$$

III. NONLINEAR RESHAPING

Each segment is represented by a vector $\boldsymbol{x} \in Omega$ in the original feature space associate with a discrete true label $y \in \Psi$. The goal of classification is to design a function $\hat{y} = f(\boldsymbol{x})$, such that the pairwise distances between the true and predicted labels $d(\hat{y}, y)$ is minimized for training samples based on desired distance metric. Whereas in reality, especially for biomedical signal, deviation analysis for samples within one class is desired as well. In other words, we introduce secondary classes for distinguishing a sample labeled as N class but deviating towards X class and a sample with the same label but no deviation tendency, where N and X stands for normality and abnormalities

respectively. For this purpose, a classifier which provides a higher resolution output is in need.To achieve this goal, several assumptions and definitions of secondary classes are given as follows.

Given a discrete time series $\boldsymbol{x_n}(n = 0, 1, 2 \dots)$ and its corresponding labels $y_n \in \Omega$, where Ω can be divided to normality N and several abnormalities ($\Omega = \{N, X_1, X_2, \dots\}$). We define secondary classes for abnormalities as follows: if $\boldsymbol{x}_{k+i} (i > 0)$ is labeled as X_a and $\boldsymbol{x}_k$ is labeled as N, then there exist a secondary class X'_a corresponds to least one sample $\boldsymbol{x}_{k+j}$ where $0 < j < i$ (as show in Eq.3). Furthermore, we assume that the probability of observing abnormality X_a at time $k + i$ knowing that $y_{k+j} = X'_a$ and $y_k = N$, where $j < i$, is higher than the probability of prior probability of observing abnormality X_a (as shown in Eq.4). To distinguish these two types of abnormalities, we call the original abnormal classes primary abnormalities and defined secondary abnormal classes the secondary abnormalities.

$$\begin{aligned} \forall \quad \boldsymbol{x}_{k|y_k=N} \quad \text{and} \quad \boldsymbol{x}_{k+i|y_{k+i}=X_a} \\ \exists X'_a, 0 < j < i \quad \text{such that} \quad \boldsymbol{x}_{k+j|y_{k+j}=X'_a} \end{aligned} \tag{3}$$

$$\begin{aligned} P(y_{k+i} = X_a | y_{k+j} = X'_a, y_k = N) \geq P(y_{k+i} = X_a) \\ \text{where} \quad 0 < j < i \end{aligned} \tag{4}$$

More importantly, some associated properties are assigned to the secondary abnormal classes defined above to enable the development of an accurate classifier. Knowing that samples from primary classes and normal class are labeled and thus can be well studied, it's feasible to define secondary abnormalities based on the known properties of these clusters. The constraints are given in Eq.5, specifying that along the trajectory that a sample evolve from normality to one primary abnormality (X_a) it is more likely to pass through the same type secondary abnormality (X'_a) than other type secondary abnormalities (X'_b, where $b \neq a$)

$$\begin{aligned} P(y_{k+j} = X'_a | y_{k+i} = X_a, y_k = N) = \\ \max\{P(y_{k+j} = X'_b | y_{k+i} = X_a, y_k = N)\} \\ \text{where} \quad b = 1, 2, \dots a, a+1, \dots \end{aligned} \tag{5}$$

To obtain a geometric topology of normal and abnormal samples to meet the constraints, we consider a symmetric topology with centered normalities and surrounding abnormalities, as illustrated in Fig.2. The properties of this configuration can be concluded as follows:

- There is no overlap or crossover between trajectories from normality $\boldsymbol{z}_{k|y_k=N}$ to different primary abnormalities $\boldsymbol{z}_{k+i|y_{k+i}=X_a}$
- The lengths of trajectories i are identical for all primary abnormalities.
- Clusters of different classes in Φ are well separated.

