Detection and Classification of R-Peak Using Naïve Bayes Classifier

S. Celin[1*], K. Vasanth[2]

[1]*Research Scholar, Sathyabama Institute of Science and Technology, Chennai.*
[2]*Professor, Vidya Jyothi Institute of Technology, Hyderabad.*

Abstract

Electrocardiogram (ECG) in classification of signals plays a major role in the diagnoses of heart diseases. The main challenging problem is the classification of accurate ECG. Here in this paper the ECG is classified into arrhythmia types. It is very important that detecting the heart disease and finding the treatment for the patient at the earliest must be done accurately. In the ECG classification different classifiers are available. The best accuracy value of 99.7% is produced by using the Bayes classifiers in this paper. ECG databases, classifiers, feature extraction techniques and performance measures are presented in the pre-processing technique. And also the classification of ECG, analysis of input beat selection and the output of classifiers are also discussed in this paper.

Keywords: ECG signal, butter worth filter, support vector machine, adaboost, ANN, naïve bayes, MIT-BIH arrhythmia database.

1. Introduction

Electrocardiogram is the abbreviation of the ECG and it is mainly used to detect the electrical activity of the heart. The hearts muscle relaxation and contraction are mainly shown. Electrocardiography is another word which is used to give or provide the condition of the heart. Analysis of the ECG waveform is used to identify the heart's normal and abnormal behavior. P, Q, R, S, T and U are some of the basic wave forms used in the ECG waveform basically. The heart conditions are analyzed by these waveforms. The most important one in ECG is depolarization and repolarization. P waveform is represented in atrial depolarization and QRS complex waveform is represented in repolarization and then T waveform of the ventricle is represented in depolarization [1]. However, even for an expertized clinician, investigating various abnormal rhythmic changes from the long ECG record is very exhausting. Hence in arrhythmia identification, computer-aided diagnosis plays a vital role and it is mainly owing to its robustness and effectiveness. In the given ECG, arrhythmia detection follows the identification of successive heartbeat classes. Therefore, in heartbeat classification, arrhythmia plays an important step in recognizing [2].
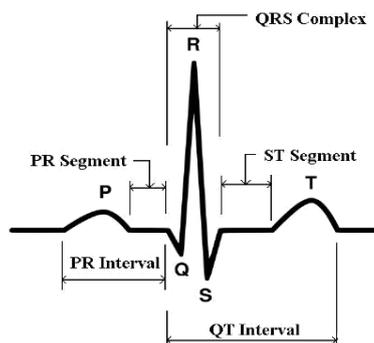

Figure 1: ECG waveform

Fig.1 shows an ECG waveform which consist of P wave, T wave, QRS complex and various intervals [3,4]. Most importantly, one heartbeat is represented by the R-R interval. The QRS detection algorithm can be used for removing baseline wonder [5-6].

For providing the system with accuracy and sensitivity, there are several existing methods for ECG waveform analysis. Wavelet, Artificial Neural network [7-10], Ada boost Classifier, Support Vector Machine, Naive Bayes Classifier are some of the various techniques based on these methods. There are several techniques in which it can analyze the ECG signal and it is described in [11-15]. In feature extraction [16] wavelet transforms is used as a tool for processing non stationary signals like ECG signals.

The ECG signals are analyzed with the combination of recurrent neural networks and eigenvectors [17]. By using particle swarm optimization and radial basis function neural network, another combination of novel method can be analyzed by the ECG signal [18]. A very popular algorithm to use in practice is the Ada Boost boosting algorithm has become popular over the last few years. Simplicity and adaptivity are the two main reasons for this popularity. Therefore we say that Ada Boost [19] is adaptive because the amount of update is chosen as a function of the weighted error of the hypotheses generated by the weak learner.

In this paper evolvable Naive bayes is presented from the ECG signal for the classification of abnormal heartbeat patterns. For personalized health monitoring the associated weights of a Naive bayes and internal structure are evolved to classify heartbeat patterns of each individual. The Naive bayes use both morphological and temporal features extracted from the ECG signal. The evolved Naive bayes produced an average of 99.7% classification accuracy of the ECG signals from the MIT-BIH Arrhythmia database [20].

The paper is organized as follows. In Section 2, describes the methodology of proposed system, In Section 3, discusses about experimental results of the proposed method. And in the section 4 discusses about conclusion.

2. Proposed Model

These are the following steps followed in Arrhythmia detection algorithm.

(a) ECG signal is pre-processed.
(b) Features extraction.
(c) The features of ECG signals are classified.

