

Data Mining Application for Health Seeker and Provider

Parul Berwal
Research Scholar
University Institute of
Engg and Technology
M.D University, Rohtak, India

Kamna Solanki
Assistant Professor
University Institute of
Engg and Technology
M.D University, Rohtak, India

ABSTRACT

Electro Cardio gram is the technique that is utilized to calculate the occurrence and consistency of heart beat. By distinguishing the overall ECG signal, doctors can easily predict that the signal is disposed to to heart attack or not. The signal processing is functioning by the computer based analysis which takes form of the alteration of the signal into another and this signal so generated is more desired than original. This research helps to identify the signal is prone to heart attack or not. This comprises the choice of some basic characteristic and comparing the neural networks outcomes with a hybrid method of ANN and FL (neural –fuzzy classifier). The outcomes so acquired after the effective comparison of each classifier states that ANFIS categorizes more perfectly than the Neural Networks.

Keywords

Electrocardiogram, Neural- Fuzzy, Neural Network, ANFIS

1. INTRODUCTION

Heart is the main part of human body. Electrocardiogram [1] is a tool utilized to determine ECG signal without entering the body. It is the electrical manifestation of the contractile movement of the heart that can be noted fast and itself. Electrodes are established on the human body skin to determine the bioelectric potentials allowable by the heart. To identify bioelectric signals given off by the heart one needs to organize the electrodes on the person's skin. Electrodes are even utilized to recognize the position of the chambers, the occurrence of any damage to the heart and information of cardiovascular condition utilized to regulate the heart beat.

An electronic system [2] assists the heart muscles to shrink and pump the blood via the heart to the lungs and the rest of the body. ECG transforms the electrical activity into the line tracings which are called waves such as P wave, QRS complex T wave and ST segment. The upper two chambers of the heart are atria and the other two are ventricles.

1. **P wave:** Atria's electrical activity.
2. **QRS complex:** It describes the motion of electrical impulses via ventricles.
3. **ST segment:** a linear, level line linking the QRS complex and T wave. It correlates to the line when ventricles contract but electricity is not produced.
4. **T wave:** correlates with the time when the lower heart chambers are resting electrically and make itself ready for the next muscle to shrink.

A pulse is discovered on the basis of the largest in the signal of PQRST of ECG signal. When PQRST and T wave's line appearances normal the heart beats are normal but if there are alterations in its line tracing then it means the heart beats are not normal. The wave pattern is obliging for the doctors to

identify the diseases utilizing ECG signal. To have the precise and fair recognition one should measure the waveform of the ECG signal properties. The characteristics enable us to recognize the physical procedure and the heart disease issue.

The following paper is organized as Section II describes the related work. Section III includes the research done and designing of the system. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

2. RELATED WORK

To categorize the heart diseases, various researchers have been working hard since a long period. The heart diseases have been classified by researchers with the help of the ECG which they are modifying day by day. One wants to elect a correct technique to boost the degree of exactness to recognize the heart disease. There are certain shortcomings in current ECG analysis in the previous research.

1. The researcher [3] introduced a new technique for the characteristic election and shape representative procedure known as Karhunen-Loeve Transform (KLT). The key characteristic to this method was that it could resolve the probability densities matter of population where the domains are not known, for the series of vectors of a problematic where domain are unknown but it was too sensitive for noise pattern of the ECG signal.
2. The author [4] presented a new method called wavelet transform which was attained by quadratic spline wavelet to discover myocardial ischemia. The technique has good method in nominal positive productivity value and high specificity. The computations required are quite more than other techniques due to alteration of the wavelet but was extended to resolve the other irregularities in the ECG signal.
3. The researcher [5] presented a multipurpose procedure to resolve 3 problems related to feature extraction that is time-frequency features, 2-time domain characteristics and 3-statistical characteristics. The rese arch deals with the complications and limitation in detecting characteristics of arrhythmia via other approaches. These 3 approaches are combined with PCA method to make sure that the redundancy issue can be minimized, thereby maximizing the accuracy level of classification.
4. The author [6] presented the paper which utilized the energy of distinct levels to restrain baseline wander from ECG signal using wavelet-transform based search algorithm. Schedules of electrodes that cause antique like noise induced by the ECG signal s are detached first. Then antique such as baseline

wanders that disturb exactness of data are detached. The present algorithm assists the individuals to remove the baseline drifts from the ECG signal that may source any obliteration to the signal and loss in clinical data or information.

