Support Vector Machine For Classification of Heartbeat Time Series Data

A. Sasirekha, P. Ganesh Kumar

Abstract—Support vector machine (SVM) is a relatively new machine learning tool and has emerged as a powerful technique for learning from data and in particular, for solving binary classification problems. In the literature several statistical-learning paradigms have been proposed for developing a heart rate variability analysis. SVM classification decision which is based on the feature extraction of Heart rate variability (HRV) analysis. Results on a real-life long-term ECG recordings of young and elderly healthy dataset show that understandable SVMs provide a anticipating tool for the prediction of heart rate signals, where as a feature of heart have been generated. Feature extraction describes a pattern or relationships between input features and output class labels directly from the data. This paper proposes several different techniques for Feature extraction. The accuracy is obtained by using the comparison of HRV features.

Keywords-QRS detection algorithm, heart rate variability (HRV), support vector machine (SVM)

I. INTRODUCTION

HEARTBEAT recorded by the electrocardiogram clearly reflects the physiological control mechanism of the autonomic nervous system on heart rate. This shows both the decreasing and increasing trends and variability of the variance of heartbeat. It is a time series data that depicts the number of heartbeat per unit of time which can vary as the body need for oxygen changes such as exercise or sleep. Heart rate variability analysis is based on measuring the variability of heart rate signals and more specifically variations per unit of time of the number of heart beats [1]. HRV analysis is applied to the estimation of the autonomic nervous balance, stress or relaxation condition and to the evaluation of mental or physiological workload. It is derived from the difference in time intervals elapsed between two consecutive heartbeats called cardio intervals or the interval between an R wave and the next R wave is the inverse of the heart rate (R-R interval) and is measured in milli seconds [2].

HRV is important because it provides a window to observe the heart's ability to respond to normal regulatory impulses that affect its rhythm. A primary focus of clinical work and research is in observing or modifying the balance in regulatory impulses from the vagus nerve and sympathetic nervous system [3]. Some researchers are focusing attention on other factors that regulate the heart, such as chemo receptors, thermo receptors, and the renin-angiotensin system. HRV is affected by several factors such as age and health status. It decreases with age, lower among people who have an inactive lifestyle and among those who have medical conditions such as coronary heart disease, hypertension and diabetic neuropathy.

Manuscript received August, 2013.

A. Sasirekha, Assistant Professor, Department of Computer Science and Engineering, Info Institute Of Engineering, India.

P. Ganesh Kumar, Assistant Professor, Department of Information Technology, Anna University of Technology, Coimbatore, India.

The analysis of HRV in individuals, including methods from nonlinear dynamics and taking the 24-hour heart rate and blood pressure (BP) variations into consideration, altogether could well have the power to become a useful diagnostic tool, particularly in mild and long-term anti-hypertensive treatments [4].

The power spectral analysis is one of the most popular non-invasive methods for monitoring the autonomic nervous system control function. It is very sensitive to even small errors like finite sampling, noise recording or misinterpretation of R-Points, in the process converting surface ECG signal into HRV signal [5]. Entropy methods exploit a symbolic representation of the HRV time series. Despite the serve reduction of information, they are able to enhance relevant features of the signal. The multiscale entropy analysis provides interesting indicators for the study of time series plot, has been proposed for the evaluation of respiratory sinus arrhythmia (RSA) [7] [11].

Statistical techniques are based on the estimation of the statistical properties of the beat-to-beat time series and describe the underlying system’s average statistical behavior over the considered time window [8]. The heartbeat time series classification is investigate with the real datasets[9] i.e. the dataset consists of long term ECG recordings of young and elderly (21-34 yr old) and twenty elderly (68-85yr old) accurately screened by the healthy subjects underwent120 min of continuous ECG and respiration of signals were collected. The R-R interval time series for each subject was then computed. Subjects were told to breathe normally and an attempt was made to maintain the respiratory rate at around 12 min. The original heartbeat time series was corrupted by zero mean white Gaussian noise [10]. The standard deviation of the noise was selected appropriately in order to obtain the signal to noise ratio (SNRs) between 0 and 5db.

