

Statistical and Learning Aided Classifier for ECG Based Predictive Diagnostic Tool

Sumi Bharali¹, Manash Pratim Sarma¹, Kandarpa Kumar Sarma² and Nikos Mastorakis³

¹ Department of Electronics and Communication Engineering, Guwahati, Assam, India

² Departments of Electronics and Communication Technology, Gauhati University, Guwahati-781014, Assam, India. email- kandarpaks@gmail.com

³ Technical University - Sofia, Sofia 1000, "Kl. Ohridski" 8, Bulgaria. email- mastor@tu-sofia.bg

Abstract. Early diagnosis and classification of long term cardiac signals are crucial issues in the treatment of heart related disorders. The available number of medical professional are not sufficient to deal with the increase patients for which design of certain machine based diagnostics tools have been accepted as a viable option. Typical Electrocardiogram (ECG) machine is helpful for monitoring the heart abnormalities only for short interval of time. Therefore, it becomes necessary to design a system which captures relevant features of the ECG signal for use with certain classifiers. In our proposed system, ECG signal elements like Q, R and S peaks are detected and heart rate estimated using Linear Discriminant Analysis (LDA), Adaptive Linear Discriminant Analysis (ALDA) and Support Vector Machine (SVM). For our work we have been used MIT BIH (Standard Arrhythmia Database).

1 Introduction

Electrocardiograph (ECG) is an instrument used in the detection and diagnosis of heart abnormalities that occur in human's body. Early diagnosis and classification of long term cardiac signals are crucial issues in the treatment of heart related disorders [1]. The available number of medical professional are not sufficient to deal with the increasing number of patients for which design of certain machine based diagnostics tools have been accepted as a viable option. ECG is used for the diagnosis of heart related disorders. Such a machine measures the electrical potentials of a patient by placing electrodes on body surface and in this way system records continuous activity of heart muscle. The study of ECG signal deals with its morphology (such as amplitude, duration, segments). Normal frequency and amplitude range of ECG signal is in between 0.05-100 Hz and 1-10 mv. Fig. 1 shows a normal ECG signal. However, the ECG signal is non-stationary; it is very difficult to analyze such a big volume of data. Therefore, it becomes necessary to design a system which captures relevant features of the ECG signal for use with certain classifiers. In our proposed system, ECG signal elements like Q, R and S peaks are detected and heart rate estimated using Linear Discriminant Analysis (LDA), Adaptive Linear Discriminant Analysis (ALDA) and Support Vector Machine (SVM). For our work we have been used MIT BIH (Standard Arrhythmia Database).

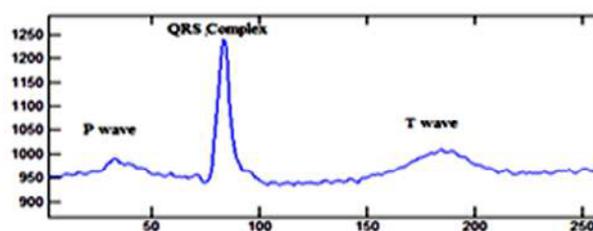


Fig.1 ECG signal.

Section II describes about the detection of QRS complex using Pam-Tompkins Algorithm and gives formulation of proposed classifiers LDA, ALDA and SVM. Section III gives description of simulation result and comparison of classifier schemes. Section IV gives the conclusion. Some of the relevant literature are [1]-[10]

2 Proposed Method

Several algorithms have already been proposed yet there are challenges which are the primary motivation behind this work for automatic detection of QRS complex. The QRS complex corresponds to the largest wave, since it represents the depolarization of the right and left ventricles, which are the chambers with the most substantial mass in the heart. Variability of heart rate is calculated from the parameter R-R interval which is the time difference between two successive R

* Corresponding author: kandarpaks@gmail.com

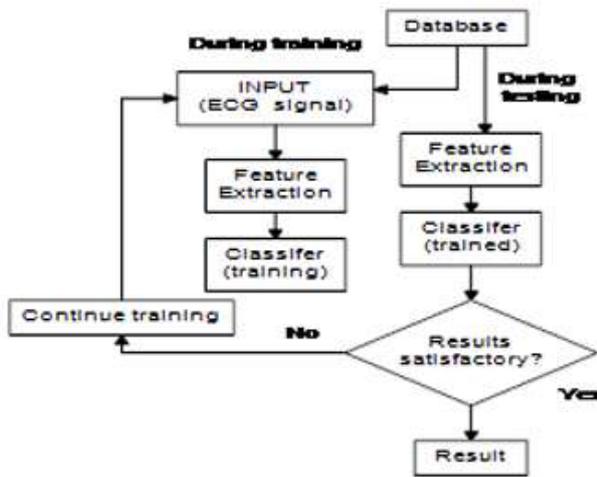


Fig.2 Overall block diagram of the system.

