

# Chapter 1

## *Kernels*-Based Nonlinear Spatial Transformation

### 1.1 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  $\Omega^d$  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 not 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.??) 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  $\Omega^d$  onto a new vector  $\mathbf{z}_k$  in a higher dimensional space denoted by  $\Omega^{d'}$ . This is achieved using a nonlinear mapping function  $\Phi^{d'} : \Omega^d \rightarrow \Omega^{d'}$ . The resulting vectors are used by the personalized classifier to identify the minor yellow alarm type out of  $\{\mathcal{N}, \mathcal{S}, \mathcal{V}, \mathcal{F}\}$ . The original text was ambiguous. Firstly, set and cluster shapes can be viewed differently, sets are not changes under the transformation although cluster geometry changes. Therefore, since we already used  $\mathcal{N}, \dots$  to represent clusters in the original space, you can use  $\mathcal{N}', \dots$  for the transformed clusters, unless you mean a set (not a cluster). In this case, sets do not changes under the transformation, and you can omit the definition  $\Phi^{d'} = \{\mathcal{N}, \mathcal{S}, \mathcal{V}, \mathcal{F}\}$  to avoid ambiguity. Secondly, you need to distinguish between two functions, one simply maps the vector in original space into a new space, and the second one processes the transformed vector and maps it to a yellow alarm. It is not clear in the text.

## 1.2 *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 [46]. 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  $\Omega$  into a Hilbert space  $\Phi$  via a nonlinear function  $\Psi()$  [?]. 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  $\sum_{x_i \in \mathcal{X}} (y_i - f(x_i))^2$  [47]. Other choices for loss function include Hingde loss, absolute loss, hit and miss loss, etc [REF].

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: better to use  $n$  instead of  $m$  if it does not conflict with other definitions.

$$\text{minimize } \frac{1}{m} \sum_{i \in \mathbb{N}} L(y_i, f(\mathbf{x}_i)) + \gamma ||f||^2, \quad (1.1)$$

where  $||f||^2$  is the squared norm of  $f$  here  $f$  is a function and norm is defined for vectors.

So add one or two sentences how to calculate it. I think for instance you mean the norm of coefficients of a polynomial function here. and the positive constant  $\gamma$ , 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  $\Phi$ . 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 [48].

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 [49]. An exhaustive search for all possible kernels is a computationally expensive and unrealistic task [?]. 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 [?]. 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 [?].

Polynomial kernel is usually applied on normalized data for its explicit expression and steady performance. However, the degree of freedom in defining a polynomial kernel is relatively high, which requires tuning a large number of parameters. In fact this statement is not so relevant here. Note that a polynomial kernel of order  $p$  can be defined as  $(1 + v^T w)^p = (1 + v_1 w_1 + v_2 w_2 + \dots)^p$ , which has only one free parameter  $p$  and all coefficients are automatically determined. However, we are going to use a polynomial function of order  $p$ , (i.e.  $f(v) = 1 + \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_i v_1^2 + \alpha_j v_1 v_2 + \dots \alpha_k v_N^p$ ) which obviously required tuning too many

free coefficients  $\alpha_i$ .

Gaussian kernel function denoted by XXX for vectors  $\mathbf{v}$  and  $\mathbf{w}$  is a very classic robust radial function, which has shown a string robustness in the case of noisy datasets [REF XXX]. 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 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:

$$\mathbf{z}_k = \Psi_{\mathbf{w}}(\mathbf{x}_k) = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d'} \end{bmatrix} \circ \begin{bmatrix} \psi_1(\mathbf{x}_k) \\ \psi_2(\mathbf{x}_k) \\ \vdots \\ \psi_{d'}(\mathbf{x}_k) \end{bmatrix}, \quad (1.2)$$

where  $\mathbf{w}$  is the vector of normalized coefficients do we have  $\|\mathbf{w}\| = 1$ . Instead of selecting kernel, the process of spatial geometry optimization is accomplished by adjusting the coefficients of fixed polynomial basis functions  $\psi(\cdot)$ . 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  $\mathbf{w} = [w_1, w_2, \dots, w_d]^T$  to achieve the ideal symmetric geometry in the reshaped feature space while maintaining the maximal separation between clusters.