- The author [7] introduced a technique for feature extraction that can extract extra spatial characteristics from the multi-channel ECG data recordings utilizing its independent module analysis and matching pursuits. The sensitivity, specificity and overall accuracy of the procedure is appreciable. There are 5 different classes of the heart diseases that are easily distinguishable via this algorithm. The only

shortcoming in the algorithm is that due to inverted T wave there is complicated segregation between ventricular PBs and PVCs.

- The researcher [8] proposed a method that is based on Hematite basis for the approximation of the ECG signal. Only the 5th order basis of Hermitian function was used. The approximation error of each segment is weighted according to the importance in the ECG complex. The role of this algorithm was that it showed that the total error obtained is half to the error shown by the other non-segmented methods. The drawback to this algorithm was that a small error in identifying can lead to the mislead diagnosis.

S.NO.	METHOD USED	RESEARCHERS INCLUDED AND YEAR	DESCRIPTION	PERFORMANCE AND ACCURACY
1	Karhunen Lo`eve Transform	(Jager 2002)	ECG analysis is the new methodology which assists to compute probability density for the sequence vectors in the domain. It is the methodology towards feature extortion and representation procedure of shape.	Accuracy observed is 80%
2	Wavelet Transform	(Thakur et al.,1993); (Li et al.1995) ; (Frau et al.2002); (Pretorius et al.2002); (Mahmoudabadi et al.2005)	It has also proved itself methodology for the time frequency analysis. It can restructure the decomposing signals and if these signals are decomposed into elementary building blocks. WT will be able to describe the local regularity of signals. It has may advantages and can preserve the energy during transformation.	Accuracy more than 90% Sensitivity attained over 90%
3	Support Vector machine	(Osowski et al.2004)	It is the reorganization system which utilizes distinct pre-processing systems for generalization of characteristics. They worked with the study that a reorganization system beats.	Accuracy attained over 80%
4	Fuzzy		Fuzzy has the countless efficiency in identifying the normal and distinct kinds of beats signifying Arrhythmias	Specificity more than 90% Sensitivity 60% to 70%
5	Fuzzy hybrid Neural network		It is simple and fast performing, good identifying and efficient technique for identifying distinct kinds of hardware.	

6	Bayesian		Bayesian can make decisions related to heart disease like doctor and it is a new tool for assessing the collected info. It comprises deep study of the info and studies our individual piece of information but it has some obstacles in telling the structure that produces the highest possible classification	Accuracy above 80%
7	Genetic algorithm		Genetic algorithms give 91% performance in sensitivity and specificity. But one needs to combine different classifiers to obtain more accuracy in results.	Specificity 90% Sensitivity more than 90%
8	Self-organizing map		In self organizing maps we get very less beat classification accuracy i.e. 76.51%.	Accuracy more than 70%

3. RESEARCH METHODOLOGY AND DESIGN

MATLAB software is used for analysis of Electro Cardiogram signals. In this chapter we will also discuss about the methods used to analyse ECG signals. The ANFIS were used in classifier tools in Fuzzy logic toolbox. It is also used in reading the raw data of ECG signal. The input ECG signal, are imported from the data files.dat and excel file.xls. Three stages in which ECG can be performed are:

- 1) Division of Pre-processing.
- 2) Feature Extraction using the various parameters.
- 3) Classification procedures using Adaptive Neuro Fuzzy Inference system trained with the back propagation gradient descent method in combination with least square analysis purpose will be explained.