waves. The QRS complex is the fundamental feature of ECG signal. The peaks detected in the signal are considered as feature vector. In this work, QRS complex waveform is detected using Pam-Tompkins algorithm which also is applicable real time environment [2]. Heart rate is calculated from the position of consecutive R-R interval. R-R interval is calculated by subtracting consecutive location of R positions.

A. System Block Diagram

The architecture is divided into two parts feature extraction and classification of Q, R, S and HR beats using three methods LDA, ALDA analysis and SVM. It is shown in fig. 2. There are two aspects first during training, the system becomes familiar with the samples and during testing, and the ability of the system is ascertained.

Next we discuss the method of extraction of QRS complex positions and calculation of the heart rate. Here, ECG signal is taken as input and processes performed for feature extraction over each normal ECG signal. As the signals are normal and database is clean, so there is no need to do the preprocessing part of the signals. Fig. 2 show systematic stages for extracting feature from ECG signal. ECG signals have undergone differentiation operation for highlighting the QRS complex and zero crossing of ECG signal. Differentiated signal is next passed through squaring the signal for making all the negative points positive. The squaring function is given by

$$y(nT) = x(nT) \tag{1}$$

The next block of Pam-Tompkins algorithm is moving average integrator. In this block, a window is specified for detection of number of peaks within the duration of single QRS complex. The equation of moving average window is given by

$$y(nT) = \left(\frac{1}{N}\right) [x(nT - (N - 1)T) + x(nT - T) + \dots + x(nT)] \tag{2}$$

where N is the number of samples in window. Window width defines the number of samples present in the QRS complex. The processed signal at this stage thresholded for identification of QRS complex. Threshold value is based on rising edge of QRS complex. Here it is taken in between 20- 25. QRS complex is further searched after 293 sample points from the previous cycle. If the signal is greater than threshold value then the position of QRS complex is chosen as the position of original ECG signal. Now plotting both original and QRS complex in a same plot we have seen that within a window, QRS complex is obtained. The position of Q and S peaks are the first and last position of window and R peak is the middle one. Now we get the location of peaks but the peaks values are not in the expected physiological range. Since each sample of the signals are digitized into 2N levels where N is the resolution of the device for collecting signal into computers the memory segment is a factor. So one way to convert digital data into physical units is to subtract the baseline and then divide by its gain for its channel to obtain physical units. So in these way features are obtained in physical range. Baseline and gain information of the signal is obtained from the MIT-BIH database. Fig. 3 shows the detected QRS complex.

The heart rate is calculated using equations as follows
 $HR=60 \times \text{sampling frequency} / (R - R \text{ interval})$
 (3)