## 1.3 Multiobjective Optimization

### 1.3.1 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  $\Omega^2$ . **This is not necessarily true, since you can choose any order of power even for a 2-D input vector. Therefore, the mapping base kernel may adopt a second-order polynomial function as follows:** We also assume for simplicity that the order of the polynomial function is 2. Therefore, we have:

$$\begin{aligned} \mathbf{x} &= [x_1 \ x_2]^T, \quad \mathbf{w} = [w_1 \ w_2 \ \dots \ w_5]^T, \quad d = 2, \quad d' = 5, \\ \psi_1(\mathbf{x}) &= x_1, \psi_2(\mathbf{x}) = x_2, \psi_3(\mathbf{x}) = x_1^2, \psi_4(\mathbf{x}) = x_2^2, \psi_5(\mathbf{x}) = x_1 x_2. \end{aligned} \quad (1.3)$$

**when you start the sentence with “where”, it means that the equation is connected to the following sentence, so it is better to add a comma to the end of equation. Otherwise, add dot “.”. I fixed some, but make sure that you do it consistently.**

Revised up to here

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

$$\begin{aligned}
o_1(\mathbf{w}) &= \frac{1}{\min_{c,d=2,\dots,p \text{ and } c \neq d} \{d(\mathbf{v}_{\mathcal{X}_c}, \mathbf{v}_{\mathcal{X}_d})\}} \\
o_2(\mathbf{w}) &= \frac{SW}{SB} = \frac{\sum_{c=1}^C \sum_{\mathbf{z} \in \mathcal{X}_c} (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})^T (\mathbf{z} - \mathbf{c}_{\mathcal{X}_c})}{\sum_{c=1}^C \sum_{d=1, d \neq c}^C (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})^T (\mathbf{c}_{\mathcal{X}_c} - \mathbf{c}_{\mathcal{X}_d})}
\end{aligned} \tag{1.4}$$

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

$$\mathbf{v}_{\mathcal{X}_i} = \mathbf{c}_N^k - \mathbf{c}_{\mathcal{X}_i} \tag{1.5}$$

Since, for some patients, the total number of a certain type of abnormal samples are very limited, abnormal samples in trainign set DS1 are utilized in calculating the two objective functions. In Eq.1.4, the abnormal cluster centroids are calculated using abnormal sample sets in training set DS1, while the normal cluster centroid is defined by the personal normal cluster.

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

### 1.3.2 Multiobjective Particle Swarm Optimization

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

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

#### Pareto Front

In order to compare different set of coefficient in this 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 [50].

## MOPSO

Among all the MOPSO algorithms in the literature, the algorithm proposed by Coello Coello and Lechug presents a better performance and lower computational complexity than most of the other MOPSO algorithms [50]. Therefore, this algorithm is implemented to solve the optimization problem in this work. One characteristic property of this algorithm is the 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 [50]. Fig.1.1, which is generated by jointly minimizing  $o_1(\mathbf{w})$  and  $o_2(\mathbf{w})$ , demonstrates the repository members are Pareto optimal than other particles and they converge to a uniform Pareto front.

With the concept of Pareto optimal, we further demonstrate the impact of applying kernel functions in this spatial reshaping problem by comparing the optimal solutions for coefficients of linear function combination and those for coefficients of polynomial kernel functions. Therefore, we first optimize the coefficients of third-order polynomial kernel functions, as formulated in Eq.1.3 and then optimize coefficients of linear features in origin feature space. The purpose is to verify if the objective functions can be fundamentally better optimized by introducing nonlinearities into the model.

As shown in Fig. 1.2, the estimated Pareto front of the nonlinear model using polynomial

