ECG Arrhythmia Classification using Least Squares Twin Support Vector Machines

Mohammad S. Refahi
Department of Electrical Engineering
Amirkabir University of Technology
msaleh.refahi@aut.ac.ir

Jalal A. Nasiri
Iranian Research Institute for Information Science and Technology(IRANDOC)
j.nasiri@irandoc.ac.ir

S.M. Ahadi
Department of Electrical Engineering
Amirkabir University of Technology
sma@aut.ac.ir

***Abstract*—Heart disease is one of the most common causes of death. Rapid diagnosis of patients with these diseases can greatly prevent them from sudden death. Today, the diagnosis of heart diseases is done by cardiologist, while achieving an automatic and accurate method for diagnosing has become a challenging issue in this area. Because small changes in the electrocardiogram signals are not recognizable with eyes, and visual disorders may be affected, artificial intelligence and machine learning algorithms can be the solution. In this paper, we use the Least Squares Twin-Support Vector Machine, which unlike ordinary support vector machine, is based on a Non-parallel margin. The results show that the method of this article is better than previous methods, and more accurate and faster for diagnosing arrhythmia.**

***Index Terms*—ECG arrhythmia, Least Squares Twin Support Vector Machines, Directed acyclic graph**

I. Introduction

The diagnosis of heart disease by electrocardiogram is one of the fastest and cheapest methods. Changes and distortions in each of the ECG[1] signal parameters can indicate a heart condition. Each of these abnormal changes is called an arrhythmia.

Because of human mistakes in diagnosing cardiac arrhythmias, achieving an automated system with high precision in the diagnosis of heart disease is very important. A typical heart rate consists of three waveforms QRS, P, and T. The magnitude and duration between each of these three waves can be considered as a feature for feature extraction.

However, one of the most popular and useful databases in this area is MIT-BIH. Researchers have used this database to test their various algorithms for arrhythmia detection and classification. In this paper, the problem of detecting arrhythmias is presented with a different approach. LS-TSVM[2] is used which is faster and more accurate than SVM[3] .In LS-TSVM instead of obtaining an optimal hyperplane, two distinct hyperplanes are obtained [1].

Several methods have been proposed for classifying of ECG signals. Among them, the most recently published works are presented. The method presented in [2] used the Pan-Tompkins algorithm to accurately extract features such as QRS complex and P wave, and employed a decision tree to classify each beat in terms of these features. A new approach to selecting the feature and classification of cardiac arrhythmias is proposed based on PSO[4]-SVM. At the first using experimental methods to show SVM method was superior, and then by using PSO introduces the generalized SVM algorithm [3].

In [4], neuro-fuzzy approach for the ECG-based classification of heart rhythms is described. Here, the QRS complex signal was characterized by Hermite polynomials, whose coefficients feed the neuro-fuzzy classifier. Detection of arrhythmia by means of ICA[5]and wavelet transform to extract important features was proposed in [5].

In [6] genetic algorithm in combination with SVM for ECG signal classification. Experimental results showed that, LS-TSVM significantly improved the recognition efficiency and classification accuracy.

The rest of this paper is organized as follows. In Section II we explain LS-TSVM. Section III covers an overview of directed acyclic graph ,and Multi-class LS-TSVM are briefly reviewed. In Sections IV. proposed our method and we explain feature extraction and introduce MIT-BIH and its important features . Experimental results are given in Section V, and Section VI contain concluding remarks.

II. LEAST SQUARES TWIN-SUPPORT VECTOR MACHINE

The main idea used in LS-TSVM, is the attempt to find two non-parallel hyperplanes[1]. Least Squares Twin-Support Vector Machines were originally designed for binary data classification and later developed for multi-class classification by using DAG[6] , OVO[7]and OVA[8]algorithms.