Support vector machine (SVMs) [12] [13] are a relatively new machine learning tool and has emerged as a powerful technique for learning from data and in particular, for solving binary classification problems. It is based on the developed theory were proposed. This made statistical learning theory not only a tool for the theoretical analysis but also a tool for creating practical algorithms for estimating multidimensional functions [14]. SVM constructs a hyper plane or set of hyper planes in high dimensional space without any assumptions on the data distribution. Accordingly SVM is called by the name Support Vector Classification (SVC) and Support Vector Regression (SVR) [15] [20]. The classification problem can be restricted to consideration of the two-class problem without loss of generality. In this problem the goal is to separate the two classes by a function which is induced from available examples. The goal is to produce a classifier that will work well on unseen examples. The linear classifier is termed the optimal separating hyper plane. Intuitively, we would expect this boundary to generalize well as opposed to the other possible boundaries. Larger and more complex classification problems have been attacked with SVC [16] [17].

The advantages of classical SVC are: a global minimum solution; relatively fast training speed for large-scale learning tasks; and sparseness in solution representation. Notably, has applied SVC to the exacting problem of face recognition, with encouraging results. In conclusion, SVC provides a robust method for pattern classification by minimizing over fitting problems. They use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. This study investigates the problem further, examines a larger number of HRV computation methods, extracts more features [18], and uses different datasets. It compares Support Vector Machine (SVM) [19] [20] during different neural network-based classifiers and studies their hardiness to noisy data. The execution of the proposed approach is tested and trained utilizing an electrocardiogram recording of heartbeat dataset [21].

II. PROPOSED MODEL FOR HEART RATE VARIABILITY ANALYSIS

The proposed methodology for the features of Heart Rate Variability (HRV) analysis is based on using Support Vector Machine (SVM) for detecting the Normal and Abnormal conditions of the given parameters, which leads to various attacks. The Support Vector Machine classifier approach for this purpose has two phases; training and testing. During the training phase, SVM classifier is trained to capture the underlying relationship between the chosen inputs and outputs. After training, the classifiers are tested with a test data set, which was used for training. Once the SVM classifiers are trained and tested, they are ready for detecting the intrusions at different operating conditions. The following issues are to be addressed while developing an SVM for Heart Rate Variability Analysis [13]:

1. Data Collection
2. Data preprocessing and representation
3. Data Normalization
4. Selection of HRV features
5. Training and Testing classification

Figure 1 shows the schematic representation of the issues to be addressed while developing a SVM model for Heart rate variability analysis.

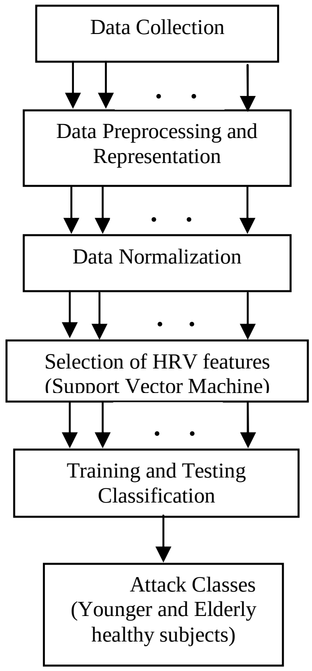

Fig 1 Proposed SVM model for HRV analysis

2.1 Data Collection

There are two ways to build HRV analysis, one is to create our own simulation classifier, and collect relevant data and the other one is by using previously collected datasets. Issues like privacy, security, and completeness greatly restrict people from generating data. The beauty of using previously collected datasets is that the results can be compared with others in the literature. Not many data sets have being collected that could built HRV analysis systems. Some of the popularly used long term electrocardiogram (ECG) recordings of younger and elderly healthy datasets are Fantasia (Disney, 1940) data set which are available in the MIT [9].

2.2 Data Preprocessing

Before training the SVM classifier, the dataset should be preprocessed to remove the redundancy present in the data and the non-numerical attributes should be represented in numerical form suitably.

