

Intelligent Arrhythmia Detection using Genetic Algorithm and Emphatic SVM (ESVM)

Jalal A. Nasiri^{a*}, Mostafa Sabzekar^a, H. Sadoghi Yazdi^a, Mahmoud Naghibzadeh^a, Bahram Naghibzadeh^b

*Commutation and Computer Research Center, Ferdowsi University of Mashhad, Mashhad, Iran **

Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran^a

School of Medicine, Mashhad University of Medical Science, Mashhad, Iran^b

{j.nasiri,sabzekar}@wali.um.ac.ir, {sadoghi, naghibzadeh}@um.ac.ir, naghibzadeh@mums.ac.ir

Abstract— In this paper, a new method of arrhythmia classification is proposed. At first we extract twenty two features from electrocardiogram signal. We propose a novel classification system based on genetic algorithm to improve the generalization performance of the SVM classifier. For this purpose, we have optimized the SVM classifier design by searching for the best value of the parameters that tune its discriminate function, and looking for the best subset of features that feed the classifier. We select appropriate features with our proposed Genetic-SVM approach. We also propose Emphatic SVM (ESVM), a new SVM classifier, with fuzzy constraints. It emphasizes on constraints of SVM formulation to give more ability to our classifier. We finally, classify the ECG signal with the ESVM. Experimental results show that our proposed approach is very truthfully for diagnosing cardiac arrhythmias. Our goal is classification of four types of arrhythmias which with this method we obtain 95% correct classification.

Keywords: ECG Arrhythmia, Support Vector Machines (SVM), Emphatic SVM, Fuzzy constraints, Genetic Algorithms, Feature reduction.

I. INTRODUCTION

For several years, the analysis of the electrocardiogram signal is the most effective and available method for diagnosing cardiac arrhythmias but the classification of an electrocardiogram (ECG) into different disease categories is a complex pattern recognition task. Computer based classifications of the ECG can achieve high accuracy and offer the potential of an affordable mass screening for cardiac abnormalities. Successful classification is achieved by finding the characteristic shapes of the ECG that discriminate effectively between the required diagnostic Categories. Conventionally, a typical heart beat is identified from the ECG and the component waves of the QRS, T and possibly P waves are characterized using measurements such as magnitude, duration and area. Datasets that are used for heart diseases involve different features. Some of them are based on laboratory experiments, while others include clinical symptoms. However, one of the most popular and useful databases is the MIT-BIH. Researchers have used this database to test their various algorithms for arrhythmia detection and classification. Several methods have been proposed for the classification of ECG signals. Among the most recently published works are those presented in [1]-[6]. The method represent in [1] based on *Fisher Linear discriminant*, they detected the RR interval duration and the distance between the occurrence of P wave and T wave.

Using these features they applied Fisher's Linear Discriminant. In [2] a SVM based method for PVC arrhythmia detection shown that has a better efficient rather than Anfis. In [3] a new approach based PSO-SVM has been proposed for feature selection and classification of cardiac arrhythmias. In [4], a neuro-fuzzy approach for the ECG-based classification of heart rhythms is described. Here, the QRS complex signal is characterized by Hermite polynomials, whose coefficients feed the neuro-fuzzy classifier. Detection of arrhythmia by means of Independent Component Analysis (ICA) and wavelet transform to extract important features is proposed in [5]. Finally, in [6], the authors present an approach for classifying beats of a large dataset by training a neural network classifier using wavelet and timing features. The authors found that the fourth scale of a dyadic wavelet transform with a quadratic spline wavelet together with the pre/post RR-interval ratio is very effective in distinguishing normal and PVC from other beats.

There are several other methods but here we focus on algorithms that similar to our work. The paper is structured as follows. In Section II we explain feature extraction and selection. Section III covers an overview of Genetic Algorithms. SVM and Multi-class SVM are briefly reviewed in Sections IV. The structure of the proposed Emphatic SVM (ESVM) is given in Section V. Section VI includes our proposed Genetic-SVM method. The effectiveness of the proposed approach is illustrated by experimental results in Section VII. Finally, Section VIII presents the concluding remarks and future work.