In order to obtain two non-parallel hyperplanes, the following optimization problems have to be solved.

$$\min_{w(1)b(1)} \frac{1}{2}(Aw^{(1)}+eb^{(1)})^T(Aw^{(1)}+eb^{(1)})+\frac{C1}{2}y^Ty \quad (1)$$

Subject to. $(Bw^{(1)}+eb^{(1)})+y=e$

$$\min_{w(2)b(2)} \frac{1}{2}(Aw^{(2)}+eb^{(2)})^T(Aw^{(2)}+eb^{(2)})+\frac{C1}{2}y^Ty \quad (2)$$

[1]Electrocardiogram
[2]Least Squares Twin-Support Vector Machine
[3]Support Vector Machine
[4]Particle Swarm Optimization
[5]Independent Component Analysis
[6]Directed Acyclic Graph
[7]One Versus One
[8]One Versus All

978-1-5386-4916-9/18/$31.00 ©2018

Subject to. $(Bw^{(2)} + eb^{(2)}) + y = e$

Vector w(i) is the coordinate of the hyperplanes and also b(i) is a bias. The matrix A denotes data of class 1 and matrix B represents class -1. A column vector of ones in real space of arbitrary dimension will be denoted by e.The variable y is the slip rate parameter. In (1) and (2), C1 and C2 are penalty parameters, which need to be adjusted for maximum accuracy . In this paper the idea of PSVM is used [7], which is an extremely fast and simple algorithm that requires only solution of a system of linear equations for generating both linear and nonlinear classifiers.

To solve equation (1) and (2), and minimize w(i)and b(i), firstly insert subject function in the objective function and make Derivative. Problems (1), (2) are solved by system of linear equations; the results are given in Formulas 3 and 4 [1].

$$[\begin{matrix} w^{(1)} \\ b^{(1)} \end{matrix}] = -(F^T F + \frac{1}{C1} E^T E)^{-1} F^T e \quad (3)$$

$$[\begin{matrix} w^{(2)} \\ b^{(2)} \end{matrix}] = (E^T E + \frac{1}{C2} F^T F)^{-1} E^T e \quad (4)$$

The matrix E is equal to [A e] and the matrix F is equal to [B e].

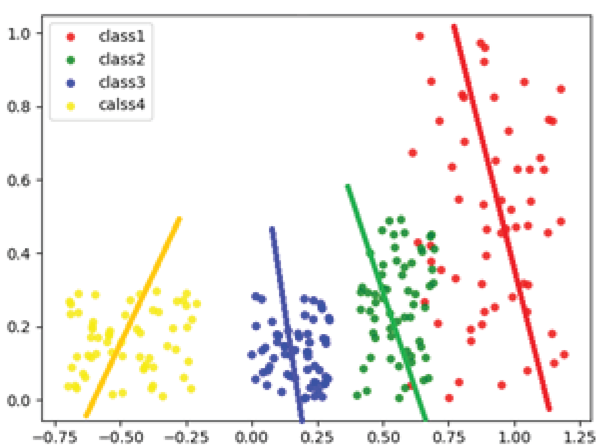

Fig. 1. Geometric representationl Multi class LSTSVM classifier

Once the weights and biases of the two non-parallels separating hyperplanes:

$$x^{'} w^{(1)} + b^{(1)} = 0 \ , \ x^{'} w^{(2)} + b^{(2)} = 0 \quad (5)$$

are obtained from (3) and (4), a new data point $x \in R^n$ is assigned to a class +1 or -1 depending on to which of the two hyperplanes, its perpendicular distance is minimum: $x^{'} w^{(1)} + b^{(1)} = 0$ or $x^{'} w^{(2)} + b^{(2)}$. It can be noted that LS-TSVM does not change the meaning of support vectors defined in TSVM, however we will not be able to identify them as we are solving primal problems instead of dual problems. For clarity, we explicitly state our linear LS-TSVM algorithm [1]. Figure 1 is provided to better illustrate multi-class LS-TSVM.

Algorithm 1 Linear LS-TSVM

Require: X train data, X' test data
  Input: $x \in \mathcal{R}^n | x \in A$, $x \in \mathcal{R}^n | x \in B$
  Output: $x \in A or B$
1. **Initialization** E =[A e] and F =[B e] .
2. Select penalty parameters C1 and C2. Usually these parameters are selected based on validation.
3. Determine the parameters of two non-parallel hyperplanes using (3) and (4).
4. Calculate perpendicular distances $|x^{'} w^{(1)} + b^{(1)}|$ and $|x^{'} w^{(2)} + b^{(2)}|$ for a new datapoint $x \in R^n$.
5. Assign the data point to class +1 or -1 based on which of the $|x^{'} w^{(1)} + b^{(1)}|$ or $|x^{'} w^{(2)} + b^{(2)}|$ is minimum.