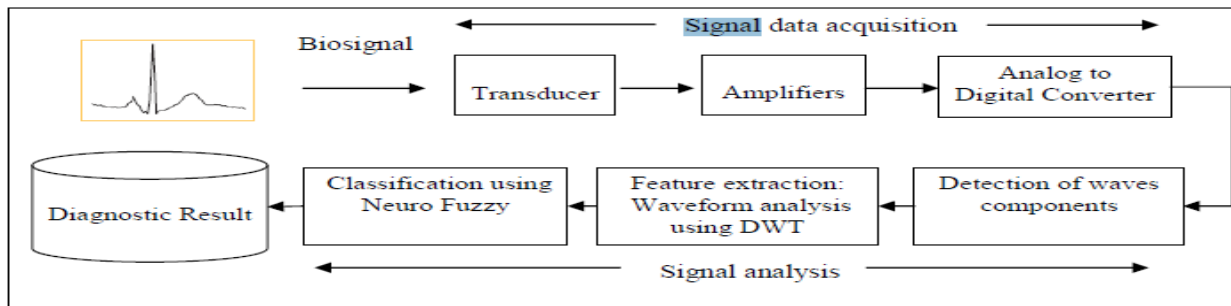


Figure 1: Overall Techniques for Electrocardiogram Analysis

For analysing and classifying ECG signal, methods are reviewed time to time and the new methods which provide accuracy are given below:

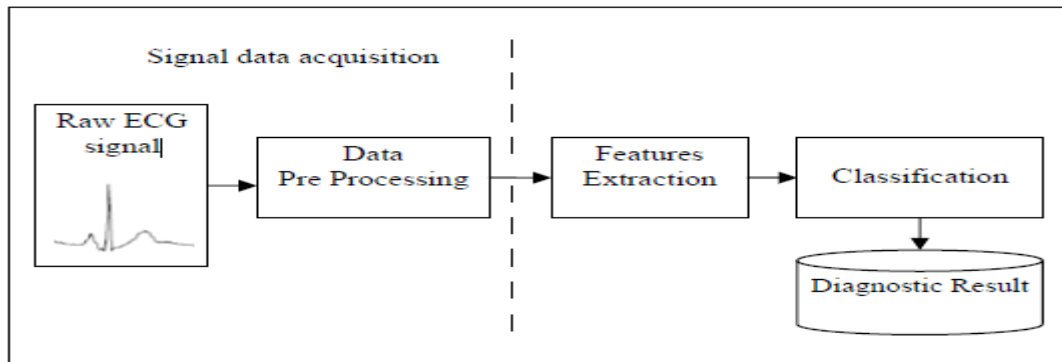


Figure 2: Electrocardiogram Analysis

3.1 Pre-processing

Pre-processing of Data is the first step in Data Mining technique. There are various uses or advantages of the Data Pre-

processing. We all know the data collected may be dirty or noisy. So pre-processing of data may be done. Other reasons of the pre-processing include:

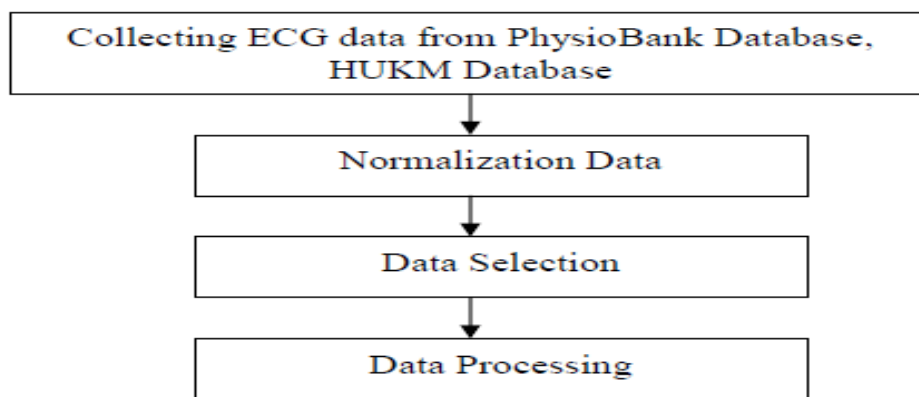


Figure 3: Pre-processing Step

3.2 Feature Extraction

Feature Extraction is the second step in data mining. This includes measuring of HRV. It is measurement of the changes in time interval between heart beats. It is also called RR variability where R is a point parallel to the range of the QRS complex of the ECG wave; and RR is the interval between consecutive Rs), and "heart period variability". To detect heart rate variations, we use TDM. Time domain methods (TDM) are depend on the beat to beat intervals and also called NN intervals which on analysis give the following parameters such as:

1. SDNN: (Standard deviation of NN intervals)- It is computed over the long periods ie,24 hour SDANN

which is computed over short periods usually taken 5 minutes.