II. FEATURE EXTRACTION AND SELECTION

In this section we will explain characteristics of extracted feature from ECG and the procedure designed for this purpose. Figure 1, presents block diagram of proposed arrhythmia classification.

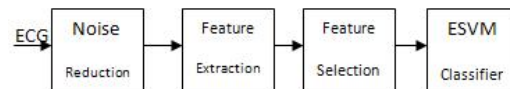


Figure 1: Block diagram of proposed arrhythmia classification

A. Dataset Description

Our experiments were conducted on the ECG data as the basic signal for classification. The annotated ECG records, available at the MIT-BIH arrhythmia database [20], have already been used frequently for the evaluation of different

classifiers in recent researches. The database has 48 records, each 30 minutes in Length. Each data was recorded in two channels, modified limb lead II and modified lead VI.

In particular, the considered beats refer to following classes: normal sinus rhythm (N), right bundle branch block (RB), left bundle branch block (LB), and paced beat (P). You can see sample of four N, RB, LB, and P in figure 2. The beats were selected from the recording of following patients, 100, 106, 107, 109, 111, 118, 202, 209, 212, 214, 215 and 217. In the table 1, you can find some information about number of beats in each category.

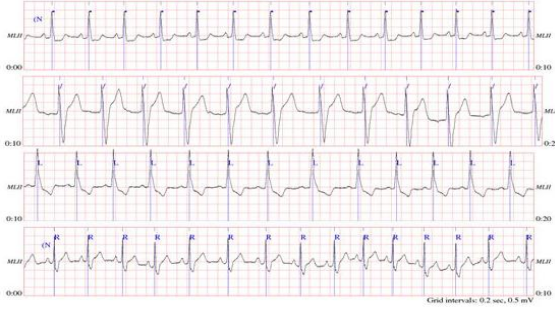


Figure 2: Sample signal of Normal, Paced, LBBB, and RBBB

Table 1: Data Set Descriptions and Numbers Used in the Simulation

Class No.	Record Example used from MIT-BIH	No. of beats used	Description
1.	100,105	243	<i>Normal(N)</i>
2.	107,217	110	<i>Paced(P)</i>
3.	111,214	600	<i>LBBB(LB)</i>
4.	118,212	450	<i>RBBB(RB)</i>

B. Noise reduction

In the first stage, we do wavelet transform for noise reduction. You can see the effect of this matter in Figure 3. First signal is the original and the second is achieved after noise reduction stage.

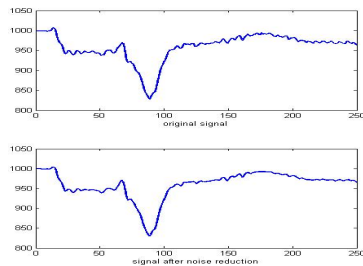


Figure 3: Sample signal, previous and after noise reduction

C. Feature Description:

Features are extracted as one feature vector for each of the beats in all records. Each vector includes one of the four possible labels. For feature extraction, we use nineteen temporal features such as R-R interval, PQ interval, PR interval, and PT interval and we use three morphological features.

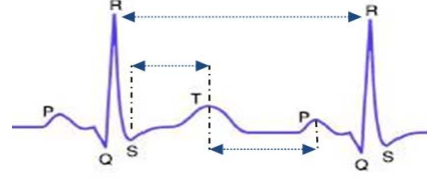


Figure 4: Sample feature: ST interval, TP interval and RR interval

Features have been extracted including the time and voltage of Q/R/S/T/P and time interval for each of five features from the next feature such as RS/ ST/ QR as a mentioned in figure 4 and also the difference of voltage in these features such as $V(Q)-V(S)$. Another feature that have considered is the time and voltage of RR. The description of the features has summarized in Table 2. $X(R)$ means the position of R in the ECG signal and $V(R)$ means the value of that position in the signal.