As shown in Fig.1 while the geometric topology of clusters in Ω depends on feature extraction stage $g()$, the original data may not form such geometric topology, a integrated reshaping method is proposed along with the new classification system as described in Fig.1. Ψ' is the new output space which include normality, primary and secondary abnormalities. The intermediate layer $h()$ is add as the reshaping function reasoned above. Note that the output layer $o()$ is designed according to the geometric topology of samples locate in the reshaped space Φ. In short, $h()$ takes the data in original feature space and reshape the them mapping to a new feature space Φ, in which reshaped data meets the constraints defined in Eq.5. Using the properties inferred by Eq.5, output layer $o()$ maps reshaped data to a certain class in the output space Ψ'

Commonly used in Support Vector Machine (SVM), the kernel method is proved to be efficient in reshaping feature space. Hence, we consider an intermediate layer as a nonlinear transformation define by Eq.6, where $\circ$ stands for element-wise production. D_Ψ is the dimensionality of transformed space Φ; $\boldsymbol{\psi}$ is the nonlinear transformation operator and $\boldsymbol{w}$ is the coefficient vector as shown in Eq.7.

$$\boldsymbol{z} = \boldsymbol{\Psi_w}(\boldsymbol{x}) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{D_\Psi} \end{bmatrix} \circ \begin{bmatrix} \psi_1(\boldsymbol{x}^T) \\ \psi_2(\boldsymbol{x}^T) \\ \vdots \\ \psi_{D_\Psi}(\boldsymbol{x}^T) \end{bmatrix} \tag{6}$$

$$z_i = w_i \times \psi_i(\boldsymbol{x}) \quad (i = 1, 2 \dots D_\Psi) \tag{7}$$

The observations $\boldsymbol{x}$ is the training data labeled as either normality N or primary abnormality X. For a certain nonlinear transformation $\boldsymbol{Psi}()$, we can tune the reshaping coefficient $\boldsymbol{w}$ in a way such that the training data in new feature space Φ has a symmetric topology described above. Therefore the optimization of $\boldsymbol{w}$ should be conclude as a multiobjective optimization problem with two objective function defined in Eq.9 and Eq.10, where cosine distances used in Eq.9 is defined in Eq.8. Assuming we have $k + 1$ classes in total, normality is noted as $N = X_0$ and the rest k abnormal classes are noted as X_i. $\boldsymbol{c_N}$ and $\boldsymbol{c_{X_i}}$ are the centroids for normality and abnormality clusters respectively and $\boldsymbol{v_{NX_i}} = \boldsymbol{c_{X_i}} - \boldsymbol{c_N}$ are the vectors pointing from normality centroid to other abnormal class centroids. In Eq.10, $\boldsymbol{c_.} = \frac{\sum_i^n \boldsymbol{z_i}}{n}$ is mean of all data in Ψ, where $n = \sum_{i=0}^k n_i$ is total number of observations in training set. Both f_1 and f_2 are calculated using transformed data $\boldsymbol{z}$ with corresponding labels.

$$d(x, y) = 1 - \frac{xy'}{\sqrt{(xx')(yy')}} \tag{8}$$

$$f_1 = \frac{1}{\min(d(v_{NX_a}, v_{NX_b}) | a, b = 1, 2 \dots k; a \neq b)} \tag{9}$$

$$f_2 = \frac{SW}{SB} = \frac{\sum_{i=0}^k (n_i - 1) s_i^2 / n - k}{\frac{1}{k-1} \sum_{i=0}^k n_i (c_i - c_.)^2} \tag{10}$$

By minimizing f_1, the overlaps and crossovers between the trajectories from normalities to different abnormalities