2. RMSSD (*root mean square of successive differences*), that represents the square root of the mean of the squares of the differences between adjacent NNs.
3. SDSD (*standard deviation of successive differences*), the standard deviation of the consecutive differences between simultaneous NNs.
4. NN50, the number of pairs of successive NNs which has a difference by more than 50 ms.

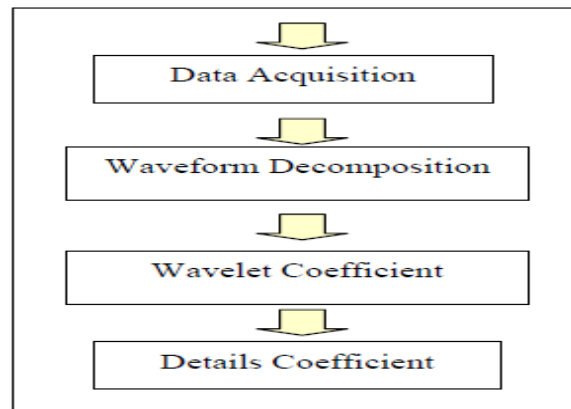


Figure 4: Feature Extraction Techniques

3.3 Classification using ANN

ANNs are very simple and basic electronic models based on the neural structural of human brain. The human brain commonly learns from its real life experiences. It has been proved that some problems that are not solvable by current computers can be easily solved by the small energy efficient packages. This kind of brain modeling promises a less technical way to evolve machine solutions for the problems. This novel approach towards computing provides a more graceful degradation at the time of system overload than its more traditional counter parts. This biologically inspired Neural network methods of computing are the future of the computing industry. Even simple functions which computer is not able to perform can be performed by any small living creature. Computer can do complex operations, computations very fast which are difficult for human brain to do whereas the task of recognizing the

pattern of pictures can be easily performed by human brain which are difficult for computer to do.

The advancement in the biological research has given us the good understanding of the thinking mechanism of the brain. These research shows that brain stores information in the form of patterns. Some of the prototypes which are very complex can be easily recognized by brain like recognizing the individual faces from different views. This work is on the basis of corresponding networks and guiding of those networks to perform specific tasks.

3.3.1 Artificial Neurons working

The main part of the NN is the Neuron. It comprises of the four components namely dendrites, axon, soma and synapses. The dendrites are used to take the inputs, combines them together, perform some operation on the input and gives out the result. Comparing with the neuron Artificial NN is built which is not so much complex as the simple neuron are but utilizes the functions used by the components of the simple biological neuron. This is just to compare the Human brain capabilities to solve the complex issues that are not computed by the previous computing methods. In ANN the inputs are considered to be taken in the form of $x(n)$ and then multiplied by the weight. Let the weight be represented by the $w(n)$. The various products

obtained are summed up and transferred to the functions to generate the result and then output.

3.3.2 Fuzzy Logic

It does not need of precise, numerical info. Input however they have the capability to control the high adaptively. It is the processing data which comes up by permitting the partial set membership in place of crisp set membership. It is successfully implemented and is adopted around the world by many. One can easily conclude by vague, noisy, imprecise or missing information. It has been accepted in Far East.

3.3.2.1 Fuzzy Sets

The collection of data is fuzzy sets in fuzzy logic based systems. In olden sets individual element can be included and can't be. Let's take an example of middle aged people. In traditional crisp set concept, people between 42 and 60 will be taken as member of set middle ages. Person will get the membership at 42. And at the age of 60. It is describing as humanistic mathematical modeling of the problem. A member from 35 -70 can be included in the set of middle age. Partial membership is used to assign to the members in the sets and then it is used to model this concept with fuzzy sets. It has uncountable truth values between one and zero and the proposals by the fuzzy logic are true. Membership sets of fuzzy can be inclusive, exclusive or there can be some degree in the fuzzy is different as it gives a chance to its elements to have membership. It returns a single degree of membership for any value in the domain so due to this we can say that the relationship is functional. Classification of Fuzzy logic is not illogical. It avoids taking the decisions that are sharp.