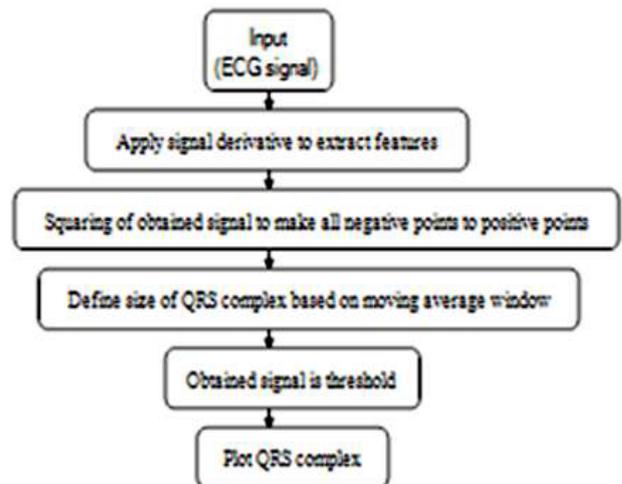


Fig.3 Evaluating QRS Complex

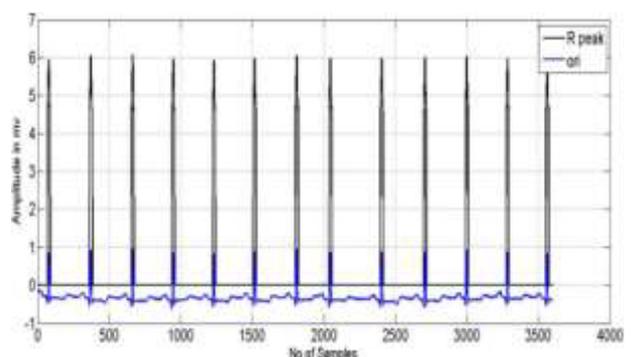


Fig. 4 QRS Complex Detected Signal no 100

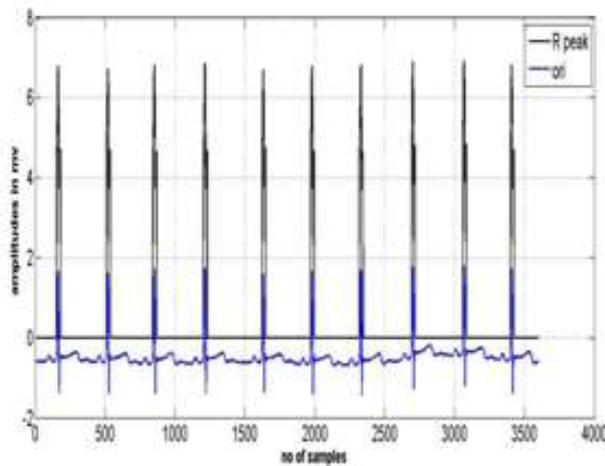


Fig.5 QRS Complex Detected Signal no 115

B. Classifier design

In this work, three classifiers are used i.e. LDA, ALDA and SVM classifier. The description is as below.

i) LDA Classifier-

LDA is a supervised method used in statistics, pattern recognition and machine learning for separating two or multiple classes into their desired classes. LDA is linear transformation technique [3]. Here classification is totally dependent on projection vector. Training data sets are projected in the direction of projection vector to maximize its separability. System model is tested by evaluating the Euclidean distance between train and test dataset. Let's say that we have m - classes and each class has n -dimensional sample points $(x_1, x_2, x_3, \dots, x_n)$. So stacking all sample values in one matrix where each row represents one feature values of different samples we get a sample matrix. Each row signifies one class. We are intend to do separation up to four classes hence we have four rows in the matrix. Fig. 6 shows the logic approaches to evaluate projection vector of LDA classifier. The system model of LDA classifier as shown in fig. 7

ii) ALDA classifier

This is a technique for prediction of how much actual and predicted data of a classifier are identical by varying its weights using weight update equation. The purpose of ALDA is to analyze [4] the closeness of the current value to the reference samples and also to minimize the error between true and obtained prediction. Fig. 8 shows logical approaches to update the weight vector. Here X 's, Y_{new} are the train data and test data respectively. Initially W is taken as random weight vector. The system model of ALDA classifier as shown in fig. 9