2.3 Data Normalization

During training of the Support Vector machine classifier, higher valued input variables may tend to suppress the influence of smaller ones. Also, if the heartbeat data is directly applied to the SVM, there is a risk of the simulated noise reaching the saturated conditions. If the noises get saturated, then the changes in the input value will produce a very small change or no change in the output value. To minimize the effects of magnitudes among inputs as well as to prevent saturation of the noise activation function, the input data is normalized before being presented to the SVM classifier. One way to normalize the data x is by using the expression:

$$x_n = \frac{(x - x_{min}) \times range}{(x_{max} - x_{min})} + starting\ value$$

where, x_n is the normalized value and x_{min} and x_{max} are the minimum and maximum values of the data.

2.4 Selection of HRV features

To make a SVM classifier to perform some specific task, one must choose number of trained and testing data set which utilizes the R-R detection Algorithm passes the signal through a low-pass and a high-pass filter in order to reduce the influence of the muscle noise which is computed by the way of statistical HRV features based on the RR signal mean value. These features are extracted by the SVM classifier.

2.5 Training and Testing Classification

Once the appropriate features of the heart rate variability signals are selected, the SVM classifier model is trained to capture the underlying relationship between the input and output using the training and testing data. In this work, trained and tested data, which constructs the support vector (data point at the margin).These classification produce the accuracy result of the SVM model.

III. REVIEW OF SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) [22] is a learning machine that plots the training vectors in high dimensional feature space, and labels each vector by its class. SVM classifies data by determining a set of support vectors, which are members of the set of training inputs that outline a hyper plane in feature space [23]. The SVM is based on the idea of structural risk minimization, which minimizes the generalization error, i.e. true error on unseen examples. The number of free parameters used in the SVM depends on the margin that separates the data points to classes but not on the number of input features. Thus SVM does not require a reduction in the number of features in order to avoid over fitting. More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high or infinite dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. In this two class classification problem we assume we are given a training data set, $X\kappa \in \mathbb{R}^n$, $d_k \in \{-1,+1\}$, and an indicator function f which is the bipolar signum function which will ultimately permit a mapping from each of the input points X_κ to the appropriate d_k. Class C_1 points will be indicated by a+1 which we refer to as positive samples: class C_2 points will be indicated by a-1 and thus called negative samples [24]. Note that linear separability implies that we can find an oriented hyper plane defined by a set of weights W and a bias ω_0 (assuming that the weight vector is not augmented) which separates the positive data points from the negative ones. For all points on the hyper plane we know that $W\cdot X+\omega_0=0$, which is the defining equation of the hyper plane. The hypothesis space under consideration is the set of functions:

$$f(X,W,\omega_0) = sign(W\cdot X+\omega_0)$$

IV. SIMULATION RESULT

This section presents the details of the simulation study carried out on Fantasia (Disney1940) Dataset [15] using the proposed method. This data set was collected by simulating a physiological signal achieves for biomedical research, operated like a real environment and being blasted with multiple attacks. Each long term ECG records of younger and elder healthy subjects contains 10 input features which is given in table 1 and one output that is labeled as either normal or as an attack. They are younger and older or elder features contains the twenty young (21 - 34 years old) and twenty elderly (68 - 85 years old) rigorously-screened healthy subjects underwent 120 minutes of continuous supine resting while continuous electrocardiographic (ECG), and respiration signals were collected; in half of each group, the recordings also include an uncalibrated continuous non-invasive blood pressure signal. Each subgroup of subjects includes equal numbers of men and women. All subjects remained in a resting state in sinus rhythm while watching the movie *Fantasia* (Disney, 1940) to help maintain wakefulness. The continuous ECG, respiration, and (where available) blood pressure signals were digitized at 250 Hz. Each heartbeat was annotated using an automated arrhythmia detection algorithm, and each beat annotation was verified by visual inspection. The R-R interval (interbeat interval) time series for each subject was then computed. The original data contain 10 samples with 72000 records. Among them only 1000 records randomly for developing the Support Vector Machine classification. The details of the records selected for training and testing the SVM classifier is given in the table II. Among these SVM classifier trained and tested data, which contain the development of QRS detection algorithm [25], in these signal passes through the low pass and high pass filter in order to reduce the influence of muscle noise, power line interference, T-wave interference. In these algorithm used to produce the several HRV features such as Statistical HRV features, Prediction based HRV features and Wavelet HRV features.