3.3.3 Classification using Neuro-Fuzzy

A hybrid of ANN and FL is Neuro-fuzzy. It uses techniques like Fuzzy sets and linguistic models which contains the set of if-then fuzzy rules that are used to integrate Neuro - fuzzy network to include the human like reasoning style of fuzzy systems. Distribution of data is first analyzed and then grouped into different membership functions. It is the most important part of fuzzy layer. The membership functions represent parameters of each heart beat and is applied as input vector to multilayer perception neural network classifier.

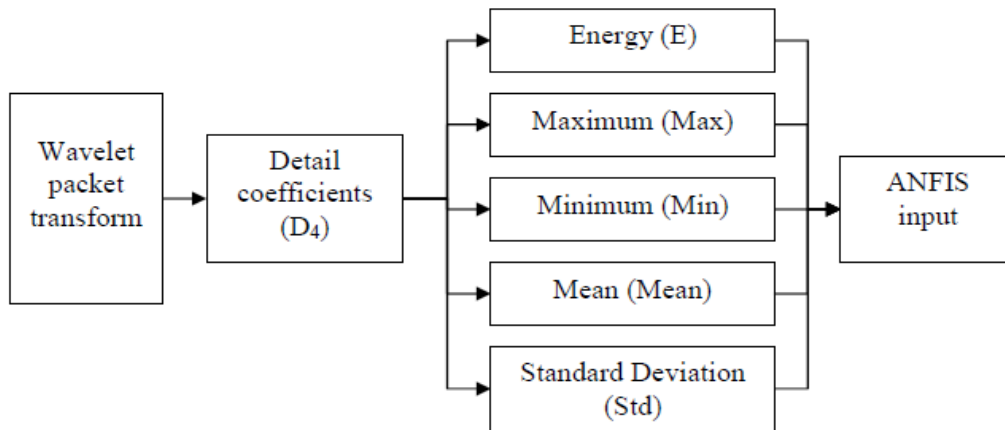


Figure 5: Analysis Framework

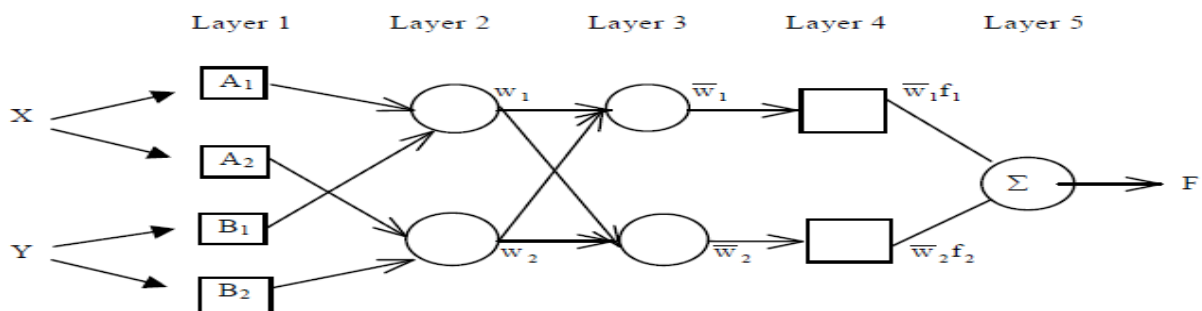


Figure 6: Working of ANFIS Model

3.4 ECG Analysis performances

The classification performance of the ANFIS model was measured to evaluate the performance of the classifiers. It was calculated based on the statistical parameters such as sensitivity, specificity and accuracy. The sensitivity, specificity and accuracy are defined as:

1. **Sensitivity** is a test to know how good the test is at detecting disease.

$$\text{Sensitivity} = \frac{\text{number of True Positives}}{\text{number of True Positives} + \text{number of False Negatives}}$$

2. **Specificity** is a test to know how good the test is at identifying normal.

$$\text{Specificity} = \frac{\text{number of True Negatives}}{\text{number of True Negatives} + \text{number of False Positives}}$$

3. **Total Classification Efficiency** is a measure of correct decisions cases/total numbers of cases.

4. EXPERIMENTAL RESULTS

Figures shows the results of Preprocessing of the ECG data, extracting the features of the ECG data, classification of ECG data using ANNs and the last is Classification using the Neuro-Fuzzy hybrid ANFIS tool. Figs. (a) shows the plotting of ECG in the form of graph (b) is the image obtained by filtered ECG signal (c) Various Feature Extracted.