Figure (2) shows the proposed model block diagram of the whole algorithm. The system is based on Wavelet Parameters and four different classifiers.
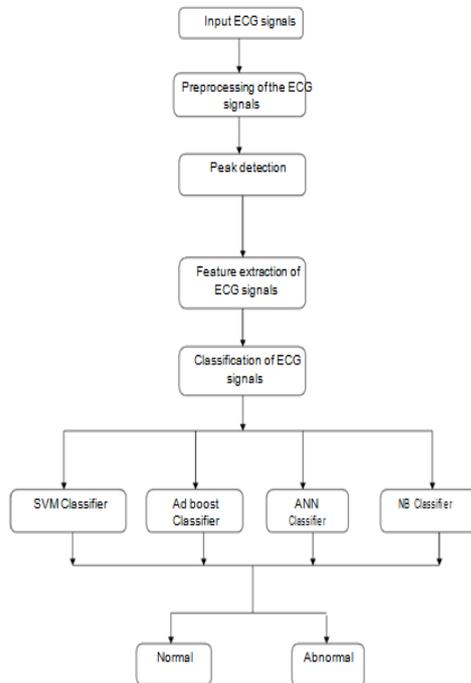

Figure 2: Block diagram of proposed system

2.1. Input ECG Signals

The database for ECG signal is taken from MIT-BIH arrhythmia database. The recording is digitized by 360 samples per second per channel with a resolution of 10mv range.

2.2. Pre-Processing

In signal and image processing, the preprocessing is one of the most important tasks. For proposed ECG classification, this is the first step. By using various filter approaches, the noise is eliminated in the input ECG signal. This is the main work of the preprocessing. For noise removal in ECG signal, the proposed work handled high pass filter, median filter, low pass filter and Butterworth filter [21]. Preprocess result is used to get the better efficiency of the ECG signal. Due to preprocessing work, the peak detection efficiency is increased. To remove the noise of power interference, baseline wander, and instrumental error, various filters approaches are used. is Therefore in the ECG signal, the elimination of baseline wander analysis is needed to diminish the irregularities in beat morphology.

2.3. Peak Detection

It is a standard technique for R peak detection normally distinguish the QRS complexes from the ECG waves. Since it is basically a high-pass filter, the derivative amplifies the higher frequencies characteristic of the QRS complex while attenuating the lower frequencies of the P and T waves.

2.4. Feature Extraction (Wavelet Parameters)

Wavelets are being useful in various fields of science and engineering such as signal and medical imaging processing etc. Wavelet research is proceeding at a rapid space and every year a new development appears, expanding the new domain of wavelet analysis. The wavelet transform gives a large number of small coefficients and a small number of large coefficients. Large coefficients mainly represent the signal values and the coefficients with smaller values represent the noise components. Some characteristics which make them useful are:

- Wavelets are localized in both time and frequency
- For analyzing non-stationary signals such as ECG and which have jumps and non-smooth features.
- Wavelet separates a signal into multi resolution components.

Two type of wavelet parameters are used they are CWT and DWT.

2.4.1. Continuous Wavelet Transform (Cwt)

The wavelet means a "small wave", which maps a function f(t) into the time-scale plane and is denoted by W (a,b) and denoted by the following equation

$$W_S(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \, h * \left(\frac{t-b}{a}\right) dt \qquad (1)$$

Where h(t) is called mother wavelet, a is the scaling parameter in y-axis and b is the shift parameter in x-axis. The signal f(t) can be recovered from the wavelet coefficients (,) by the inverse wavelet transform and is given by

$$f(t) = \frac{1}{c} \iint_{-\infty}^{\infty} W_S(a,b) h\left(\frac{t-b}{a}\right) \frac{d_a}{a^2} \, db \qquad (2)$$

With admissibility condition given as

$$c = \int_{-\infty}^{\infty} \frac{w^2}{W} dw < \qquad (3)$$

2.4.2. Discrete Wavelet Transform (DWT)

In practice, when implemented on computers, CWT must be discrete. The redundancy that is present in CWT can be reduced by discretizing the transform parameters a and b but not the time t parameter as follows.

$$a = ao^m, b = nbo \, ao^m \quad m,n \in Z \; ao \neq 1, bo \neq 0 \qquad (4)$$

For fine resolution has to be chosen sufficiently close to 1 and close to 0.

2.5. Classification of ECG Signals

Feature extraction output is given as an input to classifier. In this paper four types of classification are carried over to classify as normal and abnormal heart beat that are ANN, SVM, Adaboost and Navie Bayes.

2.5.1. SVM Classifier

Support Vector Machine (SVM) is one of the learning system and it is used mainly in classification. In 1998, SVM classifier was developed by Vapnik and for solving the supervised classification problem it is one of the most
techniques. In between training data and the decision boundary (optimal separating hyperplane) in essence, SVM classifiers is maximized to the margin, which can be formulated in a feature space as a quadratic optimization problem. The subset of patterns

those are closest to the decision boundary are called support vectors.