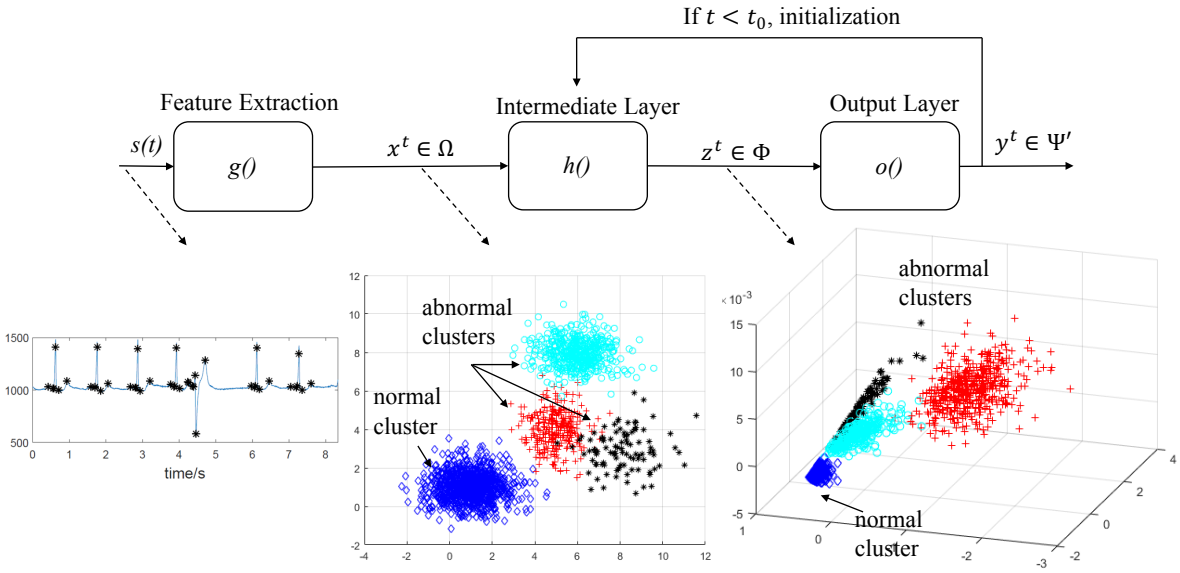

Fig. 1. Conceptual flow chart for nonlinear reshaping system

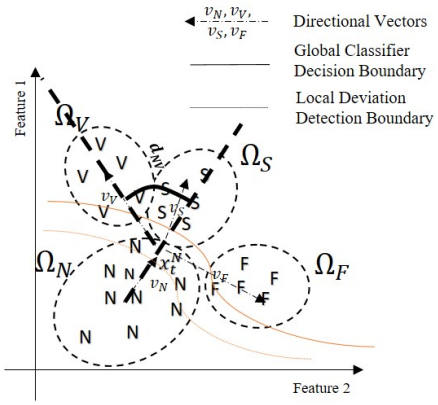

Fig. 2. A conceptual illustration for data spatial arrangement after nonlinear reshaping

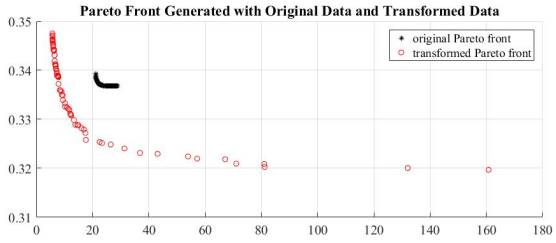

Fig. 3. Estimated Pareto front for before and after nonlinear transformation

are eliminated. Furthermore, the measure of f_1 and f_2 are both independent of dimensionality of Ψ. If the optimization method only optimize f_1, we will risk collapsing all data together. Thus adding f_2 as the second objective function is necessary for obtaining a well separated data.

Optimizing more than one objective functions at the same time is called Multiobjective Optimization. While numerous algorithms have been proposed for this subject, we adopt Multiobjective Particle Swarm Optimization (MOPSO)for its fast convergence and superior performance to other methods [15]. MOPSO approximates the Pareto front within objective function space. With two objective functions, the Pareto optimal particles can be plotted to 2-D plane. In order to compare the improvement of objective functions from original space $\boldsymbol{x}$ to the designed nonlinear space $\boldsymbol{z}$, the algorithm is tested on 2-D dummy data, as shown with illustration of feature space Ω in Fig.1, where each cluster corresponds to one class of real ECG signal.

Fig.3 shows that the estimated Pareto font of data after nonlinear transformation dominates the Pareto front of original data. The kernel used in Pareto front comparison is a 3-degree polynomial kernel. The result shows that when one objective function can not be well optimized in linear space, we may find a good solution in the higher dimensional space with nonlinear transformation. In Fig.1, the illustration of comparative cluster distributions in Ψ is generated with a 2-degree polynomial kernel for data visualization. The same conclusion can be drawn from this illustration as well, clusters are reshaped forming a structure close to the ideal case shown in Fig2.

IV. PREDICTIVE MODELING SYSTEM

In this section, a personalized predictive modeling with a hierarchical structure is proposed. As depicted in Fig. 4, the classification system is mainly composed by a global classifier(GC), a local classifier (LC) and a deviation detection stage using the methodology described in Section.III. According to the standard ECG classifier evaluation process proposed by AAMI [13], the classification system is trained and evaluated with MITDB. To be specific, MITDB is devided into two non-overlap subset: DS1 and DS2, as suggested in [17]. Each subset contains 22 records with expert annotation per cardiac cycle. DS1 serves as training set while DS2 serves as test dataset for this phase.