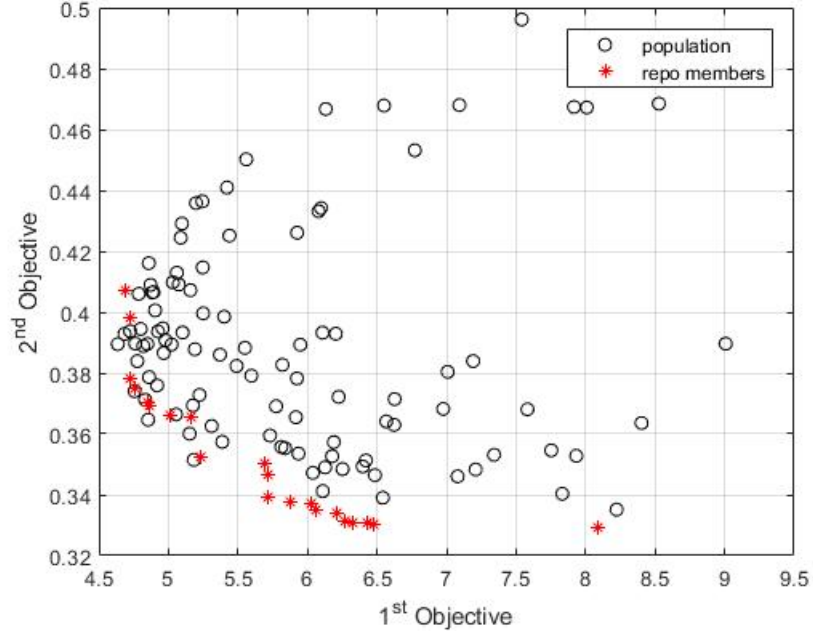


Figure 1.1: Particles stored in external repository approximate the Pareto front

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

## 1.4 Experimental Results

As mentioned in Section. 2.3, a cardiac segment is represented by an 8-dimensional vectors after feature extraction and PCA. To specify the nonlinear transformation in (1.2), the polynomial functions of order 3 is applied on the feature vectors. Moreover, taking the overfitting problem in high-dimensional space and computational cost into account, only 32 terms, which include 8 square terms  $x_i^2$ , 8 cubic terms  $x_i^3$ , 8 cross terms  $x_i x_j$  of second order

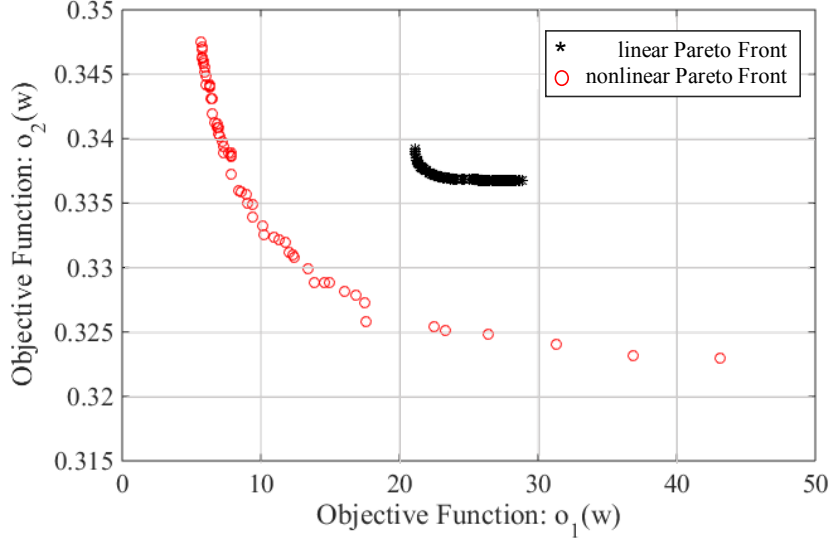


Figure 1.2: Increase of degree of freedom in optimization is proved by comparing the Pareto fronts generated by linear and nonlinear basis function

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

$$\begin{aligned} \mathbf{z} = & \{x_i^2 | i = 1, 2 \dots 8\} \cup \{x_i^3 | i = 1, 2 \dots 8\} \cup \\ & \{x_i x_j | i, j = 1, 2 \dots 8, i \neq j\} \cup \{x_i^2 x_j | i, j = 1, 2 \dots 8, i \neq j\} \end{aligned} \quad (1.6)$$

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 1.1 shows the performance of the proposed method in classifying ECG signal segments. In order to evaluate the consistency and general classification results over all recordings, the median, interquartile range (IQR), mean and standard deviation of accuracy (AC), sensitivity (SE) and specificity (SP) are presented as evaluation metrics. The results

Table 1.1: 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

are promising and the median of accuracy for all classes are in the range of 88% – 99%. Sensitivity and specificity of the proposed method exhibits similar ranges. The mean accuracy is at least 86% excluding class V. Therefore, this system is not likely to miss an important alarm or to report false alarms.

More importantly, the predictive capability of the proposed method is worthy of evaluating exclusively for it is the featuring specialty of the proposed system. In order to quantify the posterior probability of observing an abnormal signal after a preceding yellow alarm of similar type in (??), the number of predicted samples are counted as formulated in Eq 1.7:

$$\begin{aligned}
 P(\hat{y}_{k+i} = X_r | \hat{y}_k = X_y) &= \frac{\# \text{ of } y_{k+i} = X \text{ after } \hat{y}_k = X_y}{\# \text{ of true alarms after } \hat{y}_k = X_y} \\
 P(\hat{y}_{k+i} = X_r) &= \frac{\# \text{ of true alarm of type } X (y_k = X)}{\# \text{ of all true alarms}}
 \end{aligned} \tag{1.7}$$

The summary of results for all 21 test records is presented in Table. 1.2. The subsection *Probability of next abnormality (%)* in Table. 1.2 shows the confusion matrix of the proba-

Table 1.2: 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

bility of having a subsequent true abnormality of all types after observing a yellow alarm of all types along with the prior probability of observing a certain type of abnormal sample in the very last column.

