2009 Third UKSim European Symposium on Computer Modeling and Simulation

ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm

Jalal A. Nasiri [α*], Mahmoud Naghibzadeh [α], H. Sadoghi Yazdi [α], Bahram Naghibzadeh [β]

*Commutation and Computer Research Center, Ferdowsi University of Mashhad, Mashhad,Iran **
Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran [α]
School of Medicine, Mashhad University of Medical Science, Mashhad, Iran [β]
j.nasiri@wali.um.ac.ir, naghibzadeh@um.ac.ir, h_sadoghi@um.ac.ir, naghibzadeh@mums.ac.ir

Abstract— **This research is on presenting a new approach for cardiac arrhythmia disease classification. The proposed method combines both Support Vector Machine (SVM) and Genetic Algorithm approaches. First, twenty two features from electrocardiogram signal are extracted. These features are obtained semiautomatically from time-voltage of R, S, T, P, Q features of an Electro Cardiagram signals. Genetic algorithm is used to improve the generalization performance of the SVM classifier. In order to do this, the design of the SVM classifier is optimized by searching for the best value of the parameters that tune its discriminate function, and looking for the best subset of features that optimizes the classification fitness function. Experimental results demonstrate that the approach adopted better classifies ECG signals. Four types of arrhythmias were distinguished with 93% accuracy.**

Keywords: **ECG, arrhythmia, support vector machine, genetic algorithms, feature reduction.**

I. INTRODUCTION

Classification of electrocardiograms (ECG) into different disease categories is a complex pattern recognition task. However, the analysis of electrocardiogram signals is the most effective available method for diagnosing cardiac arrhythmias. Computer-based classification of ECGs can provide high accuracy and offer a potential of an affordable cardiac abnormalities mass screening. 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 used for training and test of automated classification of ECG signals include many different features. Some of them are based on laboratory experiments, while others involve 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 present in [1] is based on *Fisher Linear dicriminant*. The RR interval duration and the distance between the occurrence of P and T waves is perceived. Using these features Fisher's Linear *Discriminant* is applied. In [2] a SVM-based method for PVC arrhythmia detection is shown to be more efficient than Anfis. In [3] a new approach for feature selection and classification of cardiac arrhythmias based on PSO-SVM is proposed. 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 have 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 the normal and PVC beats from the others.

Although there are many other methods for the classification of ECG signals, the focus of this paper is on the algorithms that are most similar to the proposed method.

The paper is structured as follows. Section II describes Support Vector Machines Classification. Sections III describes the proposed method for feature extraction. Sections IV discuss feature reduction concept. The proposed genetic-SVM classification method is discussed in V. Finally, section VI and VII presents the concluding remarks and future works.

II. SUPPORT VECTOR MACHINE CLASSIFICATION

In this section, a brief description of the two and multi-class SVM classification concept is reviewed.

Support Vector Machines (SVMs) [7] are very popular and powerful in pattern learning because of supporting high

978-0-7695-3886-0/09 $26.00 © 2009 IEEE
DOI 10.1109/EMS.2009.39

dimensional data and, at the same time, providing good generalization properties. Moreover, SVMs have many usages in pattern recognition and data mining applications such as text categorization [8, and 9], phoneme recognition [10], 3D object detection [11], image classification [12], bioinformatics [13] etc. At the beginning, SVM was formulated for two-class (binary) classification problems. The extension of this method to multi-class problems is neither straightforward nor unique. DAG SVM [14] is one of the methods that have been proposed to extend SVM classifier to support multi-class classification.

A. Binary Support vector machine formulation

Let $X = \{(x_i, y_i)\}_{i=1}^{n}$ be a set of n training samples, where $x_i \in \mathcal{R}^m$ is an m-dimensional sample in the input space, and $y_i \in \{-1,1\}$ is the class label of sample x_i. SVM finds the optimal separating hyperplane (OSH) with the minimal classification errors. The linear separation hyperplane is in the form of

$$f(x) = w^T x + b$$

where w and b are the weight vector and bias, respectively. The optimal hyperplane can be obtained by solving the optimization problem (1), where ξ_i is slack variable for obtaining a soft margin while variable C controls the effect of the slack variables. Separation margin increases by decreasing the value of C.