Table I Distribution of Data

Total Number of Samples : 10		
Data Distribution	Normal	Abnormal
Training: 5	3	3
Testing: 5	2	2

The SVM model is developed using MATLAB 7.5b in Pentium 4 with 2.40 GHz processor with 256 MB of RAM. The Support Vector Machine classifier contains the three processes such as input layer, support vector (hidden layer), output layer (decision). In these support vector machine classifies the trained and tested data which is based on the data point at the margin called the support vector. The performance of the support vector during the training and testing data is shown in fig 2

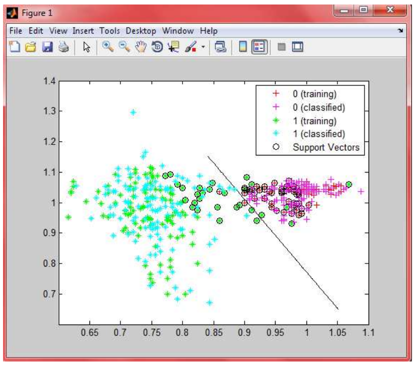

Fig 2 Training and Testing Performance of support vector

After training and testing data, the generalization performance of the support vector achieved by the result is 96.63%. The performance of the proposed Support Vector machine classifier, which is implementing the HRV features and it, is presented in the table II.

Table II Performance of proposed SVM model

Testing Performance	Normal	Abnormal	Total
No.of correctly classified samples	4	5	9
Percentage of classifier accuracy	80.4%	100%	90%

V. CONCLUSION AND FUTURE WORK

In this paper, a support vector machine classification for the heart beat time series was proposed. A simple support vector machine classifies the data with the performance of trained and tested data. The performance of the support vector machine was measured by using the tested data .Test Result shows that the proposed approach works well in detecting different attacks and is comparable to those reported in literature. Hence as an enhancement to this proposed work, future research will be directed towards the generalized discriminates methods which are used to compare with the approach of SVM.

REFERENCES

[1] M. Teich, S. Lowen, K. Vibe-Rheymer, and C.Heneghan, "Heart rate variability: measures and models," in Nonlinear Biomedical Signal Pro-cessing, vol. II, Dynamic Analysis and Modeling. New York: IEEE Press, 2001, pp. 159–213.

[2] G. Berntson, J. Bigger, D. Eckberg, P. Grossman, P. Kaufmann, M.Malik, H. Nagaraja, S. Porges, J. Saul, P. Stone, and M. van der Molen, "Heart rate variability: Origins, methods, and interpretive caveats," Psychophys-iology, vol. 34, no. 6, pp. 623–648, Nov. 1997.

[3] M. Kamath and E. Fallen, "Power spectral analysis of HRV: A noninva-sive signature of cardiac autonomic functions," Crit. Rev. Biomed. Eng.,vol. 21, pp. 245–311, 1993.

[4] R. Silipo, G. Deco, R. Vergassola, and C. Gremigni, "A characterization of HRV's nonlinear hidden dynamics bymeans ofMarkovmodels," IEEE Trans. Biomed. Eng., vol. 46, no. 8, pp. 978–986, Aug. 1999.

[5] M. Ferrario, M. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress," IEEE Trans. Biomed. Eng., vol. 53, no. 1, pp. 119–125, Jan. 2006.

[6] D. Hoyer, B. Pompe, K. Chon, H. Hardhalt, C. Wicher, and U. Zwiener, "Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales," IEEE Trans. Biomed. Eng., vol. 52, no. 4, pp. 584–592, Apr. 2005.

[7] V. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.

[8] L. Goldberger, L. A. N. Amaral, L. Glass, J.M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B.Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new re-search resource for complex physiologic signals," Circulation, vol. 101, no. 23, pp. e215–e220, Jun. 2000 circulation Electronic Pages[Online].Available:

[9] A.Bezerianos, S. Papadimitriou, and D. Alexopoulos, "Radial basis function neural networks for the characterization of heart rate variability dynamics," Artif. Intel. Med., vol. 15, no. 3, pp. 215–234, 1999.