These results validate the predictive capability of yellow alarms. For instance, the prior probability of observing a sample segment with abnormal types  $V$ ,  $S$ , and  $F$  is respectively  $\frac{96}{96+29+18} = 67\%$ ,  $\frac{29}{96+29+18} = 20\%$  and  $\frac{18}{96+29+18} = 13\%$ , based on their relative frequencies in the dataset. However, the corresponding posterior probabilities after observing a yellow alarm of type  $Vp$  are respectively  $\frac{38}{38+11+2} = 75\%$ ,  $\frac{11}{38+11+2} = 21\%$  and  $\frac{2}{38+11+2} = 4\%$ . This means that the probability of observing a real abnormal segment of type  $V$  is  $75\% - 67\% = 8\%$ , which is higher than its prior probability. The same trend holds for other yellow alarms as well. The results suggest a more in-depth study of the concept of yellow alarms for heart monitoring. Therefore, we investigate in improving the spatial transformation module of the system and meanwhile reduce the computational cost in Chapter 4.

## 1.5 Conclusions

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

Result shows that approximated Pareto front produced by kernel method in the objective function space is apparently optimal to the one which is produced by the linear combinations of original features. The results verify that the proposed method has a classification accuracy in the range of 88% – 99% for different ECG records in the test set of MIT-BIH database.

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

# 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 statistics 2010 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. D’Agostino, 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. Pongpongsri and X.-H. Yu, “An adaptive filtering approach for electrocardiogram (ecg) signal noise reduction using neural networks,” *Neurocomputing*, vol. 117, pp. 206–213, 2013.
  - [39] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, “Ecg beat detection using filter banks,” *IEEE transactions on biomedical engineering*, vol. 46, no. 2, pp. 192–202, 1999.
  - [40] D. Sadhukhan and M. Mitra, “R-peak detection algorithm for ecg using double difference and rr interval processing,” *Procedia Technology*, vol. 4, pp. 873–877, 2012.
  - [41] S. Mehta and N. Lingayat, “Svm-based algorithm for recognition of qrs complexes in electrocardiogram,” *IRBM*, vol. 29, no. 5, pp. 310–317, 2008.
  - [42] R. V. Andreão, B. Dorizzi, and J. Boudy, “Ecg signal analysis through hidden markov models,” *IEEE Transactions on Biomedical engineering*, vol. 53, no. 8, pp. 1541–1549, 2006.
  - [43] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, “A wavelet-based ecg delineator: evaluation on standard databases,” *IEEE transactions on biomedical engineering*, vol. 51, no. 4, pp. 570–581, 2004.
  - [44] S. Banerjee, R. Gupta, and M. Mitra, “Delineation of ecg characteristic features using multiresolution wavelet analysis method,” *Measurement*, vol. 45, no. 3, pp. 474–487, 2012.

- [45] Z. Zidelmal, A. Amirou, M. Adnane, and A. Belouchrani, “QRS detection based on wavelet coefficients,” *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 490–496, 2012.
- [46] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge university press, 2004.
- [47] B. Schölkopf, C. J. Burges, and A. J. Smola, *Advances in kernel methods: support vector learning*. MIT press, 1999.
- [48] T. Evgeniou, M. Pontil, and T. Poggio, “Regularization networks and support vector machines,” *Advances in computational mathematics*, vol. 13, no. 1, p. 1, 2000.
- [49] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [50] C. A. Coello Coello, “Mopso: A proposal for multiple objective particle swarm optimization,” *Proc. Congr. Evolutionary Computation (CEC’2002), Honolulu, HI, 5*, vol. 1, pp. 1051–1056, 2002.
- [51] J. E. Alvarez-Benitez, R. M. Everson, and J. E. Fieldsend, “A mopso algorithm based exclusively on pareto dominance concepts,” in *International Conference on Evolutionary Multi-Criterion Optimization*, pp. 459–473, Springer, 2005.
- [52] L. Blumenson, “A derivation of n-dimensional spherical coordinates,” *The American Mathematical Monthly*, vol. 67, no. 1, pp. 63–66, 1960.
- [53] G. W. Stewart, *Matrix algorithms volume 1: Basic decompositions*, vol. 2. Society for Industrial and Applied Mathematics, 1998.
- [54] G. Arfken, “Gram-schmidt orthogonalization,” *Mathematical methods for physicists*, vol. 3, pp. 516–520, 1985.