SVM based Methods for Arrhythmia Classification in ECG

Narendra Kohli[1], Nishchal K. Verma[1], and Abhishek Roy[2]

[1]Department of Electrical Engineering, Indian Institute of Technology Kanpur, 208016 Kanpur, India

[2]Department of Electrical and Electronics Engineering, National Institute of Technology Karnataka, Surathkal, 575025 Mangalore, India

E-mail IDs: nkohli@iitk.ac.in, nishchal@iitk.ac.in, abhishekroyn@gmail.com

Abstract--In this study, Support Vector Machine (SVM) based methods have been used to classify the electrocardiogram (ECG) arrhythmias. Among various existing SVM methods, three well-known and widely used algorithms one-against-one, one-against-all, and fuzzy decision function are used here to distinguish between the presence and absence of cardiac arrhythmia and classifying them into one of the arrhythmia groups. The various types of arrhythmias in the Cardiac Arrhythmias ECG database chosen from University of California at Irvine (UCI) to train SVM, include ischemic changes (coronary artery disease), old inferior myocardial infarction, sinus bradycardy, right bundle branch block, and others. The results obtained through implementation of all three methods are thus compared as per their accuracy rate in percentages and the performance of the SVM classifier using one-against-all (OAA) method was found to be better than other techniques. ECG arrhythmia data sets are of generally complex nature and SVM based one-against-all method is found to be of vital importance for classification based diagnosing diseases pertaining to abnormal heart beats.

KeyWords--Arrhythmias,Classification,Electrocardiogram, Support vector machine.

I. INTRODUCTION

Among various illness human being has suffered from, heart diseases are still one of the important problems known to mankind. Early detection and proper medical treatment of diseases pertaining to heart can save lives of patients in cases of sudden death [1]. Heart beat is stimulated by electric signal which basically originates from the Sino Atrial node (SA) located at the top of the right chamber or Atrium (RA) in the heart. This node is better known as heart's "natural pacemaker". Any disruption of this pacemaker may cause the heart to beat at an abnormal rate, thereby adversely impacting the circulation of blood throughout the body [2]. Arrhythmia is a kind of heart conduction system disease that causes an inefficient heart beat [3]. It represents abrupt and abnormal ECG beats. In general, arrhythmia is diagnosed by an electrocardiogram procedure. ECG signals are comprised of P wave, QRS complex, and T wave. They are designated by capital letters P, Q, R, S, and T. A typical normal ECG signal is shown in figure 2 in [2]. The main parameters included for inspection

in heart-patients are the shape, the duration, and the relationship with each other of P wave, QRS complex, and T wave components and also R-R interval. The changes in these parameters signify an illness of the heart that may happen by any reason. The entire irregular beat phases are commonly called arrhythmia and some arrhythmias are very dangerous for a patient [1]. An ECG facilitates two major kinds of information; firstly, if the time intervals on the ECG are measured, it helps in determining the duration of the electrical wave crossing the heart and consequently we can determine whether the electrical activity is normal or slow, fast or irregular. Secondly, if the amount of electrical activity passing through the heart muscle is measured, it enables a pediatric cardiologist to find out if parts of the heart are too large or are overworked [2]. Thus, physicians diagnose arrhythmia based on long-term ECG data using an ECG recording system. Physicians interpret the morphology of the ECG waveform and decide whether the heartbeat belongs to the normal sinus rhythm or to the class of arrhythmia [4]. With the various remote and mobile healthcare systems adapting ECG recorders, are being increased in number these days, the importance of a better and robust automatic arrhythmia classification algorithm is being increasingly acknowledged.

The analysis of ECG is basically recognizing its' pattern and classifying arrhythmia in real-time. To date, several researchers have made attempts to use SVM and various other classifiers to classify electrocardiograph beats. Many algorithms have been proposed over previous years for developing the automated systems to accurately classify the electrocardiographic signals. Devine and Macfarlane [5] used ANN for the detection of left ventricular strain by classifying ST-T abnormalities of the ECG. In the same year, Hu et al. [6] investigated the use of ANN for QRS detection and beat classification. They modeled the non-linear background noise using adaptive multilayer perceptron structure in order to augment the QRS complexities for its more reliable detection. Silipo et al. [7] presented a comparison work for ECG classification using two classification techniques; one with supervised; and other with unsupervised learning. Sugiura et al. developed a