Table 2: Description of features used in the simulation

Feature NO.	Description	Feature NO.	Description
1.	$X(R1)$	11.	$X(R2)$
2.	$V(R1)$	12.	$V(R2)$
3.	$X(S)$	13.	$X(R2) - X(R1)$
4.	$V(S)$	14.	$V(R2) - V(R1)$
5.	$X(T)$	15.	$X(S) - X(R1)$
6.	$V(T)$	16.	$X(T) - X(S)$
7.	$X(P)$	17.	$X(P) - X(T)$
8.	$V(P)$	18.	$X(Q) - X(P)$
9.	$X(Q)$	19.	$X(R2) - X(Q)$
10.	$V(Q)$		

The three morphological features compute by finding maximum and minimum values of that beat in ECG signal. Signal of each beat scaled so as to range between zero and one. We considered percent that are higher than 0.2, 0.5 and 0.8 as three features.

All of the obtained features are based on six features that we got them using a semiautomatic method in the first stage. We suggest first and second R point to expert using an algorithm based on maximum-minimum. Then the expert distinguishes appropriate points (R, S, T, P, Q, and R).

III. GENETIC ALGORITHMS

Genetic Algorithm (GA) is one of meta-heuristic optimization techniques, which include simulated annealing, tabu search, and evolutionary strategies. GA has been demonstrated to converge to the optimal solution for many diverse and difficult problems as a powerful and stochastic tool based on principles of natural evolution [18]. The details of our implementation of GA are described as follows:

The first step in GAs is to define the encoding allowing describing any potential solution as a numerical vector, and then you can generate a population randomly. We briefly describe some concept and operation in GA.

Selection operator: The selection process selects individuals from population directly based on fitness values [19].

Recombination: The role of the crossover operation is to create new individuals from old ones. Crossover often is a probabilistic process that exchanges information between

some (usually two) parent individuals in order to generating some new child individuals.

Mutation Operator: Mutation is applied to one individual and produces a modified mutant child.

Fitness Function: The role of the Fitness function is to measure the quality of solutions.

IV. SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) [21] is a new powerful and popular machine learning method that delivers state of the art performance in real world data mining applications and deals with high dimensional data and provides good generalization

Algorithm1: Genetic Algorithm

Input: Training Data

Output: Useful Features

Step0: initialize parameters (e.g. population size, crossover rate, mutation rate and the maximum number of population generation.)

Step1: create initial population randomly ($P(0)$).

Step2: evaluate current population (compute fitness of all chromosomes).

Step3: while (termination condition not satisfied) do [step 4-8]

Step4: select $P(t)$ from $P(t+1)$ [perform selection]

Step5: recombine $P(t)$ [perform mutation and crossover]

Step6: evaluate current population (compute fitness of all chromosomes).

Step7: $t = t + 1$

Step8: go to Step 3

Algorithm1: Pseudo code for GA

properties. It also, has the potential to handle very large feature spaces and large classification problems [23]. SVMs have been applied on many fields, such as text classification, image classification, and bioinformatics and so on. In many ECG classification and arrhythmia detection such as [2, 3] SVM applied. In SVM, the original input space is mapped into high dimensional feature space. The optimal separating hyperplane is found by exploiting the optimization theory.

A. SVM-based classification

Assume $S = \{(x_i, y_i)\}_{i=1}^n$ is the training set where n is the number of input samples, $x_i \in \mathcal{R}^m$ is an m -dimensional input vector, and $y_i = \{-1, +1\}$ is the label of x_i . For linearly separable input data, we can determine a hyperplane

$$f(x) = w^T x + b = 0 \quad (1)$$

where w is an m -dimensional vector and b is a scalar. Function $\text{sign}(f(x_i))$ is the decision function for testing sample x_i . Considering the noise with slack variables ξ_i and error penalty term $C \sum_{i=1}^n \xi_i$, the optimal hyperplane can be found by solving the following quadratic problem:

$$\begin{aligned} & \text{Minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ & \text{subject to} \quad y_i(w^T x_i + b) \geq 1 - \xi_i \\ & \quad \quad \quad \xi_i \geq 0, \quad i = 1, \dots, n \end{aligned} \quad (2)$$

The problem (2) can be simplified by converting it with Karush-Kuhn-Tucker (KKT) conditions into the equivalent Lagrange dual problem

$$\begin{aligned} & \text{Maximize} \quad Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ & \text{subject to} \quad \sum_{i=1}^n y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n. \end{aligned} \quad (3)$$

where α_i are nonnegative Lagrange multipliers. The decision function is given by:

$$\begin{aligned} D(x) &= \text{sign}(w^T x + b) \\ &= \text{sign}\left(\sum_{i \in U} \alpha_i y_i x_i^T x + b\right), \end{aligned} \quad (4)$$

where b is given by

$$b = y_i - w^T x_i, \quad (5)$$

and U is the set of support vector indices.