In a support vector machine, the optimal hyperplane is obtained by maximizing the generalization ability of the SVM. Howevver, if the training data are not linearly separable, the obtained classifier may not have high generalization ability, eventhough the hyperplanes are determined optimally. To enhance linear separability, the original input space is mapped into a high-dimensional dot-product space called the feature space. Now using the nonlinear vector function $\varphi(x) = (\varphi_1(x), \dots, \varphi_l(x))^T$ that maps the m-dimensional input vector x into the l-dimensional feature space, the OSH in the feature space is given by

$$f(x) = w^T \varphi(x) + b,$$

The decision function for a test data is:

$$D(x) = sign(w^T \varphi(x) + b).$$

The optimal hyperplane can be found by solving the following quadratic optimization problem:

$$Minimize \ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$subject\ to \ \ y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, \ \ i = 1, \dots, n$$

B. Multi Class Support vector machine

As described before, SVMs are intrinsically binary classifiers, but, the classification of ECG signals often involves more than two classes. In order to face this issue, a number of multiclass classification strategies can be adopted [15], [16]. The most popular ones are the one-against-all (OAA) and the one-against-one (OAO) strategies.

The one against one constructs $\frac{n(n-1)}{2}$ decision functions for all the combinations of class pairs. Experimental results in [17] indicate that the one-against-one is more suitable for practical use. (More details appeared in [17]). We use OAO for ECG multi class classification.

III. FEATURE EXTRACTION AND SELECTION

In this section we will explain the characteristics of the extracted feature from the ECG signals and the procedure designed for the extraction. Figure 1, presents the block diagram of the proposed arrhythmia classification. Details of what is done at each stage are explained later.

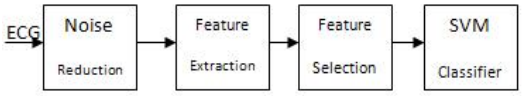

Fig 1: Block diagram of proposed arrhythmia classification

A. Dataset Description

Our experiments were conducted on the ECG data as the basic signal for classification. In recent researches, the annotated ECG records, available at the MIT-BIH arrhythmia database [20], have been widely used for the evaluation of the performance of different classifiers. The database has 48 records with each record being an ECG signal for the duration of 30 minutes. Each data was recorded in two channels, modified limb lead II and modified lead VI.

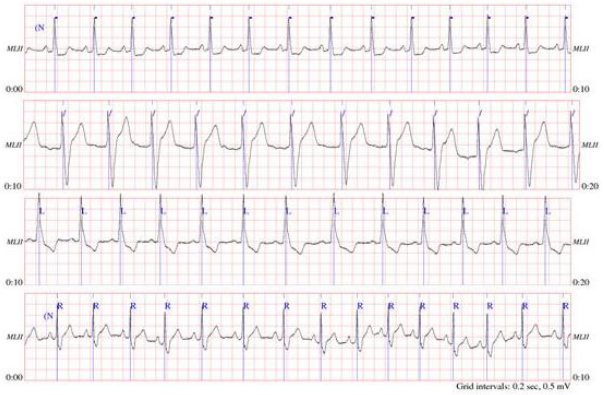

Fig 2: Sample signal of Normal, Paced, LBBB, RBBB

In particular, the considered beat types refer to following classes: normal sinus rhythm (N), right bundle branch block (RB), left bundle branch block (LB), and paced beat (P). In Figure 2, sample of four N, RB, LB, and P beats are noticeable. The beats were selected from the recording of

following patients, 100, 106, 107, 109, 111, 118, 202, 209, 212, 214, 215 and 217.

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	*Normal(N)*
2.	107,217	110	*Paced(P)*
3.	111,214	600	*LBBB(LB)*
4.	118,212	450	*RBBB(RB)*

B. Noise reduction

In the first stage of feature extraction, a wavelet transformation is performed in order to reduce noises. The wavelet transform allows processing non-stationary signals such as ECG signal. This is possible Figure 3a is the presents the original signal and Figure 3b is the same signal after reducing noises.