SVM is analyzed by

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \qquad (5)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \qquad (6)$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \qquad (7)$$

In the above three equation TP denotes the number of true positive samples, FN indicates the number of false negative samples, TN denotes the number of true negative samples and FP indicates the number of false positive samples. These TP, TN, FP and FN are used for classification and it is defined as
FP: Normal class classifies as abnormal
TP: Abnormal class classifies as abnormal
FN: Abnormal classifies as normal.
TN: Normal class classifies as normal

2.5.2. Ada Boost Classifier

In machine learning algorithms, ensemble classifiers are playing a predominant role in recent times. Especially, in various applications these classifiers are used to address the class imbalance problem [22,23]. By introducing and aggregating several classifiers, the main aim of ensemble classifier is to reduce the misclassification rate (error rate) of a weak classifier. The basic idea is to get predictions of several classifiers on the original data and by combining all the various predictions, this makes it a strong classifier. Bootstrap aggregating and Boosting are the foremost strategies in ensemble classifiers. The commonly used learner is decision tree based learners. In our work, as base learner we use adaptive boosting (AdaBoost) ensemble implementation with the random forest. Freund has been developed the AdaBoost algorithm. It is one of the first practical boosting algorithm. The basic idea of AdaBoost can be written as a weighted combination of the weak learners as follows:

$$H(x) = \sin(\sum_{t=1}^{T} \alpha_i \quad h_t(x)) \qquad (8)$$

Where h ( ) is the weak classifier, weight chosen such that error rate is miminum.

2.5.3. Ann Classifier

ANN stands for Artificial Neural network. Artificial neural network (ANN) are basically an inspired network which is inspired by the human brain in an organization of the neuron. Also it consists of a decision making process which is more useful in the application of the areas like classification, pattern recognition, etc. The decision making process in the Artificial Neural Network is of holistic. This is mainly based on the aggregation of the entire input pattern. And also the conventional computer has to wade through the individual data elements in the processing of data elements individually for the arrival of the conclusion.

Due to their massively parallel structure, the neural networks derive their power, and it has the ability to learn from the experience. For fairly accurate classification, they can be used to input data into categories, by which they are previously trained for doing so. Depending on the efficacy of training, the accuracy of the classification is done, which in turn depends upon the depth and rigor of the training. In the form of connection weights the knowledge gained by the learning experience is stored, which can be used to make decisions on fresh input.

2.5.4. Naive Bayes Classifier

Classification describes more important data classes and it is a form of data analysis which extracts the models. Because of high accuracy, prediction rate and automatic method to search for hypothesis, Classification has numerous applications, including performance prediction, target marketing, fraud detection, manufacturing, and medical diagnosis. Algorithms like NB, SVM, decision trees, hiking etc. are used for diverse sorts of classification and prediction. The classification of ECG signal as Normal and Abnormal implemented in Rapid Miner tool, the Naïve Bayes (NB) classifier is used in this phase. The Class membership probabilities such as the probability that a given tuple belongs to a particular class is predicted in the NB classifiers which is a statistical classifiers. High accuracy and speed is exhibited in the Bayes classifiers and it is also applied to large databases. Naïve Bayes classifier is independent of the values of the other attributes and it is assumed that the effect in the attribute value is in a given class. This assumption is called as the class conditional independence. This is made to simplify the computations involved and also in a sense, it is considered as the naïve.

Table 1: Performance comparison

Literature	Features	Classifier	Performance measures in %
Hu et al. [25]	Time domain features	Mixture of experts (MOE)	94.0 ACC, 82.6 SEN, 97.1 SPE, 77.7 PPV
De Chazal et al. [26]	Morphological + heart beat intervals + R–R intervals of ECG	Morphological + heart beat intervals + R–R intervals of ECG	73 SEN, 96 SPE, 45 PPV
Kutlu and Kuntalp [27]	HOS of wavelet packet decomposition (WPD) + morphological features + energy spectral density of DFT	kNN	85.59 SEN, 99.56 SPE, 95.46 PPV
Kutlu and Kuntalp [28]	HOS of WPD	K-nearest neighbor (KNN)	90 SEN, 98 SPE, 92 PPV
Ye et al. [29]	(DWT + Independent component analysis (ICA)) + R-R intervals	SVM	86.4 ACC
Das and Ari [30]	S-transform (ST) +DWT +temporal	Multilayer perceptron (MLP)-NN	97.5 ACC
Afkhami et al. [31]	Statistical features + R-R intervals	Gaussian mixture model-expectation maximization (GMM-EM)	96.15 ACC, 93.3 SEN, 88.99 PPV
Proposed Work	Statistical features + R-R intervals	SVM, Adaboost, ANN, Naïve bayes	87.5 ACC, 93ACC, 94 ACC, 99.7 ACC