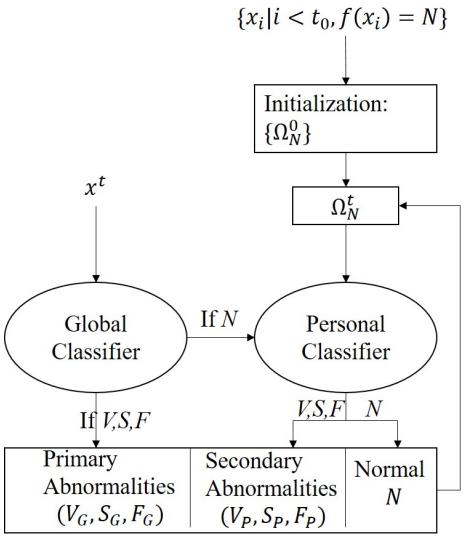

Fig. 4. Flow chart of the system

Regarding to the segmentation method described in section II, all cardiac cycles are firstly mapped into four classes: Normal, Supraventricular, Ventricular and Fusion [22]. Accordingly, a segment is labeled as Normal, if all member cycles are labeled as N, otherwise it is labeled as the only abnormal class (V, S or F) within the segment. If a segment includes two or more different abnormal classes, the segment is discarded. As a result, the total amount of segments for each class in each subset is presented in Table. I

A. Global Classifier (GC)

We first build a Global Classifier(GC) using all segments for every class in training set DS1 (denoted as $\Omega_{DS1}^N, \Omega_V, \Omega_S, \Omega_F$ for class N, V, S and F respectively). Taking the simplicity and efficiency into consideration, a k-Nearest Neighbors (kNN) (k is equal to 10 in this work) is selected. Nevertheless, the choice of GC does not have impact on the following steps.

For each patient, every segment in the received signal is processed by GC first, which attaches one of four major labels $N, V, S,$ or F to it. Denoting the segment at time t as x^t, the corresponding major label is denoted by G_t. For segments with abnormal labels ($G_t \in V_G, S_G, F_G$), the system adopts major labels as output ($y' = G_t$). Whereas the ones labeled as N_G, will be further processed by Local Classifier.

B. Local Classifier

1) Deviation Detection: Local Classifier is composed of two sequential steps: deviation detection and re-classification by nonlinear reshaping. These two steps are equivalent to a binary classifier (N or not N) and a multiclass classifier of three classes (V_p, S_p, F_p).

The deviation detection step is initialized by the first 20% of segments with true N label, marked as Ω_N^0 in Fig.2. Similarly, the abnormal clusters (Ω_V, Ω_S, Ω_F) consist of samples with corresponding labels in DS1. While in our experiment some record has limited number of N class segments, the system incorporates nearest neighbors of $\boldsymbol{c_N}$ from Ω_{DS1}^N into Ω_N^0 until we have 200 samples to initialize the classifiers. In order to perform the local analysis, the following distance metrics are calculated when a new sample segment x^t is labeled as N_g by GC.

$$R_{max} = \max_{x_i \in \Omega_N^t, x_j \in \Omega_N^t} \{\sqrt{(x_i - x_j)^2}\} \tag{11}$$

$$D_N(x^t) = \underset{x_N \in \Omega_N^t}{\text{median}} \{\sqrt{(x^t - x_N)^2}\} \tag{12}$$

$$D_V(x^t) = \underset{x_V \in \Omega_V}{\text{median}} \{\sqrt{(x^t - x_V)^2}\} \tag{13}$$

$$D_S(x^t) = \underset{x_S \in \Omega_S}{\text{median}} \{\sqrt{(x^t - x_S)^2}\} \tag{14}$$

$$D_F(x^t) = \underset{x_N \in \Omega_F}{\text{median}} \{\sqrt{(x^t - x_F)^2}\} \tag{15}$$

$$D_{max}(x^t) = \max_{x_N \in \Omega_N^t} \{\sqrt{(x^t - x_N)^2}\} \tag{16}$$

In order to trace the evaluation of normal heartbeat, Ω_t^N represents a dynamic cluster including aggregated normal samples until time t. A segment is labeled as $p_t = N$ when it meets Eq.17 and Eq.18 at the same time, otherwise the sample is mapped to one secondary abnormal classes. α is a tuning parameter, it controls the balance between classifier sensitivity and specificity. The detailed descriptions and the mechanisms of updating Ω_t^N can be found in [14].