486

fuzzy logic based method for detecting cardiac arrhythmias and discriminating ventricular arrhythmias [8]. Acharya et al. used heart rate variability as the base signal and implemented ANN and fuzzy equivalence relation for the classification of four cardiac arrhythmias [9]. Song et al. [4] proposed SVM based arrhythmia classification with the reduction of feature dimensions by linear discriminant analysis (LDA). Since, a SVM is known to have the advantage of offering remarkable performance of classification [4]; we have chosen three well-known and most widely used SVM based methods in this study for classification of standard arrhythmia dataset and thereby comparing their accuracy rates obtained for best results.

This paper is organized into V Sections. A brief discussion on the support vector machine is presented in Section II. In section III we review several multi-class SVM methods, such as one-against-one, fuzzy decision function, and one-against-all. Section IV presents the results of all the classification methods including the results of comparison for arrhythmia classification in ECG on the basis of percent accuracy rate obtained. Finally, the conclusions are drawn in Section V.

II. SUPPORT VECTOR MACHINE

SVMs were designed for binary classifications and its algorithm can be better understood with a mathematical explanation and example as discussed in [10]. Let $S = \{(x_1, y_1), (x_2, y_2), ..., (x_i, y_i)\}$ be a training set where x_i are m-dimensional attribute vectors, $y_i \in \{-1, +1\}$, $y_i = -1$, and $y_i = +1$ for class 1 and class 2, respectively.

According to [11], the SVMs classifier is defined as follows:

$$D(x) = w^T \Phi(x) + b = 0 \qquad (1)$$

where $\Phi(x)$ is a mapping function, w^T is a vector in the feature space, and b is a scalar.

To classify the data linearly separable in the feature space, the decision function satisfies the following condition:

$$y_i (w^T \Phi(x) + b) \geq 1$$

$$\text{for} \quad i = 1, ..., l \qquad (2)$$

Among all the separating hyperplanes, the optimal separating hyperplane with maximal margin between two classes can be formed as follows:

$$\min_{w,b} J(w,b) = \frac{1}{2} w^T w \qquad (3)$$

subject to (2). If the training data are nonlinearly separable, the hard margin constraints are relaxed by introducing slack variables ξ_i in (2) as follows:

$$y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$$

$$\text{for} \quad i = 1, ..., l \qquad (4)$$

$$\xi_i \geq 0$$

$$\text{for} \quad i = 1, ..., l \qquad (5)$$

In order to obtain the optimal separating hyperplane, we should minimize

$$\min_{w,b,\xi_i} J(w, b, \xi_i) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{l} \xi_i \qquad (6)$$

As per (4) and (5), parameter γ determines the tradeoff between the maximum margin and the minimum classification error. The optimization problem of (6) is a convex quadratic program and can be solved using Lagrange multiplier method. By using Lagrange multipliers α_i and β_i ($i = 1, 2.....l$), the Lagrangian function can be constructed as follows:

$$L(w, b, \alpha_i, \xi_i, \beta_i,) = J(w, b, \xi_i)$$
$$- \sum_{i=1}^{l} \alpha_i \{y_i [w^T \phi(x_i) + b] - 1 + \xi_i\} - \sum_{i=1}^{l} \beta_i \xi_i \qquad (7)$$

According to the Kuhn–Tucker theorem, the solution of the optimization problem using Largarngian function is as follows:

$$w = \sum_{i=1}^{l} \alpha_i y_i \phi(x_i) \qquad (8)$$

The training examples (x_i, y_i) with nonzero Lagrangian coefficients α_i are called support vectors. By solving the following convex quadratic programming problem, we can find the α_i coefficients.

$$\max [-\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j (\phi(x_i)^T \cdot \phi(x_j)) \alpha_i \alpha_j + \sum_{i=1}^{l} \alpha_i] \qquad (9)$$

subject to

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \qquad i = 1, ..., l \qquad (10)$$

$$0 \leq \alpha_i \leq \gamma, \qquad i = 1, ..., l \qquad (11)$$

By substituting (8) into (1), the classifier can be obtained. Given new input $x, f(x)$, can be estimated by using (12). If $f(x) > 0$, the sample belongs to class 1; otherwise class 2

$$f(x) = \text{sgn}\{\sum_{i=1}^{l} \alpha_i y_i \cdot (\phi(x_i)^T \cdot \phi(x)) + b\} \qquad (12)$$