3. Result and Discussions

The proposed method uses ECG signal and it is taken from MIT-BIH database and classify the output as normal or abnormal signal. The input signal is preprocessed by using filter methods such as low pass, high pass and butter worth filter. Detect the peak for the QRS waveform and extract the features for the HRV signal. Classify the signal using four techniques such as ANN, SVM, Naïve Bayes, Adaboost classifier. The experimental result is simulated by using Matlab 2016a.
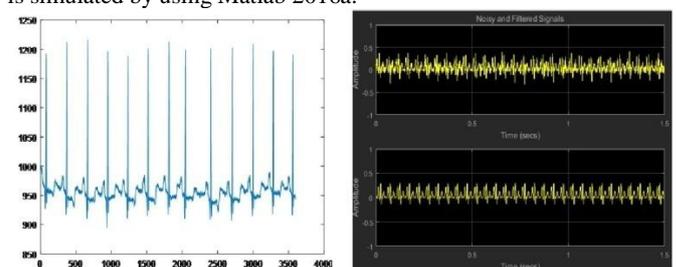
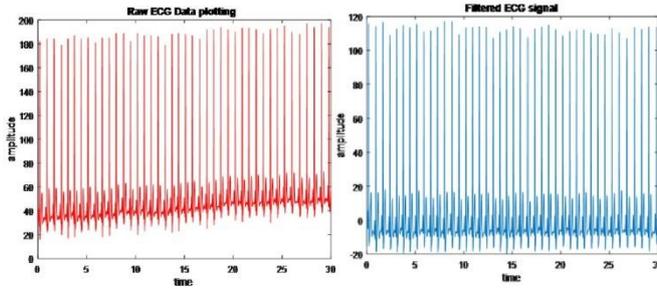**Figure 3:** Original ECG, noise signal for normal signal

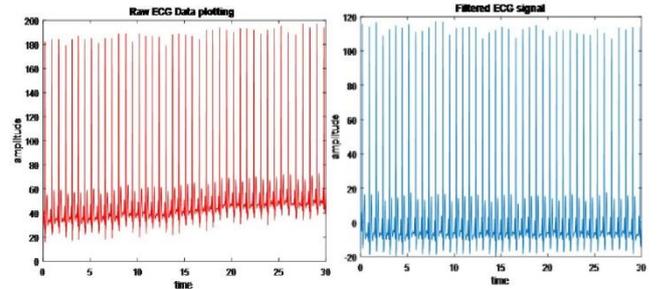**Figure 4:** Raw ECG data plotting, Filtered ECG signal for normal signal

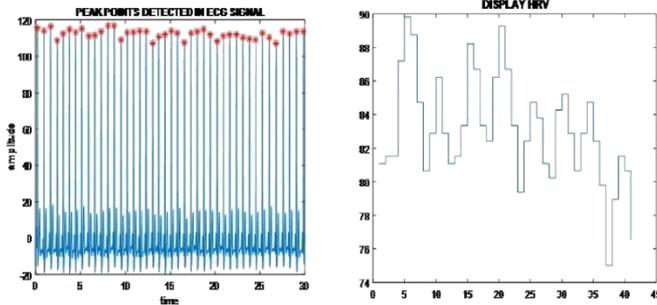**Figure 8:** Raw ECG data plotting, filtered ECG signal for abnormal signal

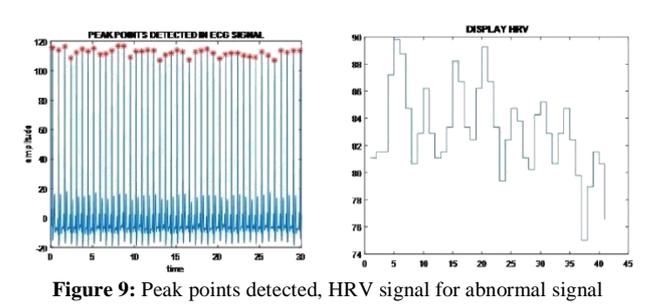**Figure 5:** Peak points detected, HRV display for normal signal

Figure 9: Peak points detected, HRV signal for abnormal signal

Figure 3 to 6 shows the ECG signal preprocessing and peak detection of the normal signal. The ECG signal is filtered by using low, high, and butter worth filter to remove the unwanted noise in the signal. Raw ECG is plotted for the signal and peak points are detected for the QRS waveform using peak detection algorithm. In this the peaks are detected for the signal and HRV (Heart Rate Variable) signal is displayed. After HRV signal is displayed, point care is detected for the signal and finally PRV (Pulse Rate Variable) is detected for the ECG signal. The same process is repeated from figure 7 to figure 10.
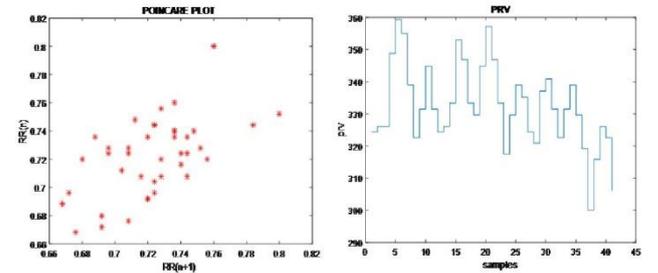
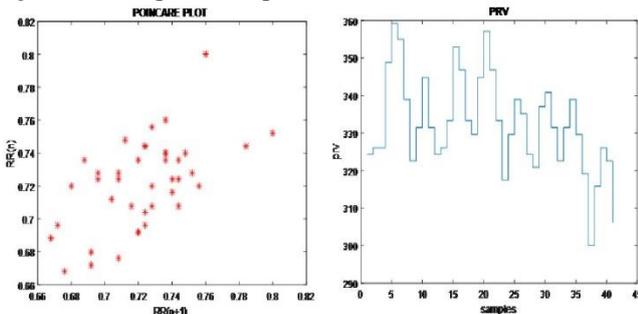**Figure 10:** Point care plot and PRV for the abnormal signal