B. Multi-class SVMs

The formulation of SVM is based on a two-class classification problem. How to effectively extend it for multi-class classification is not unique and is an on-going research issue [24]. The most famous types of support vector machines that handle multi-class problems are:

- One-against-all support vector machines [22],
- One-against-one (pairwise) support vector machines [25].

In one-against-all support vector machines, a k -class problem is converted to k two-class problems. For i -th two-class problem, class i is separated from the remaining classes. In pairwise support vector machines, the k -class problem is converted to $k(k-1)/2$ two-class problems which cover all pairs of classes. A problem with both mentioned methods is unclassifiable regions. One way to solve this problem is to use fuzzy membership functions [26]. Another way is Direct Acyclic Graph (DAG) SVM [27] that that uses a decision tree in the testing stage. Training of a DAG is the same as conventional pairwise SVMs. We will use this method in our multi-class problem.

V. THE PROPOSED EMPHATIC SVM

In this section we propose a new structure for support vector machines and then use it for arrhythmia detection. Whereas in the training phase of the SVM (2) a constraint is assigned to each sample, our primary question is that can we investigate the importance degree of samples in the constraint which is ascribed to each sample. To answer this question we use fuzzy inequality in each constraint of the training samples in order to give more flexibility and relaxation to each constraint satisfaction. Note that slack variables ξ_i in conventional SVM cannot play this role because they are the unknowns of the system not the input variables.

The proposed method is obtained by modifying the conventional SVM (2) into the following formulation:

$$\begin{aligned} \text{Minimize } Q(w, b, \xi) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to } y_i(w^T x_i + b) &\geq 1 - \xi_i, \quad i = 1, \dots, n \\ w &\in \mathcal{R}^m, \xi = (\xi_1, \xi_2, \dots, \xi_n), \xi_i \geq 0, i = 1, \dots, n \end{aligned} \quad (6)$$

The symbol \geq means that we like to permit some violations in the satisfaction of the constraints. The fuzzy greater than or equal symbol defines membership functions

$$\mu_i: \mathcal{R}^{m+1+n} \rightarrow (0,1], i = 1, \dots, n.$$

According to the use of the representation theorem of fuzzy sets, consider a linear membership function for the i -th constraint (Figure 5),

$$\mu_i(w, b, \xi) = \begin{cases} 1, & \text{if } y_i(w^T x_i + b) \geq 1 - \xi_i \\ \frac{y_i(w^T x_i + b) - 1 + \xi_i + d_i}{d_i}, & \text{if } 1 - (\xi_i + d_i) \leq y_i(w^T x_i + b) \leq 1 - \xi_i \\ 0, & \text{if } y_i(w^T x_i + b) \leq 1 - (\xi_i + d_i) \end{cases} \quad (7)$$

Note that μ_i is function of an m -dimensional vector w , a scalar b , and an n -dimensional vector ξ .

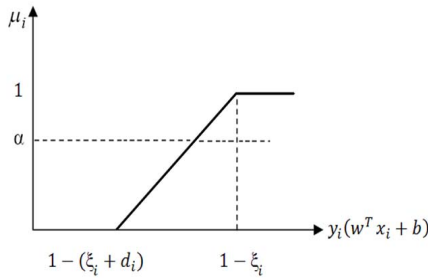


Figure 5: membership function μ_i

For each constraint $i, i=1,2,\dots,n$, of (6),

$$X_i = \{(w, b, \xi) \in \mathcal{R}^{m+1+n} | y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, n\}, \quad (8)$$

where $\xi = (\xi_1, \xi_2, \dots, \xi_n)$.