Where

$$\text{sgn}(x) = \begin{cases} 1, x > 0 \\ 0, x \leq 0 \end{cases}$$

In (12), the pair wise inner product in the feature space can be computed from the original data items using a kernel function [12], [13], and the kernel function can be denoted by

$$K(x, x_i) = \phi(x)^T \cdot \phi(x_i) \qquad (13)$$

487

The typical kernel functions include polynomial kernel functions and radial basis functions (RBFs). In this way, $f(x)$ can be rewritten as follows:

$$f(x) = \mathrm{sgn}\{\sum_{i=1}^{l} \alpha_i y_i \cdot K(x_i, x) + b\} \qquad (14)$$

III. MULTICLASS CLASSIFICATION ALGORITHM

The multiclass classification problem refers to assigning each of the observations into one of k classes. In this section, we have introduced the one-against-one, FDF and one against all methods. The figures in this section can be found in [14]. At first, we introduce the one-against-one method.

Assume $S = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\}$ is a training set, where $x_i \in R^m$ and $y_i \in (1, 2, ... k)$. For the one against-one method [15], one needs to determine $k(k-1)/2$ classifiers for the k - classes problems. The optimal hyperplane with SVMs for class i against class j is

$$D_{ij}(x) = w_{ij}^T \phi(x) + b_{ij} = 0,$$

$$i < j, 1 < j \leq k, 1 \leq i < k$$

where w^T_{ij} is a vector in the feature space, $\Phi(x)$ is a mapping function, and b_{ij} is a scalar.

Here the orientation of the optimal hyperplane is defined as per the following equation:

$$D_{ij}(x) = -D_{ji}(x) \qquad (15)$$

A. One-against-One method

For the input vector, one computes

$$D_i(x) = \sum_{j \neq i, j=1}^{k} \mathrm{sgn}(D_{ij}(x)) \qquad (16)$$

and classifies x into the class

$$\arg\max_{i=1,...,k}(D_i(x))$$

$$(17)$$

B. Fuzzy Decision Function Method

In the FDF method [16], for the input vector x , the 1-D membership function $m_{ij}(x)(i, j = 1, 2, ..., k)$ in the directions orthogonal to the optimal separating hyperplanes $D_{ij}(x) = 0$ is defined as follows:

$$m_{ij}(x) = \begin{cases} 1, & 1 \leq D_{ij}(x) \\ D_{ij}(x), & otherwise \end{cases}$$

In [16], the membership functions $m_i(x)$ are given by

$$m_i(x) = \min_{j=1,...,k} m_{ij}(x)) \qquad (18)$$

Using (18), sample x is classified into the class

$$\arg\max_{i=1,...k}(m_i(x)) \qquad (19)$$

C. One-against-all method

For a k class problem, the one-against-all method constructs k SVM models. The ith SVM is trained with all of the training examples in the ith class with positive labels and all other examples with negative labels. The final output of the one-against-all method is the class that corresponds to the SVM with the highest output value [17]. Thus, by solving the optimization problem in (3)-(5) using all the training samples in the dataset, the decision function of the ith SVM is

$$D_i(x) = w_i^T \phi(x) + b_i$$

The input vector **x** will be assigned to the class that corresponds to the largest value of the decision functions. The class of **x** = $\arg\max_{i=1,...,k}(D_i(x))$.

IV. RESULTS

The standard multivariate ECG dataset taken here is chosen from the University of California at Irvine (UCI) Cardiac Arrhythmias database [18]. The experiments were conducted on a personal computer with 1.5 GHz CPU and 1 GB of RAM. Initially this database contained 452 instances and 279 attributes, of which 206 attributes are linear valued and the rest are nominal. But owing to the presence of many missing values and also zero valued columns, it became imperative to preprocess and resize the dataset while maintaining the reliability and relevance of the dataset. Thus, columns containing all zeroes or all missing values were removed first from the dataset, followed by the removal of rows having missing values in the dataset and removal of classes having insignificant number of instances. This resized the dataset to total 377 instances and 166 attributes, distributed into 6 classes with class 1 referring to 'normal' ECG, class 2 to 5 referring to different classes of arrhythmia and class 6 referring to the rest of unclassified ones. Table I shows different arrhythmia classes with number of instances belonging to each of the classes in the dataset.