III. DIRECTED ACYCLIC GRAPH

Most common methods for classifying multiple classes are OVO, OVA, and a DAG. Since in DAG the problem of unclassifiable regions is solved, in this paper we use DAG. DAG is very similar to OVO , but in this method, $(K-1)$ LS-TSVM algorithm is executed, while in OVO $(K.\frac{K-1}{2})$ algorithm executed. K is the number of dataset classes. In the experimental step, a DAG is used, where it has $(K.\frac{K-1}{2})$ nodes and K edges. In fact, each node of this graph is a binary classifiers. By starting from the root of the graph in each node, the classifier is evaluated, and according to the output value, it moves to the left or right of the graph. Fig. 2 shows the structure of DAG .

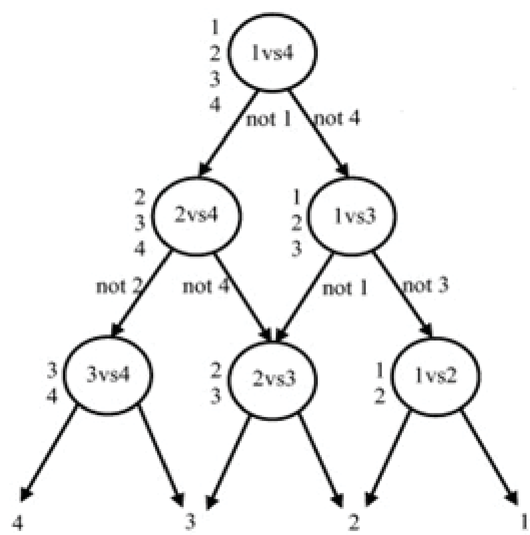

Fig. 2. Directed Acyclic Graph LS-TSVM classifier

IV. FEATURE EXTRACTION

Figure 3,First, by improving the SNR ratio, the ECG signal is prepared for signal processing. Then the main points of each electrocardiogram signal are determined. The resulting dataset contains 22 attributes and 520 samples.Then the LS-TSVM classifier is trained and then used DAG was used to classifying data.

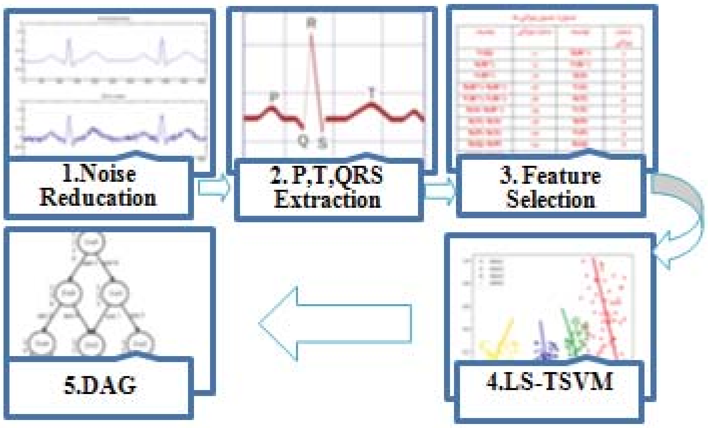

Fig. 3. Block diagram of proposed arrhythmia classification

A. *MIT−BIH Data base*

Includes 48 double-channel ECG recordings, which sampled from, 47 people who have been studied at the Boston Hospital Laboratory. Recordings are stored at 360 samples per second, with a precision of 11 bits at 10 mV. People who have been studied , they were either healthy, or had one of the three issues of right bundle branch block, left bundle branch block, and paced rhythm. Fig. 4 shows four different electrocardiograms are related to normal, right bundle branch block,left bundle branch block, and paced rhythm.

B. *Feature description*

For each signal, 22 important features, which extracted from the waveform by a cardiologist. The most important of these features are the distance between each of the points R1, R2, T, S and Q from each other ,as well as the heart rate, etc. which are shown in Table I. For example, X(R1) means the duration of R1 and V(R1) means the voltage of R1.Choosing these features was based on how the electrocardiograms are read by cardiologist and what they considered diagnosis diseases[8].