Taking $X = \bigcap_{i \in I} X_i$, where $I = \{1, \dots, n\}$, then (6) can be written as

$$\text{Minimize } \left\{ Q(w, b, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \mid (w, b, \xi) \in X \right\}. \quad (9)$$

It is clear that $\forall \alpha \in (0,1]$, an α -cut of the constraint set will be the classical set

$$X(\alpha) = \{(w, b, \xi) \in \mathcal{R}^{m+1+n} \mid \mu_X(w, b, \xi) \geq \alpha\},$$

where $\mu_X(x) = \inf \{\mu_i(x), i \in I\}$. In this way $X_i(\alpha)$ will denote an α -cut of the i -th constraint.

The optimal solution of (7) for a given $\alpha \in (0,1]$ is:

$$\begin{aligned} S(\alpha) &= \{(w, b, \xi) \in \mathcal{R}^{m+1+n} \mid \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ &= \{\text{Min } \frac{1}{2} \|w'\|^2 + C \sum_{i=1}^n \xi'_i, (w', b', \xi') \in X(\alpha)\} \end{aligned} \quad (10)$$

As $\forall \alpha \in (0,1]$,

$$\begin{aligned} X(\alpha) &= \\ \bigcap_{i \in I} \{(w, b, \xi) \in \mathcal{R}^{m+1+n} \mid y_i(w^T x_i + b) &\geq r_i(\alpha), \xi_i \geq 0, i = 1, \dots, n\} \end{aligned} \quad (11)$$

with $r_i(\alpha) = 1 - \xi_i - d_i(1 - \alpha)$, thus we have the following problem:

$$\begin{aligned} \text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to } y_i(w^T x_i + b) &\geq 1 - \xi_i - d_i(1 - \alpha), \quad i = 1, \dots, n \\ \xi_i &\geq 0, \quad i = 1, \dots, n \end{aligned} \quad (12)$$

Similar to the conventional SVM, we first convert this constrained problem into the equivalent unconstrained one. Introducing the nonnegative Lagrange multipliers β_i and γ_i , we obtain:

$$\begin{aligned} Q(w, b, \xi, \beta, \gamma) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ &- \sum_{i=1}^n \beta_i \{y_i(w^T x_i + b) - 1 + \xi_i + d_i(1 - \alpha)\} \\ &- \sum_{i=1}^n \gamma_i \xi_i \end{aligned} \quad (13)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T$ and $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^T$.

For the optimal solution, the following Karush-Kuhn-Tucker (KKT) conditions are satisfied:

$$\frac{\partial Q(w, b, \xi, \beta, \gamma)}{\partial w} = 0, \quad \text{i.e., } w = \sum_{i=1}^n \beta_i y_i x_i, \quad (14)$$

$$\frac{\partial Q(w, b, \xi, \beta, \gamma)}{\partial b} = 0, \quad \text{i.e., } \sum_{i=1}^n \beta_i y_i = 0, \quad (15)$$

$$\frac{\partial Q(w, b, \xi, \beta, \gamma)}{\partial \xi} = 0, \quad \text{i.e., } \beta_i + \gamma_i = C. \quad (16)$$

$$\beta_i \{y_i(w^T x_i + b) - 1 + \xi_i + d_i(1 - \alpha)\} = 0 \quad (17)$$

$$\gamma_i \xi_i = 0 \quad (18)$$

$$\xi_i \geq 0, \quad \beta_i \geq 0, \quad \gamma_i \geq 0 \quad (19)$$

where $i=1, \dots, n$.

Thus substituting (14), (15), and (16) into (13), we obtain the following dual problem. Maximize

$$Q(\beta) = \sum_{i=1}^n \beta_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_i \beta_j y_i y_j x_i^T x_j - \sum_{i=1}^n \beta_i d_i (1 - \alpha)$$

$$= \sum_{i=1}^n \beta_i (1 - d_i + d_i \alpha) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_i \beta_j y_i y_j x_i^T x_j \quad (20)$$

subject to the constraints:

$$\sum_{i=1}^n \beta_i y_i = 0, \quad 0 \leq \beta_i \leq C, \quad \text{for } i = 1, \dots, n.$$

The decision function is given by:

$$D(x) = \text{sign}(w^T x + b) = \text{sign}\left(\sum_{i \in U} \beta_i y_i x_i^T x + b\right) \quad (21)$$

and b is given by:

$$b = y_i - \sum_{i \in U} \beta_i y_i K(x_i, x_j) \quad (22)$$

where U is the set of support vector indices.