[10] B.Aysin, L. Chaparro, I. Grav´ e, and V. Shusterman, "Orthonormal basis partitioning and time frequency representation of cardiac rhythm dynamics," IEEE Trans. Biomed. Eng., vol. 52, no. 5, pp. 878–889, May 2005.

[11] A.Alexandridi, C. D. Stylios, and G. Manis, "Neural networks and fuzzy logic approximation and prediction for HRV analysis," presented at the Eur. Symp. Intel. Technol., Hybrid Syst. Implement. Smart Adapt. Syst., Oulu, Finland, Jul. 2003.

[12] G.Pajares and J. M. de la Cruz, "On combining support vector machines and simulated annealing in stereovision matching," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 34, no. 4, pp. 1646–1657, Aug. 2004.

[13] R.Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," IEEE Trans. Biomed. Eng., vol. 52, no. 5, pp. 828–838, May 2005.

[14] G.Manis, S. Nikolopoulos, A. Alexandridi, and C. Davos, "Assessment of the classification capability of prediction and approximation methods for HRV analysis," Comput. Biol.Med., vol. 37, no. 5, pp. 642–654, 2007.

[15] T.N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Bir-baunner, and B. Scholkopf, "Support vector channel selection for BCI," IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[16] M.Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-Data set IIb: Support vector machines for the P300 speller paradigm," IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1073–1076, Jun. 2004.

[17] S.Osowski,L.T.Hoai, andT.Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," IEEE Trans. Biomed. Eng., vol. 51, no. 4, pp. 584–589, Apr. 2004.

[18] I.G¨ uler and E. D. ¨ Ubeyli, "Multiclass support vector machines for EEG signals classification," IEEE Trans. Inf. Technol. Biomed., vol. 11, no. 2, pp. 117–126, Mar. 2007.

[19] L.Ramirez, N. Durdle, D. Hill, and J. Raso, "A support vector classifier approach to predicting the risk of progression of adolescent idiopathic scoliosis," IEEE Trans. Inf. Technol. Biomed., vol. 9, no. 2, pp. 276–282, Jun. 2005.

[20] K.I.Kim,K. Jung, S.H. Park, andH. J.Kim, "Support vector machines for texture classification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 11, pp. 1542–1550, Nov. 2002.

[21] I.El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcifications," IEEE Trans. Med. Imag., vol. 21, no. 12, pp. 1552–1563, Dec. 2002.

[22] G.Georgoulas, C. Stylios, and P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal clas-sification using support vector machines," IEEE Trans. Biomed. Eng., vol. 53, no. 5, pp. 875–884, May 2006.

[23] A.Kampouraki, C. Nikou, and G. Manis, "Robustness of support vector machine-based classification of heart rate signals," in Proc. IEEE Conf. Eng.Med. Biol. Soc. (EMBS 2006),NewYork,Aug.–Sep., pp. 2159–2162.

[24] R. Kondor and T. Jebara, "A kernel between sets of vectors," in Proc. 20th Int. Conf. Mach. Learn. (ICML), Washington, DC, 2003, pp. 361–368.

[25] J. Pan andW. J. Tompkins, "A real-time QRS detection algorithm," IEEE Trans. Biomed. Eng., vol. 32, no. 3, pp. 230–236, Mar. 1985.

A.Sasirekha is currently working as a Assistant Professor in Info Institute of Engineering, coimbatore. She received her B.Tech degree in Information Technology, M.Tech degree in Mainframe Technology under Anna University Chennai. Her research interest is in Data Mining and Bioinformatics.

P. Ganesh Kumar is currently working as a Assistant Professor in the Department of Information Technology, Anna University of Technology, Coimbatore. From July 2003 to August 2008, he worked as a Lecturer in the Department of Information Technology, Kalasalingam University, Krishnankoil. He received his BTech degree in Information Technology from the University of Madras in May 2003 and MS (By Research) degree in Information Technology from Anna University Chennai in September 2008. He has published 6 international journals and 15 international conferences. His research interest is application of soft computing techniques in data mining, and bioinformatics.