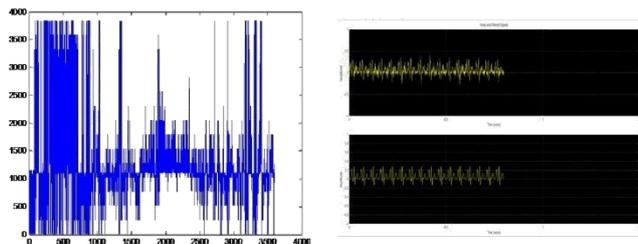**Figure 6:** Point care plot and PRV for the normal signal

Figure 11: Waveform of CWT and DWT

Figure 7: Original ECG, noise signal for abnormal signal
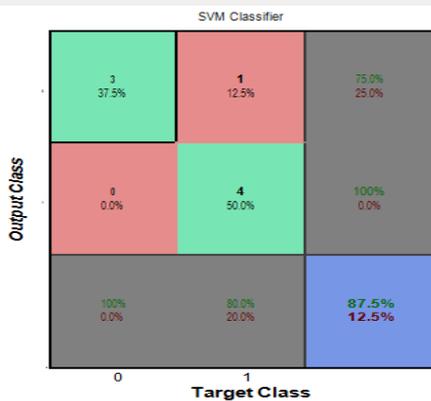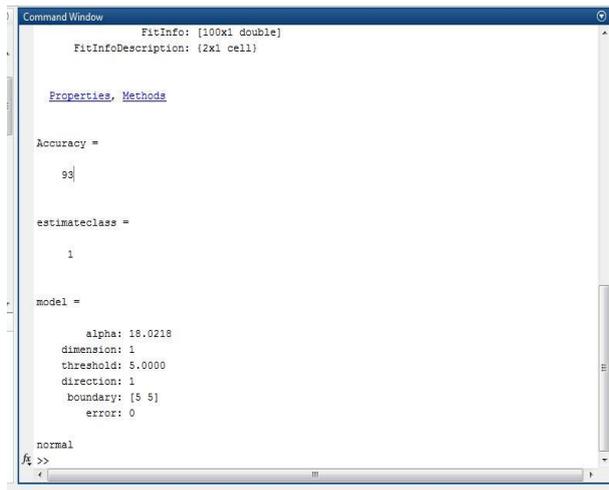
Figure 12: Confusion matrix of SVM classifier

Figure 14: Neural network classifier tool box

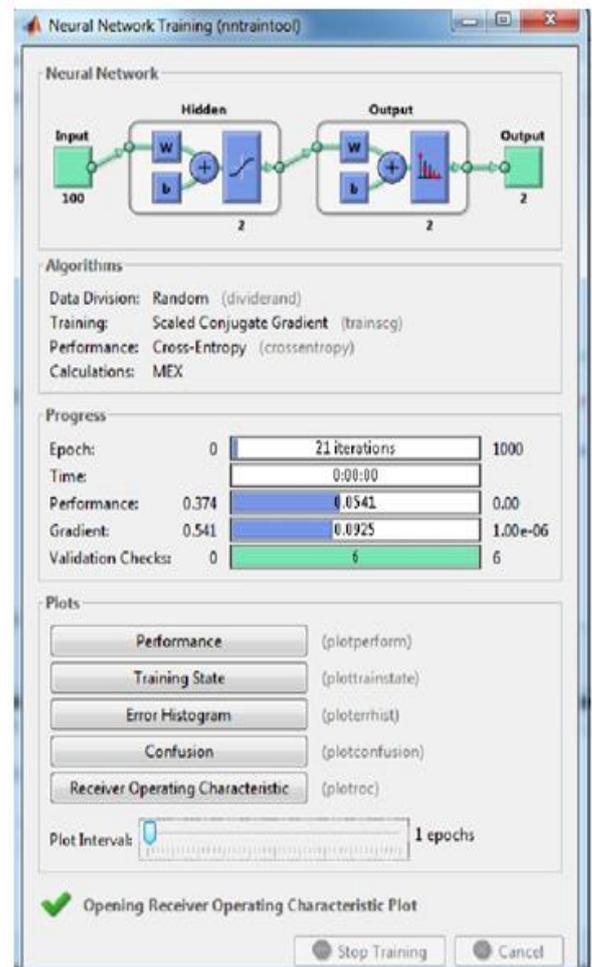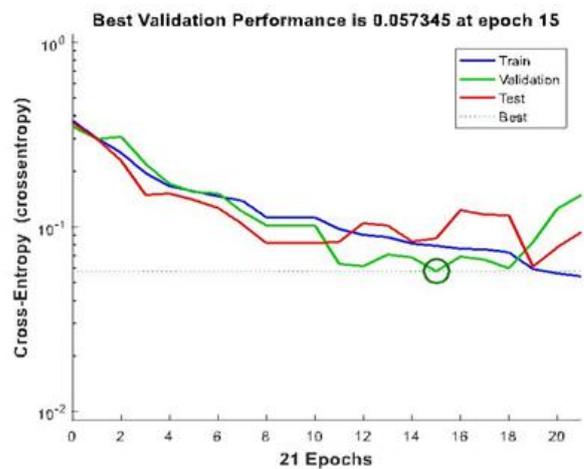**Figure 13:** Adaboost classifier