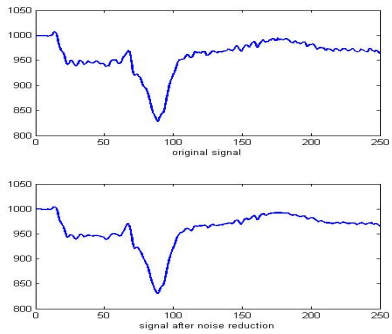

Fig 3a,3b: Sample signal, previous and after noise reduction

C. Feature Description

For each signal nineteen temporal features such as R-R interval, PQ interval, PR interval, and PT interval and three morphological features are recognized. These features are manually extracted for each beat and put into a separate vector. Each vector is tagged with one the four possible labels N, P, LB, RB.

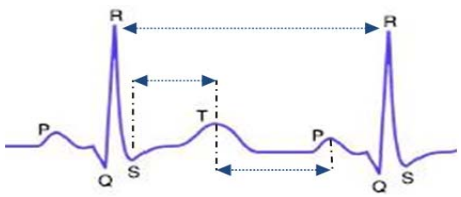

Fig 4: sample features, ST interval, TP interval and RR interval

Features have been extracted including the time and voltage of Q/R/S/T/P and time interval for each of 5 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: Features Descriptions 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 by computing the maximum and the minimum values of a beat in ECG signal. Signals of each beat are scaled, using the following formula, such that the range of every signal is between zero and one.

$$f(t) = \frac{f(t) - min\,(t)}{max(t) - min\,(t)}$$

The minimum and maximum voltages between the first and the second R feature is computer first and the normalization action is performed [0 1]. As mentioned before, we considered percent that are higher than 0.2, 0.5 and 0.8 as three features. Six of the 22 features, called basic features, are: R1, S, T, P, Q, R2 and the rest are called derived features. The derived features are calculated using the basic features via a semiautomatic procedure. 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).

IV. FEATURE REDUCTION

Many studies, in the field of data analysis and feature selection, suggest that not all the features are useful for classification [2, 3, and 5]. On the contrary, some features may act as noises and, hence, reduce the classification accuracy. In this research, two different feature reduction approaches are adopted. The studied showed that a meta-heuristic-based approach has better performance for classification of ECG arrhythmia than statistical method [3].

A. Principal component analysis

Principal Component Analysis (PCA) is a statistical method for reducing the dimensions of the data [21]. It

selects a set of variables that are uncorrelated with each other and, at the same time, each one is linear combination of the original variables. Principal components are derived from the original data such that the first principal component accounts for the maximum proportion of the variance of the original data set, and subsequent orthogonal components account for the maximum proportion of the remaining variance. The process steps of PCA are as follows:

Step 1) Compute the mean vector of data.
Step 2) Compute the covariance matrix of data.
Step 3) Compute the eigenvalue and eigenvector matrix of covariance matrix.
Step 4) Form the components using the eigenvectors of the covariance matrix as weighting coefficients.

It remains to be shown that the PCA classifier performs well for all datasets, rather, it may not perform well for some datasets [3]. Some researches [22, 23 and 24] show that the PCA isn't powerful at analyzing nonlinear structure data. It seems, the existence of noisy data, abnormal features ranges such that the range of some features is [0 1] while for others is [0 1000], and low variance of some important features are reasons for the weak performance of PCA. You can see the experimental result by PCA in table 4.

B. Genetic Algorithm

Genetic Algorithm is one prospective option for feature reduction. Other meta-heuristic optimization techniques such as simulated annealing tabu search, and evolutionary strategies are also candidates for this purpose. GA has been demonstrated to converge to a semi-optimal solution for many diverse and difficult problems as a powerful and stochastic tool based on principles of natural evolution [18]. In many application it is used for feature reduction and feature weighting [25] . The details of our implementation of GA are described as follows:

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

The first step in any GA algorithm is to define the encoding to allow describing a potential solution as a numerical vector, and then try to generate a population randomly. We briefly describe some concepts and operations in GA.
Selection operator: The selection process directly selects individuals from the current population based on the fitness values of every chromosome [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 a fitness function is to measure the quality of a solution.