Table I. Arrhythmia classes with corresponding number of instances in the dataset

Class	Class name	No. of instances
1	Normal	237
2	Ischemic changes (Coronary Artery Disease)	36
3	Old Inferior Myocardial Infarction	14
4	Sinus bradycardy	24
5	Right bundle branch block	48
6	Others	18

488

10	2^5	69.15	62.77	72.87
11	2^6	69.15	62.77	72.87
12	2^7	69.15	62.77	72.87
13	2^8	69.15	62.77	72.87

All the SVM based methods taken here were trained by half of the total dataset chosen fairly from the main dataset ensuring representation of all classes present in the required percentage. The remaining half of the main dataset was used for testing and analysis purpose. Table II shows the representation of each of the classes in the training and the testing datasets.

Table II. Representation of each of the classes in the training and the testing datasets

Data	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Total
Initial	237	36	14	24	48	18	377
Train	119	18	7	12	24	9	189
Test	118	18	7	12	24	9	188

We applied the SVM based methods using Gaussian kernel. The kernel parameter σ and the regularization parameter lambda λ were empirically optimized in Table III and Table IV respectively, by minimizing the error rate on the validation dataset in order to obtain the best accuracy rate in terms of percentage and thus ensuring better classification of the arrhythmia dataset.

Table III The accuracy rate (in %) wrt σ value

No.	σ	OAO	FDF	OAA
1	2^{-4}	62.77	62.77	62.77
2	2^{-3}	62.77	62.77	62.77
3	2^{-2}	62.77	62.77	62.77
4	2^{-1}	62.77	62.77	62.77
5	2^0	62.77	62.77	62.77
6	2^1	62.77	62.77	62.77
7	2^2	62.77	62.77	62.77
8	2^3	62.77	62.77	69.15
9	2^4	**69.15**	62.77	**72.87**
10	2^5	64.89	62.77	68.61
11	2^6	62.77	62.77	63.30
12	2^7	62.77	62.77	62.77
13	2^8	62.77	62.77	62.77

Table IV The accuracy rate (in %) wrt λ value at best σ value

No.	λ	OAO	FDF	OAA
1	2^{-4}	3.72	62.77	68.09
2	2^{-3}	3.72	62.77	68.09
3	2^{-2}	3.72	62.77	68.09
4	2^{-1}	54.26	62.77	**73.40**
5	2^0	**69.68**	62.77	72.34
6	2^1	69.15	62.77	72.87
7	2^2	69.15	62.77	72.87
8	2^3	69.15	62.77	72.87
9	2^4	69.15	62.77	72.87

A. Comparison of performance

The classification results of arrhythmia in ECG database indicate that among various SVM based methods, one-against-all algorithm shows the highest percentage of accuracy rate. The system was trained and optimized by classifying the dataset at various σ values. That σ value was chosen and fixed, which gives highest percentage of accuracy rate; and next, the system was converged by classifying it again at various λ values. The range of values taken for σ and λ values are, $\sigma = [2^{-4}, 2^{-3}, 2^{-2}, ..., 2^7, 2^8]$ and $\lambda = [2^{-4}, 2^{-3}, 2^{-2}, ..., 2^7, 2^8]$. It is found that performance wise one-against-one algorithm trails behind one-against-all method in ECG classification but its results could be competitive at times. FDF showed poor and constant classification results on ECG dataset. To remove ambiguity over having any potential to improve its performance, FDF method was used for classification with σ value ranging from 1 to 5000 but no improvement in accurate rate was observed. Clearly, OAO method resulted in the highest accuracy rate but in general, the very high accuracy rate is difficult to obtain. This could be due to the presence of a particular class sweeping maximum share of number of instances in the total dataset. This reflects the further potential of OAO method to give even higher results in cases of ECG datasets with more uniform distribution, thereby ensuring better training of the system.