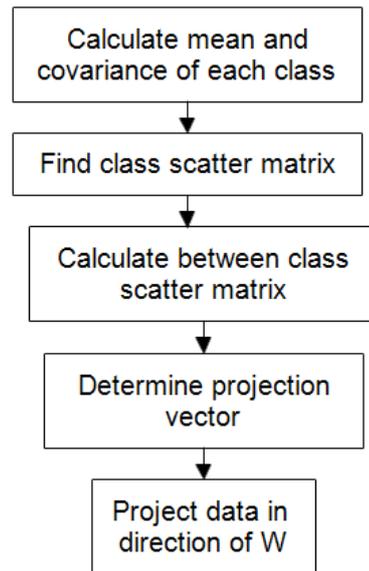


Fig.6 Logical steps to find projection vector of LDA

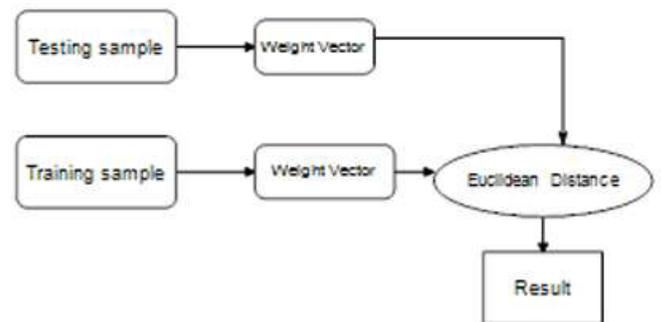


Fig.7 Block diagram of LDA classifier

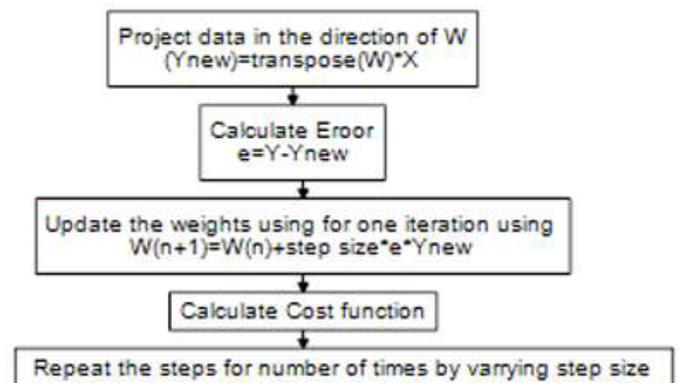


Fig.8 Block diagram for obtaining weight vector

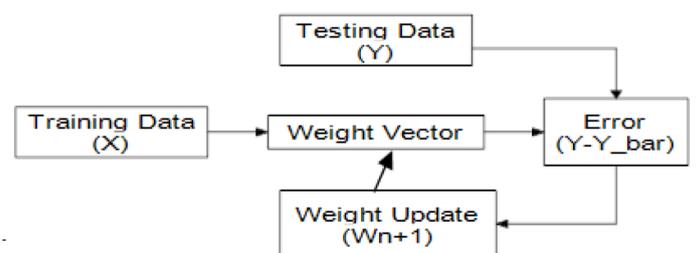


Fig.9 Block diagram of ALDA Classifier

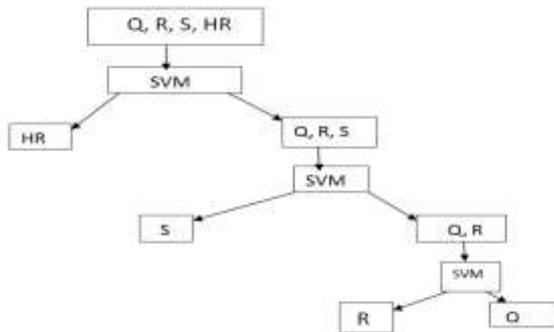


Fig.10 Block Diagram of SVM Classifier

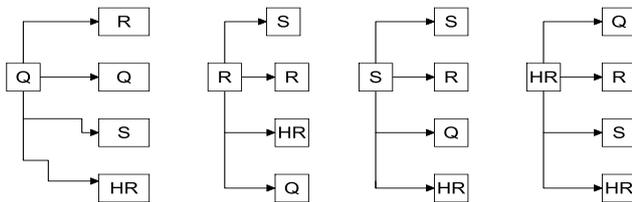


Fig.11 Euclidean distance computed using LDA classifier

iii) SVM classifier

Support Vector Machine is a supervised learning machine [6]. The main objective of SVM is to separate of two or multiple classes based on hyper-plane. With SVM kernel function, we can transform non-linear data into higher dimensional space. SVM focuses on data points which are difficult to classify. Such difficult points are close to decision boundary or hyper planes identified by support vectors. The proposed multiclass SVM uses three SVM to classify Q, R, S and HR beats of normal patient. The system model of SVM classifier as shown in fig. 10

3 Experimental Details and Results

Here we discuss the results derived. Initially database is divided into two sets testing and training data sets one for testing and other for training. Sizes of both the data sets are $(N \times M)$, where N stands for number of features and M stands for feature length. Calculated features are fed in to LDA classifier. Sizes of both data sets are made equal for calculating the Euclidean distance for each sample value of testing and training data sets. Both sets are projected in the direction of projection vector. Initially the projection vector is found to be one. So taking one as weight vector both testing and training data sets are projected in the direction of weight vector and finally computing the Euclidean distance for measuring similarity score among features. Euclidean distance is computed among this combination as shown in fig. 11. Before finding Euclidean distance data should be in normalized form, otherwise small feature values are dominated by higher values and gives fluctuating results.