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

All of the obtained features are based on six features which 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).[8]

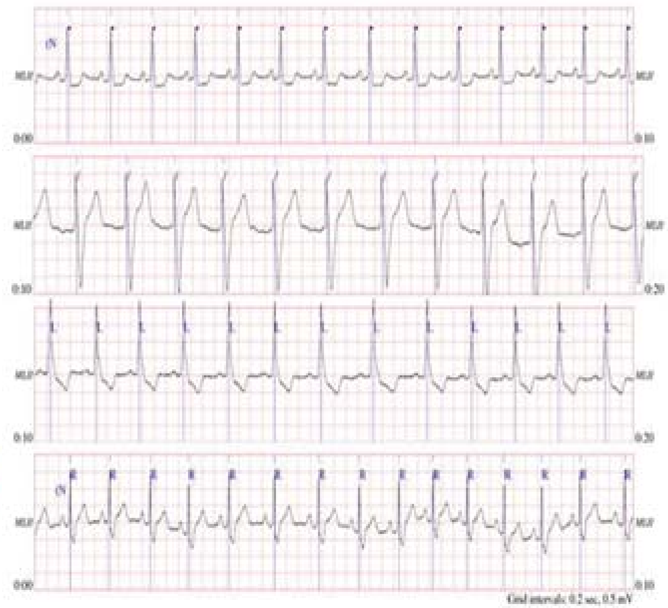

Fig. 4. Sample signal of Normal, Paced, LBBB, and RBB

TABLE I
FEATURES DESCRIPTIONS

Description	Feature number	Description	Feature number
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)		

V. EXPERIMENTAL RESULTS

In this study, all the experiments were implemented in Python 3.6 on Windows7 with an Intel Core i3 CPU (1.90GHz) with 4-GB RAM. To obtain signal information and extract outward properties,MATLAB was used to specify the QRS, P,T points, and then extracts features. The classification performance of the proposed multiclass classifier is affected by the choice of hyperplanes parameters. DAG LS-TSVM classifiers include two penalty parameters c1 and c2 .

For linear LS-TSVM our parameter ranges were C1, C2$\in$ $\{2^{(-15)}, ., 2^{(20)}\}$ and in nonlinear LS-TSVM used a third degree polynomial kernel.

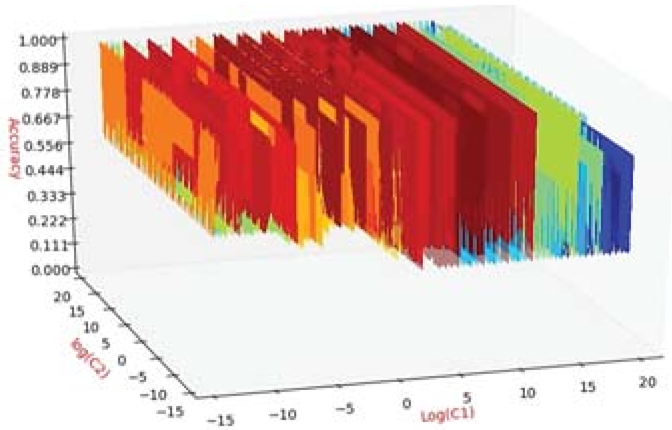

Fig. 5. The influence of the penalty parameter (C1 = C2) in accuracy

Then, using grid search, we found the best parameters. Fig. 5 shows the influence of parameters on the predictive accuracy of the proposed classifiers for MIT-BIH dataset. From Fig. 5, it is clear that the influence of C1 on the accuracy is more as compared to C2. For low values of penalty parameter in range $(2^{(-5)}, ., 2^{(10)})$ DAG LS-TSVM classifiers show better results.

For evaluation of the proposed method, 80% of all data in the MIT-BIH dataset were used for training the composed system and the rest were used for the evaluation.Then in order to validate data ,we used tenfold cross validation. Also, for better understanding of wrong classification results,a confusion matrix of the output has been shown, in TABLE II.

TABLE II
CONFUSION MATRIX

class	1	2	3	4
1	98.18	0.90	0	0.90
2	0.90	93.63	1.8	3.63
3	0	0	1	0
4	0	6.06	0	93.93

We have tested two multi-classifiers, which were obtained by extending the formulation of binary LS-TSVM on the basis of OVA and DAG methods. OVA LS-TSVM classifier separates one class from the remaining classes. However, unclassifiable regions exist in this approach. This classifier also leads to the class imbalance problem.[11]

DAG LS-TSVM classifier,however, solves the unclassifiable regions problem. The training phase of DAG LS-TSVM is similar to OVO LS-TSVM classifier but it takes less time to predict the class for a given data point.In testing phase, DAG LS-TSVM classifier is faster among all three methods. It also has better generalization ability and obtains highest or comparable accuracy with MIT-BIH dataset. DAG LS-TSVM classifier is four times faster than traditional DAG SVM which also shows its superiority[1].