Figure 15: performance plot

Figure 11 shows the confusion matrix output of SVM classifier and figure 12 shows the Adaboost classifier output. To classify the ECG signal into normal or abnormal signal totally 16 ECG signal are taken from the database. The accuracy of the SVM classifier is 87.5% and the accuracy of the Adaboost classifier is 93%. Adaboost classifier is adding a SVM classifier with adaboost classifier is the final output of the adaboost classifier. Figure 13 shows the neural network classifier tool box and figure 14 to 18 shows the various characterics of the ANN classifier. The overall accuracy of the ANN classifier is 94% .Total number of iteration used for the process is 21.Figure 19 shows the naïve bayes classifier output and the accuracy of the classifier is 99.7%. Naïve bayes classifier shows the better accuracy compared to other classifier.
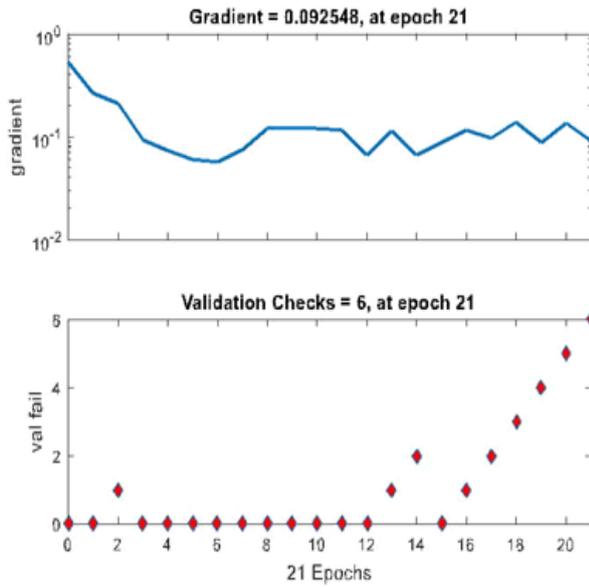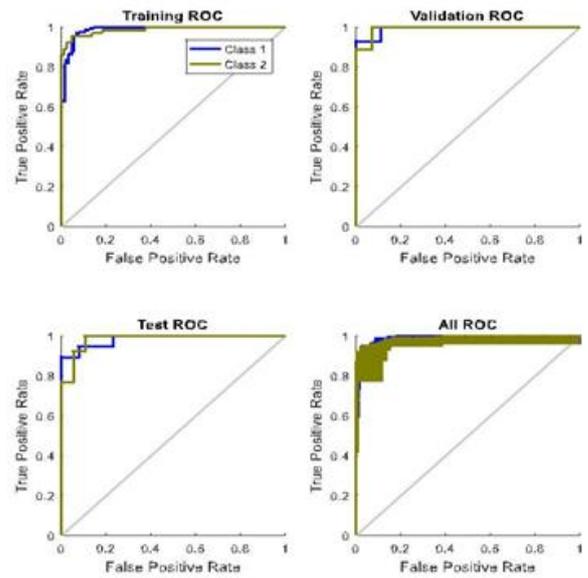
Figure 16: Training state plot

Figure17: Error histogram plot

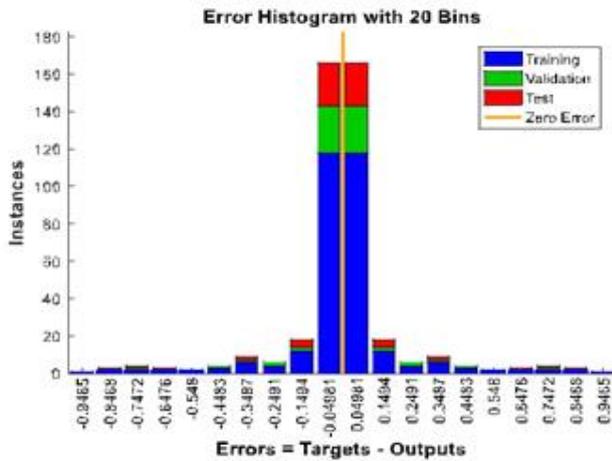**Figure 18:** Confusion matrix for ANN

Figure 19: Receiver Operating Charactertics Plot

Figure 20: Percentage of correct and incorrect classification

Figure 21: Naïve Bayes classifier

Totally four classifier is used to classify the ECG signal database such as ANN, SVM, Adaboost and Naïve bayes classifier. These classifiers classify the ECG signal into normal or abnormal ECG signal. The classifier accuracy and error values of all the classifier are compared. Table 2 shows the experimental comparison table for classifier accuracy and error values.[14]

Table 2: Experimental Comparison of Proposed work.