Confusion matrix takes simply two values one of which is the target values of each feature of test data and the other is the predicted value of each feature generated

after calculating Euclidean distance between two vectors. Here diagonal elements represent the true prediction each class given test data. In Table 2 some parameters like sensitivity, specificity, accuracy, misclassification rate and precision are calculated. Overall accuracy of the system is found to be is 88.9 %. For calculating the five parameters the following equations are used:

$$\text{Sensitivity} = TP / (TP + FN) \tag{4}$$

$$\text{Speccificity} = TN / (TN + FN) \tag{5}$$

$$\text{Accuracy} = TP + TN / (TN + FP + TP + FN) \tag{6}$$

$$\text{Misclassification Rate} = FP + FN / (TN + FP + TP + FN) \tag{7}$$

$$\text{Precision} = TP / (TP + FP) \tag{8}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative. In Table 1, confusion matrix values are shown derived from LDA classifier during Q, R, S and HR estimator. The next sets of results are derived using ALDA classifier. Initially a random weight vector is taken and projected the first sample point of training data set (X) in the direction of given weight vector. After that an error is calculate between testing data set and projected data set for updating the weight vector. The obtained new weight vector is now ready projecting the next sample point of training data. Similarly, again the error difference between known and predicted sample points for updating the next weight vector. This process will continue for every sample point of training and testing data sets. Now, for repeating these steps for multiple iteration, we take the last weight vector of the respective cycles of entire data set and feed this vector in to the next cycle. In this way, we will get mean square error for every cycle of iteration. The MSE is used as the cost function. The results show variation when step size of weight update is varied. First we take a step size of 0.004. Fig. 12 shows cost function for 10 numbers of iteration by taking step size 0.004. Table 3 show confusion matrix generated for step size 0.004 using ALDA for Q, R, S and HR calculator. Results are also derived for step sizes of 0.001 and 0.0001. Results are shown in fig.s 13 and 14 and confusion matrix values are shown in Tables 4 and 5.

From fig.s 12 to 14 we see that with decreasing step size and increasing the iteration, system is responding well and error difference between known and unknown data sets are becoming less. This process is the basis of ALDA. Table 6 shows some parameters are calculated for step size .004, 0.001 and 0.0001. Table X shows different parameters calculated using SVM.

Table 1 : Confusion matrix of LDA with Q, R, S and HR estimate

Actual/ Predicted	Q	R	S	HR
Q	31	0	7	0
R	3	35	0	0
S	7	0	31	0
HR	0	0	0	38

Table 2 : Parameters obtained from the LDA classifier

Parameters	LDA
Sensitivity	0.887
Specificity	0.962
Accuracy	.887
Misclassification Rate	0.2233
Precision	0.8827

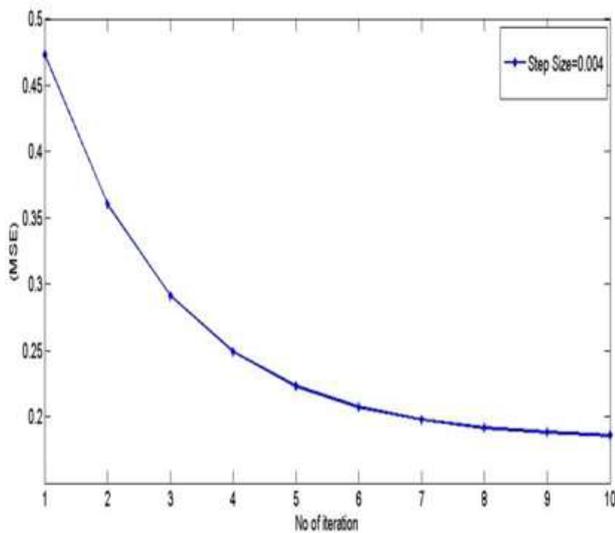


Fig.12 Cost function obtained for step size 0.004 and 10 number of iteration

Table 3. Confusion matrix

Actual/ Predicted	Q	R	S	HR
Q	29	0	9	0
R	0	35	0	3
S	7	0	31	0
HR	0	0	0	38

In this paper, feature classification has been done using LDA, ALDA and SVM technique. Overall accuracy of the LDA system is found to be 88.7%. We also define a cost function to quantify the error between actual and predicted data by varying its step size and updating the weight vector for various numbers of times. The best result come when step size is change to 0.0001 and iterates the system for 550 numbers of