In fact, we have changed constraints formulation of the SVM problem for our purposes and name this scheme Emphatic constraint Support Vector Machine (ESVM). Constraints of ESVM have more relaxation than traditional SVMs because of their fuzzy inequalities. In this system, d_i and α are meaningful parameters. Each constraint is given a specific d_i which acts as a tolerance to the corresponding sample. In fact, the feasible region is extended for finding the unknown variables (w, b, ξ_i) . Note that, slack variables ξ_i are not user defined and are computed during the training phase. Therefore, we cannot control noisy or outlier samples directly or give importance degree to specific samples using ξ_i . If the same d_i is assigned to all constraints, the system can equally tolerate crossing over any sample. On the other hand, if different d_i s are assigned to different constraints, it means we have assumed a different degree of importance to samples; similar to Fuzzy SVM [28]. Larger d_i causes the corresponding sample x_i to be less important and to be able to consider this data as noise or outlier. It then plays a less important role in determining the separating hyperplane. For ESVM we need a subsystem to determine d_i . We used Circle Method [29] which is a geometric based model for giving importance degree to each sample.

Also, α is another user defined parameter in RSVM formulation. It is the level at which the membership degree of the fuzzy inequality of constraints, μ_i , is cut. A larger value for α means our certainty in the whole set of data is higher and vice versa. Note that, if we have high certainty in the training samples, we should not permit constraint violations. It is clear that $(1 - \alpha)$ indicates the uncertainty of user in the accuracy of collected samples. This new SVM formulation as nonlinear optimization problem with fuzzy inequality constraints adds useful concepts to conventional SVMs.

We will use this new structure of SVM, namely ESVM, for classifying the ECG signal.

VI. THE PROPOSED GENETIC-SVM

In this section, as mention above, we describe the proposed Genetic-SVM system for the feature selection. The aim of this system is to select the subset of features automatically for optimizing the SVM classifier. You can see the overall procedure of algorithm in the following flowchart.

A. Genetic set up:

The first step in GAs is to define the encoding allowing describing any potential solution as a numerical vector, we use a vector of (0 and 1) with length of 22 (the number of features) which 0 and 1 is for the omitted and selected features respectively. At first, randomly we generate 50 chromosomes as a population. We use Roulette Wheel Selection for the cross- over and also we apply *Swap mutation*. This operator simply changes the position of two samples at random. The probability parameter of mutation is equal 0.1.

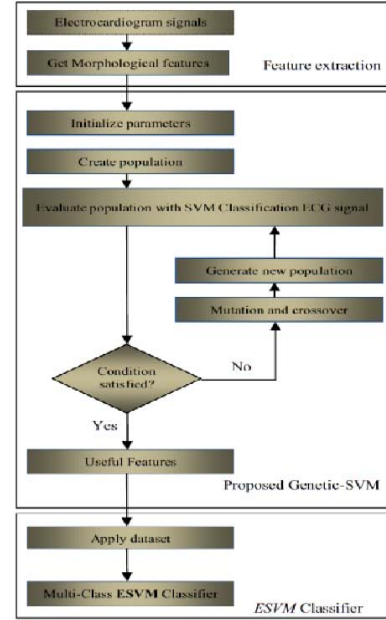


Figure 6: The overall of our proposed arrhythmia detection algorithm

The choice of the fitness function is important because it is on this basis that the Genetic evaluates the goodness of each candidate solution for designing our SVM classification system. In this paper, we shall explore the correction rate of ECG signal classification.