V. CONCLUSION AND FUTURE WORK

This paper presents classification of arrhythmia in ECG database with various popular and widely used SVM based methods. We have applied some well known classification techniques, viz., one-against-one, fuzzy decision function, and one-against-all, to distinguish between the presence and absence of cardiac arrhythmia and classify it in one of the six groups. SVM is a strong classifier and using various methods it can deliver high performance even with smaller learning data [4]. The obtained results clearly indicate OAA to be the most preferred method for arrhythmia classification in ECG. Seeing the potential to give dominating results over other methods, we are working towards improvisation of OAA method and in future, expecting to implement on various other standard datasets as well.

489

REFERENCES

[1] Y. Ozbay and B. Karlik, "A Recognition of ECG Arrththmia Using Artificial Neural Networks," Proceedings – 23rd Annual conference, IEEE/EMBS, Oct. 25-28, 2001, Istanbul, Turkey.

[2] B. Anuradha and V.C. Veera Reddy, "Cardiac Arrhythmia Classification Using Fuzzy Classifier," Journal of Theoretical and Applied Information Technology.

[3] J. Kim, H.S. Shin, K. shin, and M. Lee, "Robust Algorithm for Arrhythmia Classification in ECG Using Extreme Learning Machine," BioMedical Engineering OnLine, 2009, 28th October 2009, **8**:31.

[4] M.H. Song, J.Lee, S.P. Cho, K.J. Lee, and S.K. Yoo, "Support Vector Machine Based Arrhythmia Classification Using Reduced Features," International Journal of Control, Automation, and Systems, vol. 3, no. 4, pp. 571-579, Dec. 2005.

[5] B. Devine and P. W. Macfarlane, "Detection of Electrocardiographic left ventricular strain using neural nets," Med. Biol. Eng. Comput, 31, pp. 343-348, 1993.

[6] Y. H. Hu, W. J. Tompkins, J. L. Urristi, and X. A. Valtino, "Application of artificial neural networks for ECG signal detection and classification," Journal of Electrocardiology, 26, pp. 66-73, 1993.

[7] R. Silipo, G. Bortolan, and C. Marchesi, "Supervised and unsupervised learning for diagnostic ECG classification," IEEE Eng. in Medicine & Biology, pp.1054, 31 Oct.-3 Nov., Amsterdam, The Netherlands, 1996.

[8] Sugiura, T. Hirata, H. Harada, Y. and Kazui, T., "Automatic discrimination of arrhythmia waveforms using fuzzy logic", Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 20, no. 1, 1998.

[9] Acharya U R, Subbanna Bhat P, Iyengar S S, Rao A and Dua S, "Classification of heart rate using artificial neural network and fuzzy equivalence relation", Pattern Recognit., vol. 36, pp. 61–68, 2003.

[10] B. Liu, Z. Hao, and E.C.C. Tsang, "Nesting One-Against-One Algorithm Based on SVM's for Pattern Classification," IEEE Transactions on Neural Networks, Vol. 19, No. 12, Dec. 2008.

[11] V. N. Vapnik, The Nature of Statistical Learning Theory. London, U.K.: Springer-Verlag, 1995.

[12] M. Aizeman, E. Braverman, and L. Rozonoer, "Theoretical foundations of potential function method in pattern recognition learning," Autom. Remote Control, vol. 25, pp. 821–837, 1964.

[13] S. Saitoh, Theory of Reproducing Kernels and Its Applications. Harlow, U.K.: Longman, 1988.

[14] S. Abe, Support Vector Machines for Pattern Classification. London, U.K.: Springer-Verlag, 2006.

[15] U. H. G. Kreßel, "Pairwise classification and support vector machines," in Advances in Kernel Methods-Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge, MA: MIT Press, 1998, pp. 255–268.

[16] D. Tsujinishi and S. Abe, "Fuzzy least squares support vector machines for multiclass problems," Neural Netw., vol. 16, pp. 785–792, 2003.

[17] R. Debnath, N. Takahide, and H. Takahashi, "A decision based one-against-one method for multi-class support vector machine," Pattern Anal. Appl., vol. 7, pp. 164-175, 2004.

[18] P. M. Murphy and D. W. Aha, UCI Machine Learning Repository Database, 2010 [Online]. Available: [http://archive.ics.uci.edu/ml], Irvine, CA: University of California, School of Information and Computer Science.