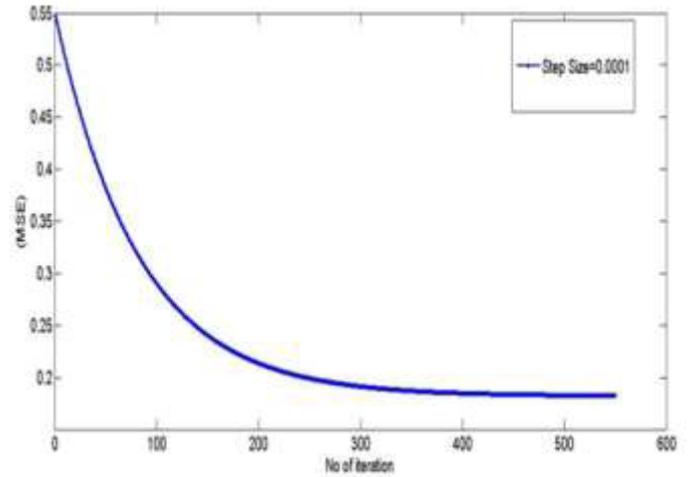


Fig.13 Cost function obtained for step size 0.001 and 20 number of iteration

Table 3- Confusion matrix

Actual/ Predicted	Q	R	S	HR
Q	28	0	10	0
R	0	35	0	3
S	6	0	32	0
HR	0	0	0	38

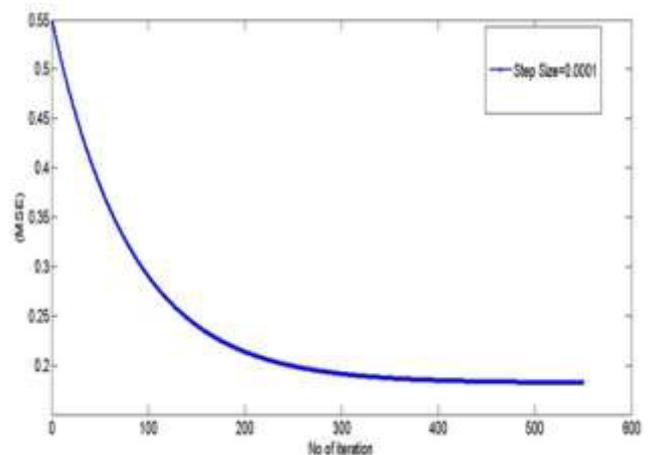


Fig.14 Cost Function obtained for step size 0.0001 and 550 number of iteration

Table 4 : Confusion matrix

Actual/ Predicted	Q	R	S	HR
Q	31	0	7	0
R	0	35	0	3
S	4	0	34	0
HR	0	0	0	38

Table 5 : Parameters calculated for a LDA classifier

Parameters	ALDA Step size=0.004	ALDA Step Size=0.001	ALDA Step size=0.0001
Sensitivity	0.8749	0.8747	0.9082
Specificity	0.958	0.9585	0.9541
Accuracy	0.8749	0.88	.9078
Misclassification Rate	0.23	0.2495	0.1835
Precision	0.876	0.8776	.9101

Table 6 : Confusion matrix from IRST node (HR beat detected)

Actual/Predicted	No HR (Remaining)	Yes HR
No HR (Remaining)	101	0
Yes HR	0	6

Table 7 : Confusion matrix from second node (S beat detected)

Actual/Predicted	NoS (Remaining beat)	Yes S
NoS (Remaining beat)	62	8
Yes S	14	17

Table 8 : Parameters calculated from SVM

Parameters	SVM
Sensitivity	0.911
Specificity	0.9535
Accuracy	0.9455
Misclassification Rate	0.059
Precision	0.9015

Finally, another set of results derived using SVM. Initially data is divided into two groups and one group labeled as +1 and another as -1. In fig. 10, first node contains all the features to classify the classes into their receptive groups. So after training, labeling and testing the system six sample points are detected as belong to one group i.e. (+1) and 101 sample points were detected as belong to another group (-1). Now 101 test sample points are from test set data for testing remaining classes. At second node system is trained and labeled as +1 (when train data<-0.48) and -1 (when train data>-0.48). At third node again SVM is applied to the remaining trained dataset and misclassified data obtained from the second node. Here system is trained and labeled as +1 if (train data is greater zero) and as -1 (when train data is less than zero but greater than -0.48). In Tables 6 and 7, confusion matrices for first node for predicting HR and second node with S beat detected are shown respectively. Table 8 includes some of the parameters derived from the SVM. Table 9 includes a comparison of the performance.