Statistical analysis of the performance of each classifier also confirms that the DAG LS-TSVM classifier is the best performing classifier. As shown in TABLE III, in this application, linear LS-TSVM has better result than non-linear LS-TSVM.

TABLE III
THE ARRHYTHMIA CLASSIFICATION RESULTS

Methods	**Optimal parameters**	**Best accuracy**	**Running Time(s)**
DAG Linear LS-TSVM	C1= -7 C2= -3	**97.1**	**30**
DAG Nonlinear Polynomial LS-TSVM	C1= -10 C2=0 Degree=3	70	850
SVM Polynomial	C=8 Degree=2	91.1	900
OVA Linear LS-TSVM	C1=-2 C2=-2	95.5	40

In this study,we have investigated and compared several approaches such as: neural networks, SVM, Decision tree, and some improved SVMs in the accuracy. These results indicate that our proposed method significantly improved the recognition efficiency and classification accuracy and also had a good generalization ability as compared to the previous studies.

TABLE IV
COMPARISON OF OUR RESULTS WITH PREVIOUSLY STUDIES

Methods	**Year**	**Accuracy**	**Author**
PSO-SVM	2008	89.7	Melgani[3]
Decision tree	2014	90.38	Park[2]
Feature selection	2011	94.86	Llamedo[11]
Genetic-SVM	2010	96	Khazaee[6]
E-SVM	2009	94.23	Nasiri[8]
PSO-SVM	2014	96.06	Khazaee[10]
neural network	2006	96.82	Inan[9]
Proposed Method	**2018**	**97.1**	Refahi

VI. Conclusion and Future Work

As LS-TSVM has a better generalization ability and faster computational speed, we extended the binary LS-TSVM to DAG LS-TSVM and proposed a fast ECG heartbeat recognition method .

In this paper, all the data features were used and there was no feature reduction. However,as a future work, with feature selection, higher accuracy can be reached based on algorithms such as PSO and PCA . For more accurate classification results also, Fourier transform of the signal can be used for the purpose of obtaining more feature for more accurate classification results.

References

[1] M. Arun Kumar and M. Gopal,*"Least squares twin support vector machines for pattern classification"*, Expert Systems with Applications, vol. 36, no. 4, pp. 7535-7543, 2009.

[2] Park and K. Kang,*"PcHD: Personalized classification of heartbeat types using a decision tree"*, Computers in Biology and Medicine, vol. 54, pp. 79-88, 2014.

[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"*, IEEETrans. Instrum. Meas., vol. 52, no. 4, pp. 12241231, 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] A. Khazaee, A. Ebrahimzadeh, *" Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features"*. Biomed.Signal Process. Control.5(4), 252263 (2010)

[7] G. Fung and O. Mangasarian, *"Multicategory Proximal Support Vector Machine Classifiers"*, Machine Learning, vol. 59, no. 1-2, pp. 77-97, 2005.

[8] Jalal A. Nasiri Mahmoud Naghibzadeh, H. Sadoghi Yazdi ,Bahram Naghibzadeh , *"ECG Arrhythmia Classification with Support Vector Machines and GeneticAlgorithm "*, Third UKSim European Symposium on Computer Modeling and Simulation 2009.

[9] O.T. Inan, L. Giovangrandi, G.T. Kovacs, Robust,*" neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features"*, IEEE Trans. Biomed.Eng. 53(12), 25072515 2006.

[10] A. Khazaee, A.E. Zadeh, *" ECG beat classification using particle swarm optimization and support vector machine"*, Front. Comput. Sci. 8(2), 217231 ,2014.

[11] M. Llamedo, J.P. Martinez, *" Heartbeat classification using feature selection driven by database generalization criteria. "*, Trans. Biomed. Eng. 58(3), 616625 ,2011.

[12] Tomar, Divya, and Sonali Agarwal. *"A Comparison on Multi-Class Classification Methods Based on Least Squares Twin Support Vector Machine."*, Knowledge-Based Systems, vol. 81, 2015, pp. 131147., doi:10.1016/j.knosys.2015.02.009.