B. Classification of ECG with genetic and ESVM

The procedure describing the proposed SVM classification system is as follows:

- Step 1) generates randomly an initial population of size 50.
- Step 2) for each chromosomes of the population, train $\frac{n(n-1)}{2}$ SVM Classifier.
- Step 3) using OAO (multi-class SVM) for computing fitness of each chromosome (subset of features).
- Step 4) select individuals from population directly based on fitness values and regenerate new individuals from old ones.
- Step 5) If the maximum number of iteration is not yet reached, return to step 2.
- Steps 6) select the best fitness as optimal subset feature.
- Steps 7) apply the optimal feature to dataset.
- Step 8) Classify the ECG signal using the proposed Emphatic SVM (ESVM).

VII. EXPERIMENTAL RESULT

For the evaluating of proposed method we use 50% of all data for training and the rest for test. At first we classify four type of ECG signal without any feature selection. In fact we applied the SVM classifier directly on the entire originally feature space. You can find the result in the Table 3 and Table 4.

Table 3: The arrhythmia classification results with SVM (1), Genetic-SVM (2), and Genetic-ESVM (3) with polynomial kernel

	<i>P,LR</i>	<i>P,LL</i>	<i>P,N</i>	<i>LR,LL</i>	<i>LR,N</i>	<i>LL,N</i>	<i>OVERALL</i>
1.	98.18	100	50	99.38	54.55	67.33	79.23
2.	95.45	88.67	84	100	72.64	96.67	82.31
3.	95.45	89.33	91	99.38	79.09	94	84.62

Table 4: The arrhythmia classification results with Genetic-SVM (1) and Genetic-ESVM (2) with linear kernel

	<i>P,LR</i>	<i>P,LL</i>	<i>P,N</i>	<i>LR,LL</i>	<i>LR,N</i>	<i>LL,N</i>	<i>OVERALL</i>
1.	95.45	99.33	99	100	89	98	93.46
2.	99.09	99.33	97	100	87.27	98	94.23

In the next stage we run our proposed Genetic-SVM and find the best subset of features with SVM fitness function then we detect arrhythmia with use of ESVM. The Genetic-ESVM with linear kernel has a better result in general and shown that our proposed method been very powerful for arrhythmia classification. You can compare Genetic-ESVM, Genetic-SVM and SVM approaches in Figure 7 and Figure 8.

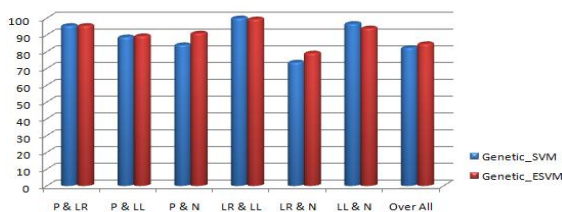


Figure 7: Genetic-SVM vs. Genetic-ESVM approach with polynomial kernel

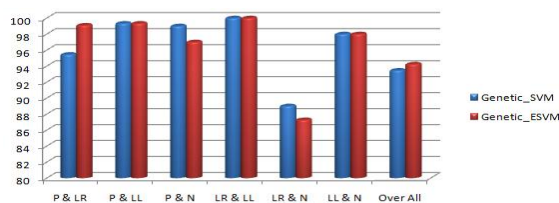


Figure 8: Genetic-SVM vs. Genetic-ESVM approach with linear kernel

REFERENCES

- [1] M. Elgendy, M. Jonkman, F. D. Boer, "Premature Atrial Complexes Detection Using the Fisher Linear Discriminant", 7th IEEE Int. Conf. on Cognitive Informatics (ICCI'08), 2008.
- [2] A. Gharaviri, F. Dehghan, M. Teshnelab, H. Abrishami, "comparison of neural network anfis, and SVM classification for PVC arrhythmia detection", Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.
- [3] F. Melgani, Y. Bazi, "Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 5, SEPTEMBER, 2008.