V. THE PROPOSED GENETIC –SVM

In this section, the proposed Genetic-SVM system for the classification of ECG signals is described. The aim of this system is to automatically select a proper subset of features for optimizing the SVM classifier. The upcoming figure shows the procedure followed.

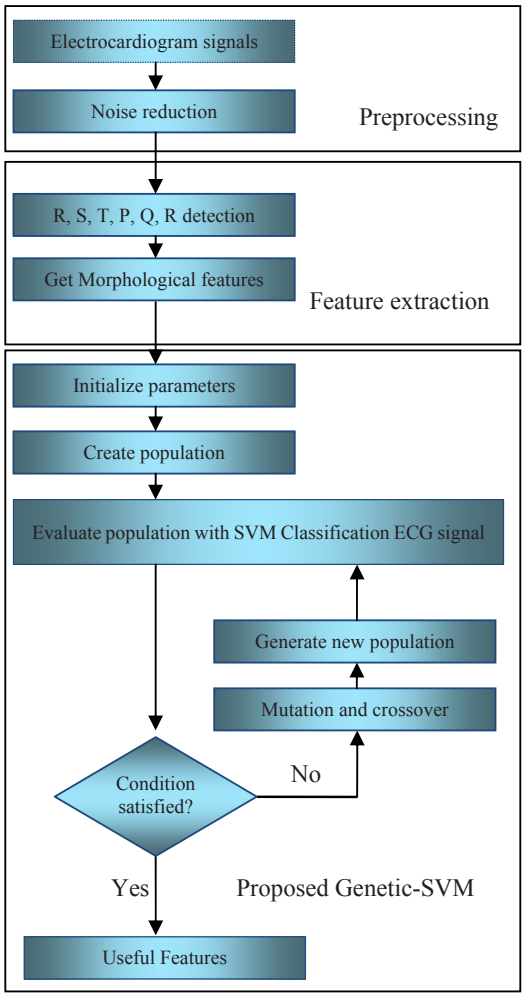

Fig 5: the Proposed Genetic –SVM approach

A. Genetic set up:

The first step in a GA algorithm is to define the encoding procedure which allows describing any potential solution as a numerical vector. A vector of length 22 with each component being either zero or one is taken for the encoding. A zero-value for the component means the corresponding feature is omitted and a one-value means vice versa. In the following experiments, the original population consisted of 50 randomly selected chromosomes. Roulette Wheel Selection and *Swapping* are used for cross-over and mutation operations, respectively. Swap operation randomly changes the position of two samples. The probability parameter of mutation is taken to be 0.1.

The choice of the fitness function is important because it is on this basis that the Genetic Algorithm 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. SVM Classification with geneticalgorithm

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

Step 1) Randomly generates an initial population of size 50.
Step 2) For each chromosomes of the population, train $\frac{n(n-1)}{2}$ SVM Classifiers.
Step 3) Using OAO (multi-class SVM), compute the fitness of each chromosome (subset of features).
Step 4) Directly Select some individuals from the current population based on the 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 chromosome with the best fitness value as the desired subset of features.
Step 7) Classify the ECG Signals with the trained SVM.

VI. EXPERIMENTAL RESULT

For the evaluation of the proposed method, 50% of all the data of the MIT-BIH dataset are used for training the composed system and the rest are used for the evaluation. In the first experiment, the selection stage was omitted and the SVM classifier was directly applied to the entire original feature space. The results of this experiment are presented in the Table 3, 4 and 5.

Table 3: The Arrhythmia classification result using ***SVM*** with Linear (1.) and Polynomial (2.) kernel.

	P,LR	*P,LL*	*P,N*	*LR,LL*	*LR,N*	*LL,N*	*OVERALL*
1.	98.18	100	50	99.38	54.55	67.33	79.23
2.	93.64	98	54	99.38	53.46	67.33	78.08

Table 4: The Arrhythmia classification result using ***PCA_SVM*** with Linear (1.) and Polynomial (2.) kernel.