$$D_{max}(x^t) \leq \alpha R_{max} \tag{17}$$

$$\min(D_N(x^t), D_V(x^t), D_S(x^t), D_F(x^t)) = D_N(x^t) \tag{18}$$

2) re-classification: The methodology presented in Section.III is incorporated in the system through this step. Using Ω_N^0 together with Ω_V, Ω_S, Ω_F as initialization data to represent the original data space, the nonlinear reshaping coefficient $\boldsymbol{e}$ is obtained by optimizing f_1 and f_2 with MOPSO for each patient individually. At the same time, the abnormal cluster (Ω_V, Ω_S, Ω_F) is projected to the same nonlinear space, as the new sample x^t is then transformed to z_t in a higher dimensional nonlinear space with Eq.6. Since f_1 is optimized in MOPSO, we adopt cosine distance to label P_t at this stage. As illustrated in Fig.2, v_N, v_S, v_V, v_F are indicates the relative deviating direction for sample z_t. Therefore the deviations for each abnormal class are calculated as $deviations_{X=V,S,F} \leftarrow d(v_N, v_X)$ and finally P_t takes the class which minimize the cosine distance. The details of this stage and structure of entire system are shown in the following algorithm.

$n \leftarrow 0.2 \times size(\Omega^N)$
$\Omega_N^0 \leftarrow \Omega^N(1:n) \cup \Omega_N^{DS1}(1:200-n)$
for $t = [1, \infty]$ **do**
  $x^t \leftarrow g(t)$ {*feature extraction and dimension reduction*}
  $G_t = mKNN(x^t)$
  if $G_t! = N_G$ **then**

$y \leftarrow G_t$
else
calculate $\{D_N(x^t), D_V(x^t), D_S(x^t), D_F(x^t), D_{max}(x^t), R_{max}\}$
condition1=$D_{max}(x^t) \leq \alpha R_{max}$
condition2=$\min(D_N(x^t), D_V(x^t), D_S(x^t), D_F(x^t)) = D_N(x^t)$
if condition1 is true and condition2 is true **then**
$P_t = N$
else
$z_t = h(x^t)$ {*nonlinear reshaping*}
$v_V \leftarrow c_V - z_t$
$v_S \leftarrow c_S - z_t$
$v_F \leftarrow c_F - z_t$
$v_N \leftarrow z_t - c_N$
$deviations_{X=V,S,F} \leftarrow d(v_N, v_X)$
$P_t = argmin\{deviations\}$
$y \leftarrow P_t$
end if
end if
end for

V. RESULT

In the experimental test, the original feature space is 8-dimensional. Thus we conducted a set of experiment with different nonlinear transformation. Taking the computational cost in high dimensional space into consideration and also the correlation between nonlinear terms, we present the result with the following nonlinear transformation (Eq.19), where we set the number of cross terms equals to the number of power terms. Therefore, the total dimensionality of Φ is $8 + 8 + 16 = 32$ dimensions.

$$\begin{aligned} \boldsymbol{z} = &\{x_i^2 | i = 1, 2 \dots 8\} \cup \{x_i^3 | i = 1, 2 \dots 8\} \cup \\ &\{x_i x_j | i, j = 1, 2 \dots 8, i \neq j\} \cup \\ &\{x_i^2 x_j | i, j = 1, 2 \dots 8, i \neq j\} \end{aligned} \tag{19}$$

Performance of the proposed system in this paper is tested with DS2 excluding record 232, for this record has only 7 N class segments. In total, 21 records are tested. For class N, we show the median, interquartile range (IQR), mean and standard deviation of Accuracy (AC), Sensitivity (SE); similarly for abnormal classes (V, S and F), we show Accuracy (AC) and Sensitivity (SE).