- [4] T. H. Linh, S. Osowski, and M. L. Stodolowski, "On-line heart beat recognition using Hermite polynomials and neuron-fuzzy network," IEEE Trans. Instrum. Meas., vol. 52, no. 4, pp. 1224-1231, Aug. 2003.
- [5] A. Azemi, V.R. Sabzevari, M. Khademi, H. Gholizadeh, A. Kiani, Z. Dastgheib, "Intelligent Arrhythmia Detection and Classification Using ICA", Proceedings of the 28th IEEE, EMBS Annual International Conference, New York City, USA, Aug 30-Sept 3, 2006.
- [6] T. Inan, L. Giovangrandi, and J. T. A. Kovacs, "Robust neural network based classification of premature ventricular contractions using wavelet transform and timing interval features," IEEE Trans. Biomed. Eng., vol. 53, no. 12, pp. 2507-2515, Dec. 2006.
- [7] Vapnik, V., "The Nature of Statistical Learning Theory", New York: Springer-Verlag, 1995.
- [8] Joachims, T., "Text categorization with support vector machines: Learning with many relevant features", Technical report, University of Dortmund, 1997.
- [9] Wang, T.-Y., Chiang, H.-M., "Fuzzy support vector machine for multi-class text categorization", Information Process and Management, 43, 914-929, 2007.
- [10] Salomon, J., "Support vector machines for phoneme classification", M.Sc Thesis, University of Edinburgh, 2001.
- [11] Pontil, M., Verri, A., "Support Vector Machines for 3D Object Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 6, 1998.
- [12] Takeuchi, K., Collier, N., "Bio-Medical Entity Extraction using Support Vector Machines", Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine, 57-64, 2003.
- [13] Foody, M.G., Mathur, A., "A Relative Evaluation of Multiclass Image Classification by Support Vector Machines", IEEE Transactions on Geoscience and Remote Sensing, 42, 1335- 1343, 2004.
- [14] Platt, J., Cristianini, N., Shawe-Taylor, J., "Large margin DAGs for multiclass classification", Advances in Neural Information Processing Systems 12. MIT Press, 543-557, 2000.
- [15] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machine," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778-1790, Aug. 2004.
- [16] C.-W.Hsu and C.-J. Lin, "A comparison of methods formulticlass support vector machines," IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 415-425, Mar. 2002.
- [17] Hsu, C.-W., Lin, C.-J., "A comparison of methods for multiclass support vector machines", IEEE Trans. Neural Networks 13(2), 415-425, 2002.
- [18] Gen, M. and Yun.Y, soft computing approach for reliability optimization: state of the art survey, In Proceedings of the Reliability engineering and system safety, 2006, 1008-1026.
- [19] Eiben, A.E. and Smith, J.E. Introduction to Evolutionary Computing, Springer, 2nd printing, 2007.
- [20] R. Mark and G. Moody MIT-BIH Arrhythmia Database 1997 [Online]. Available <http://ecg.mit.edu/dbinfo.html>.
- [21] V. Vapnik, The Nature of Statistical Learning Theory, New York: Springer-Verlag, 1995.
- [22] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, 1998.
- [23] S. Poyhonen, A. Arkkio, P. Jover, and H. hyotyniemi, Coupling pairwise support vector machines for fault classification, Control Engineering Practice, Vol. 13, 759-769, 2005.
- [24] C.W. Hsu and C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks, Vol. 13 (2), 415-425, 2002.
- [25] U. H.-G., Krebel, Pairwise Classification and Support Vector Machines, In B.Schölkopf, C. J. C. Burges, and A. J. Smola, Eds., Advances in Kernel Methods: Support Vector Learning, 225-268, MIT Press, Cambridge, MA, 1999.
- [26] S. Abe, Support Vector Machines for Pattern Classification, Spinger-Verlag London Limited, 2005.
- [27] J. Platt, N. Cristianini, J. Shawe-Taylor, Large margin DAGs for multiclass classification, Advances in Neural Information Processing Systems 12. MIT Press, 543-557, 2000.
- [28] C. F. Lin and S. D. Wang, Fuzzy Support Vector Machine, IEEE Trans. on Neural Networks, vol. 13, no. 2, 464-471, Mar. 2002.
- [29] L. Chu, and C. Wu, "A Fuzzy Support Vector Machine Based on Geometric Model," Proceedings of the fifth World Congress on Intelligent Control and Automation, Hangzhou, P.R. China, pp.1843-1846, June 15-19, 2004.