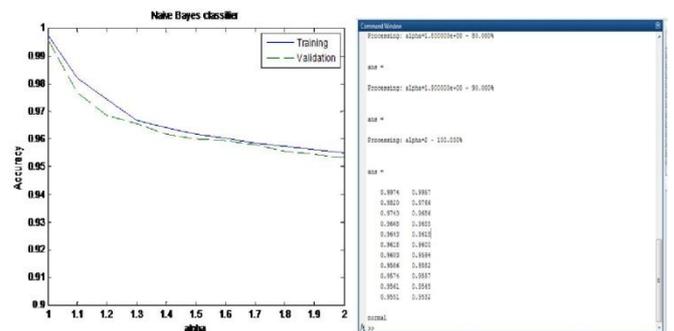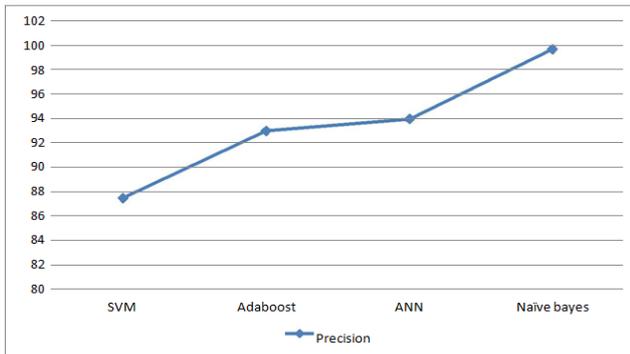| Classifier	Training %	Precision
SVM	100	87.5
Adaboost	100	93
ANN	100	94
Naive bayes	100	99.7

Figure 22: Comparison of classifier accuracy and error

Figure 20 shows the classifier accuracy and error comparison, the accuracy of the Naïve bayes classifier is high compared to other classifier. If the error value is low the accuracy is high.[13]

4. Conclusion

In this paper of various types classifier is used to classify the ECG signal into normal or abnormal ECG signal .The input image is preprocessed by various filtering technique to remove high frequency noise. Peak value is detected by peak detection algorithm and HRV signal is detected. Features are extracted for signal and classify the signal by ANN, SVM, Adaboost and Naive bayes classifier. From experimental results the accuracy of the SVM, Adaboost, ANN and Naïve Bayes classifier is 87.5% ,93%,94% and 99.7%.Naive Bayes classifier has the higher accuracy compared to other classifier.

References

[1] Sörnmo L & Laguna P, *Bioelectrical signal processing in cardiac and neurological applications*, Academic Press, (2005).

[2] De Chazal P, Celler BG & Reilly RB, "Using wavelet coefficients for the classification of the electrocardiogram", *IEEE 22nd Annual International Conference of the Engineering in Medicine and Biology Society,* (2000), pp.64-67.

[3] Cornelia G & Romulus R, "Ecg signals processing using wavelets", *University of Oradea: Electronics Department, Oradea, Romania*, (2005).

[4] Saxena SC, Kumar V & Hamde ST, "Feature extraction from ECG signals using wavelet transforms for disease diagnostics", *International Journal of Systems Science*, Vol.33, No.13, (2002), pp.1073-1085.

[5] Gautam R & Sharmar A, "Detection of QRS complexes of ECG recording based on wavelet transform using Matlab", *International Journal of Engineering Science and Technology*, Vol.2, No.7, (2010), pp.3038-3044.

[6] Akshay N, Jonnabhotla NAV, Sadam N & Yeddanapudi ND, "ECG noise removal and QRS complex detection using UWT", *IEEE International Conference On Electronics and Information Engineering (ICEIE)*, Vol.2, (2010), pp.V2-438.

[7] Espiritu-Santo-Rincon A & Carbajal-Fernandez C, "ECG feature extraction via waveform segmentation", *IEEE 7th International Conference on Electrical Engineering Computing Science and Automatic Control (CCE)*, (2010), pp.250-255.

[8] Barro S, Fernandez-Delgado M, Vila-Sobrino JA, Regueiro CV & Sanchez E, "Classifying multichannel ECG patterns with an adaptive neural network", *IEEE Engineering in Medicine and Biology Magazine*, Vol.17, No.1, (1998), pp.45-55.