More importantly, the predictive capacity of the designed system worth evaluating separately. One approach is to count the number of samples labeled with secondary abnormal classes with a upcoming sample marked as true abnormal class. We define the probability of predicting abnormality X in Eq.20. Here we only take the samples which is classified as abnormal while its true label is N. These samples should not be considered as false alarm, since it can predicts the upcoming abnormalities.

$$\begin{aligned} probability\ of\ abnormalities(\theta) = &\frac{n_\theta}{n_V + n_S + n_F} \\ &\theta \in \{V, S, F\} \end{aligned} \tag{20}$$

TABLE III
OVERALL RESULTS FOR NORMAL AND ABNORMAL CLASSES.

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

TABLE IV
PREDICTIVE POWER OF YELLOW ALARMS:A YELLOW ALARM INCREASES THE CHANCE OF OBSERVING A RED ALARM OF THE SAME TYPE.

	Number of next abnormality				Probability of next abnormality (%)			
secondary abnormalities	V_p	S_p	F_p	Total	V_p	S_p	F_p	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
Total	4	20	0	24	100	100	100	100

The aggregated result for all 21 test records is in Table.IV. The last column of the Table.IV shows the probability of having the specific true abnormality X after any secondary abnormality. It's compared with the probability of having the same type of abnormality after a secondary abnormality of its own type. For instance, without the label of secondary abnormality, we have 67% of chance to experience an upcoming V, 20% for an upcoming S and 13% for a F. With the label of secondary abnormalities, the probability of having a true abnormal segment with the same type are enhanced to 75%, 29% and 25% for V, S, F, respectively. The result proves the trajectory assumption of predicting heart problem.

VI. CONCLUSIONS

In this paper, we proposed a novel patient-specific ECG classification system which classify heartbeats into seven categories, namely three primary abnormalities, three secondary abnormalities and normality. We reasoned this design with secondary abnormalities regarding to the geometric trajectory of sample beats deviating from normality towards abnormalities. This assumption was proved using the results in Table.IV. There is a higher probability to observe the same abnormality on the patient if there were this secondary abnormality recognized by our system.

While the trajectory theorem is proved in this paper, it has the assumption that clusters, both normal and abnormal, are convex in the transformed nonlinear space Φ. With this assumption, we will be able to capture the deviating direction correctly with cosine distance. This requires the new features

to be uncorrelated. Therefore, for other dataset, using another nonlinear transformation may be necessary.

Above all the proposed methodology demonstrates a potential to add detailed information about sample deviation upon conventional machine learning algorithm. We tested this system with ECG signal and observed promising results, but this method is not bound to this application. Further tests on other type of data will be conducted by out team as well.

ACKNOWLEDGMENT

REFERENCES

[1] 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*. IEEE, 2015, pp. 714–721.

[2] L. Lim and B. Yee, "Coach's companion-athlete's health monitoring system," *University of California, Berkeley*, 2005.

[3] 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.

[4] A. Dallali, A. Kachouri, and M. Samet, "Fuzzy c-means clustering, neural network, wt, and hrv for classification of cardiac arrhythmia," *ARPN Journal of Engineering and Applied Sciences*, vol. 6, no. 10, p. 2011, 2011.

[5] Z. Zidelmal, A. Amirou, D. Ould-Abdeslam, and J. Merckle, "Ecg beat classification using a cost sensitive classifier," *Computer methods and programs in biomedicine*, vol. 111, no. 3, pp. 570–577, 2013.

[6] 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.

[7] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 891–900, Sept 1997.

[8] I. Eisenstein, J. Edelstein, R. Sarma, M. Sanmarco, and R. H. Selvester, "The electrocardiogram in obesity," *Journal of electrocardiology*, vol. 15, no. 2, pp. 115–118, 1982.

[9] L. S. Green, R. L. Lux, C. W. Haws, R. Williams, S. C. Hunt, and M. 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] J. CHEN, H. PENG, and A. RAZI, "Remote ecg monitoring kit to predict patient-specific heart abnormalities."

[15] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Transactions on evolutionary computation*, vol. 8, no. 3, pp. 256–279, 2004.

[16] 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.

[17] 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, no. 7, pp. 1196–1206, July 2004.

[18] 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.

[19] L. Schamroth, "An introduction to electrocardiography," 1982.

[20] 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.

[21] G. B. Moody, "Evaluating ecg analyzers," *WFDB Applications Guide*, 2003.

[22] 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, Aug 2011.