Table 9 : Comparison of classifier performance in terms of accuracy

Classifier		
Linear Discriminant Analysis	Adaptive Linear Discriminant Analysis	Support Vector Machine
88.7 %	90.8%	94.55%

In this paper, features classification has been done using LDA, ALDA and SVM technique. Overall accuracy of the LDA system is found to be 88.7%. We also define a cost function to quantify the error between actual and predicted data by varying its step size and updating the weight vector for various numbers of times. The best result come when step size is change to 0.0001 and iterates the system for 550 numbers of times and accuracy of the system is increased to 90.8%. Finally a combination of both linear and non-linear classifier SVM is used to classify both linear and non-linear data using linear kernel, polynomial kernel and RBF kernel. SVM yield 94.55% when kernel parameters of RBF are set to 0.9 for sigma and 0.2 for penalty parameter C. Hence comparing three methods we find that the SVM provides better accuracy compared to LDA and ALDA classifier.

4 Conclusion

Here, Pam-Tompkins algorithm is used for features extraction of ECG Signal. With this algorithm Q, R and S peak is searched for every cycle of ECG signal and the heart rate of patients. The calculated features vectors are fed to the LDA, ALDA and SVM classifier. The outcome of this work is that results derived from LDA and ALDA are compared with that obtained from SVM while estimating QRS and HR of patients taking ECG signals. The proposed approach shall be helpful in designing a diagnostic support tool.

References

- [1] E. J. S. Luz, W. R. Schwartz, G. C. Chavez and D. Menotti, "ECG-based Heartbeat Classification for Arrhythmia Detection: A Survey", in proceeding of Computer Methods and Programs in Biomedicine, Vol.127, pp 2-22, April 2016.
- [2] A. Kaushik and P. Sabherwal, "Monitoring Electrocardiogram using Android based Smart phone", in

preceding of Annual IEEE India Conference (INDICON), pp 1-6, 2015.

- [3] S. Balakrishnama, A.Ganapathiraju, “Linear discriminant analysis-a brief tutorial”, in Institute for Signal and information Processing , vol. 18, pp. 1-8, 1998.
- [4] M. H. Hayes, “Statistical Digital Signal Processing and Modelling”, 1st edition, Wiley, 1996.
- [5] U. Desai, R. J. Martis, C. G. Nayak and G. Seshikala, ”Machine Intelligent Diagnosis of ECG for Arrhythmia Classification Using DWT, ICA and SVM Techniques”, in proceeding of IEEE INDICON, pp 1-4, 2015.
- [6] A. S. Thakare, R. Ghongade and Prof. A. S. Barhatte, “QRS Complex Detection and Arrhythmia Classification using SVM”, in preceding of International Conference on Communication, Control and Intelligent Systems (CCIS), pp 239-243, 2015.
- [7] E. M. Imah, F. A. Afif and M. Ivan.Fanany, ”A Comparative Study on Daubechies Wavelet Transformation, Kernel PCA and PCA as Feature Extractors for Arrhythmia Detection Using SVM”, in proceeding of IEEE Region 10 Conference, pp 5-9, 2011.
- [8] P. Sarma, S. R. Nirmala and K. K Sarma, “ ECG Classification using Wavelet Subband”, in preceding of International Conference on Signal Processing and Integrated Networks (SPIN), pp 785-790, 2014.
- [9] P. Sarma, S. R.Nirmala and K. K.Sarma, “ Classification of ECG using Some Novel Features”, in Proceeding of IEEE ICETACS, 2013.
- [10] K. Lochan, P. Sah and K. K. Sarma, “Innovative feature set for retinopathic analysis of diabetes and its detection”, IEEE NCETACS, 2012.