[9] Abibullaev B & Seo HD, "A new QRS detection method using wavelets and artificial neural networks", *Journal of medical systems*, Vol.35, No.4, (2011), pp.683-691.

[10] Kannathal N, Acharya UR, Lim CM, Sadasivan PK & Krishnan SM, "Classification of cardiac patient states using artificial neural networks", *Experimental & Clinical Cardiology*, Vol.8, No.4, (2003), pp.206 –211.

[11] Van Alste JA & Schilder TS, "Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps", *IEEE Transactions on Biomedical Engineering*, Vol.12, (1985), pp.1052-1060.

[12] Hosseini HG, Reynolds KJ & Powers D, "A multi-stage neural network classifier for ECG events", *IEEE 23rd Annual International Conference of the Engineering in Medicine and Biology Society*, Vol.2, (2001), pp.1672-1675.

[13] Foo SY, Stuart G, Harvey B & Meyer-Baese, A, "Neural network-based EKG pattern recognition", *Engineering Applications of Artificial Intelligence*, Vol.15, No.3-4, (2002), pp.253-260.

[14] Maglaveras N, Stamkopoulos T, Diamantaras K, Pappas C & Strintzis M, "ECG pattern recognition and classification using non-linear transformations and neural networks: a review", *International journal of medical informatics*, Vol.52, No.1-3, (1998), pp.191-208.

[15] Sternickel K, "Automatic pattern recognition in ECG time series", *Computer methods and programs in biomedicine*, Vol.68, No.2, (2002), pp.109-115.

[16] Güler İ & Übeylı ED, "ECG beat classifier designed by combined neural network model", *Pattern recognition, Vol.38, No.2, (2005), pp.199-208.*

[17] Übeyli ED, "Combining recurrent neural networks with eigenvector methods for classification of ECG beats", *Digital Signal Processing, Vol.19, No.2, (2009), pp.320-329.*

[18] Korürek M & Doğan B, "ECG beat classification using particle swarm optimization and radial basis function neural network", *Expert systems with Applications, Vol.37, No.12, (2010), pp.7563-7569.*

[19] Ingole DT, Kulat K & Ingole MD, "Feature Extraction via Multi resolution Analysis for ECG Signal", *IEEE First International Conference on Emerging Trends in Engineering and Technology, (2008), pp.659-664.*

[20] Hao W & Luo J, "Generalized multiclass adaboost and its applications to multimedia classification", *IEEE Conference on Computer Vision and Pattern Recognition Workshop, (2006), pp.113-113.*

[21] Mark R & Moody G, *MIT-BIH Arrhythmia Database Directory, (2012).*

[22] Stallin S, Rajkumar P & Rajendran K, "Reduction of Noises in ECG Signal by Various Filters", *International Journal of Engineering Research & Technology, Vol.3, No.1, (2014).*

[23] Li T & Zhou M, "Ecg classification using wavelet packet entropy and random forests", *Entropy, Vol.18, No.8, (2016), pp.1-16.*

[24] Galar M, Fernandez A, Barrenechea E, Bustince H & Herrera F, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol.42, No.4, (2012), pp.463-484.*

[25] Oza NC & Tumer K, "Classifier ensembles: Select real-world applications", *Information Fusion, Vol.9, No.1, (2008), pp.4-20.*

[26] Hu YH, Palreddy S & Tompkins WJ, "A patient-adaptable ECG beat classifier using a mixture of experts approach", *IEEE transactions on biomedical engineering, Vol.44, No.9, (1997), pp.891-900.*

[27] De Chazal P, O'Dwyer M & Reilly RB, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", *IEEE transactions on biomedical engineering, Vol.51, No.7, (2004), pp.1196-1206.*

[28] Kutlu Y & Kuntalp D, "A multi-stage automatic arrhythmia recognition and classification system", *Computers in Biology and Medicine, Vol.41, No.1, (2011), pp.37-45.*

[29] Kutlu, Y & Kuntalp D, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients", *Computer methods and programs in biomedicine, Vol.105, No.3, (2012), pp.257-267.*

[30] Ye C, Kumar BV & Coimbra MT, "Heartbeat classification using morphological and dynamic features of ECG signals", *IEEE Transactions on Biomedical Engineering, Vol.59, No.10, (2012), pp.2930-2941.*

[31] Das MK & Ari, S, "ECG beats classification using mixture of features", *International scholarly research notices, (2014).*

[32] Afkhami RG, Azarnia G & Tinati MA, "Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals", *Pattern Recognition Letters, Vol.70, (2016), pp.45-51.*

[33] Z Iskakova, M Sarsembayev, Z Kakenova (2018). Can Central Asia be integrated as asean? Opción, Año 33. 152-169.

[34] A Mukanbetkaliyev, S Amandykova, Y Zhambayev, Z Duskaziyeva, A Alimbetova (2018). The aspects of legal regulation on staffing of procuratorial authorities of the Russian Federation and the Republic of Kazakhstan Opción, Año 33. 187-216.