	P, LR	*P, LL*	*P,N*	*LR,LL*	*LR, N*	*LL, N*	*OVERALL*
1.	98.2	98.7	53.7	99.4	54.6	66.7	80.00
2.	99.2	99.4	54.6	97.5	54.6	66.7	79.26

Table 5: The Arrhythmia classification result using ***Genetic_SVM*** with Linear (1.) and Polynomial (2.) kernel.

	P,LR	*P,LL*	*P, N*	*LR, LL*	*LR,N*	*LL,N*	*OVERALL*
1.	95.45	99.33	99	100	89	98	**93.46**
2.	95.45	88.67	84	100	73.64	96.67	82.31

In the next experiment, the proposed genetic-SVM classifier is used to the best subset of features that can optimize the SVM classifier. It was concluded that the proposed genetic-SVM with a linear kernel performs generally better. performance in general and shown that our proposed method been very powerful for arrhythmia classification. Figure 7 compares the genetic-SVM approach with plain SVM classifier.

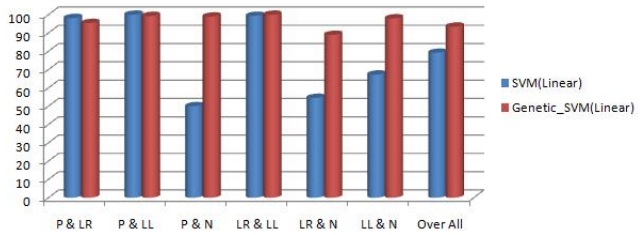

Fig 6: SVM vs. Genetic –SVM approach with linear kernel

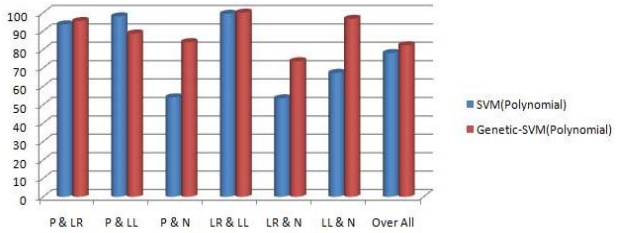

Fig 7: SVM vs. Genetic –SVM approach with polynomial kernel

VII. CONCLUSION AND FUTURE WORK

In this paper, a new method based on support vector machines for feature selection of electrocardiogram signals is proposed. Experimental results show that feature selection greatly improves the quality of classification. This is because some features may act as noises and negatively affect the outcome. In the absence of efficient deterministic algorithm, meta-heuristic approaches are considered for

feature reduction. A flexible and effective GA-SVM method for classification of ECG signal was presented and showed to be effective on MIT-BIH dataset to recognize arrhythmia classes.

For future works, the weight of each selected feature on the classification of ECG signals will be determined.

ACKNOWLEDGEMENT

This work has been partially supported by Iran Telecommunication Research Center (ITRC), Tehran, Iran. This support is gratefully acknowledged.

REFERENCES

[1] M. Elgendi, 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," comparision 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, andM. L. Stodoloski, "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 Processinge 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] R. O. Duda, P. E. Hart and D. G. Stork, Pattern classification (2nd Edition), Wiley-Interscience, 2000
[22] B. Schölkopf, A. Smola, and K.Müller, "Kernel principal component analysis", In Int. Conf. on Artificial Neural Networks, 1997, pp. 583–588.
[23] T. Takiguchi and Y. Ariki, "Robust Feature Extraction using Kernel PCA", In Int. Conf. on Acoustics, Speech and Signal Processing, 2006, pp. 509-512.
[24] A. Lima, H. Zen, Y. Nankaku, C. Miyajima, K. Tokuda and T. Kitamura, "On the Use of Kernel PCA for Feature Extraction in Speech Recognition", In eurospeech 2003, 2003, pp. 2625-2628.
[25] H.Moeinzadeh, M-M.Mohammadi, A.Akbari, B.Nasersharif, " Evolutionary-Class Independent LDA As a Pre-Process for Improving Classification ", 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009), Montreal, Canada, 2009.