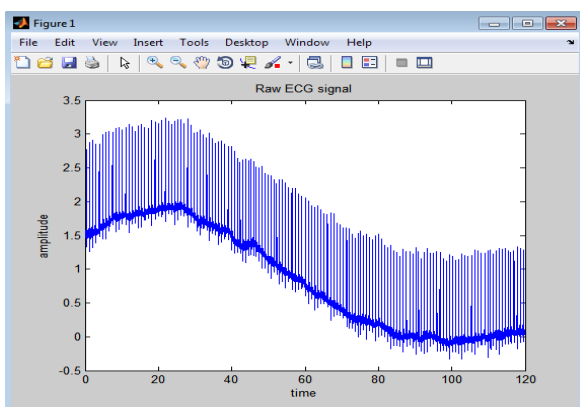


Figure 7: Plotting of ECG in the form of graph

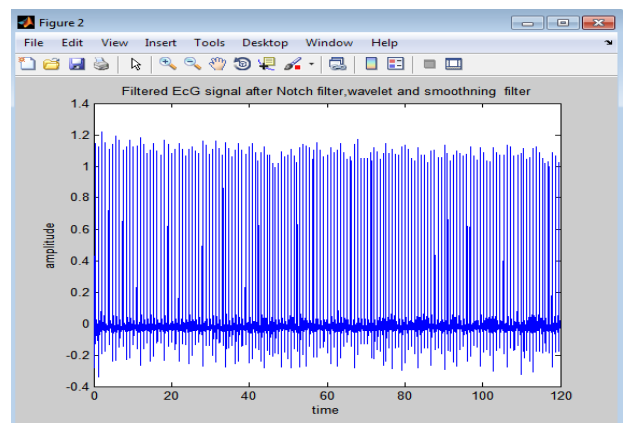


Figure 8: Image obtained by filtered ECG signal



Figure 9: Various Feature Extracted.

3.1 Classification using ANNs

In this Peak points are detected in ECG signal such that the values that can lead to heart attack are only taken. Peak signals in graph tell us the values that can lead to heart attack.

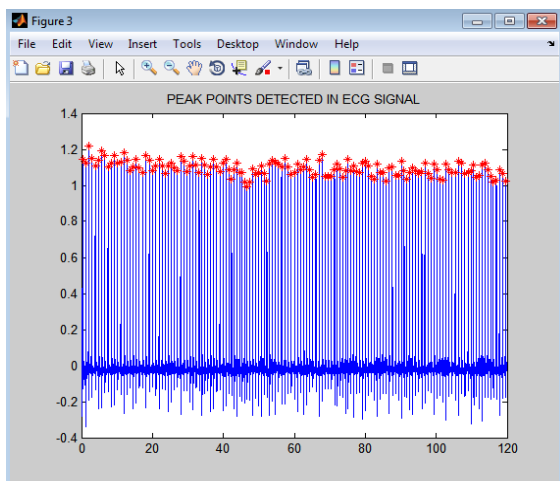


Figure 10(a): Peak Points detected in ECG signals

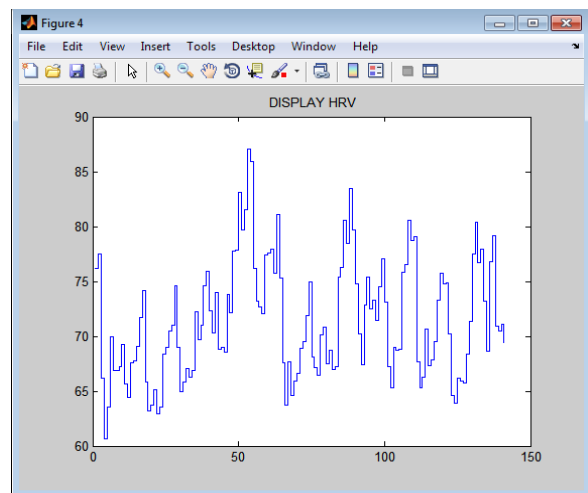


Figure 10(b): Heart Rate Variations

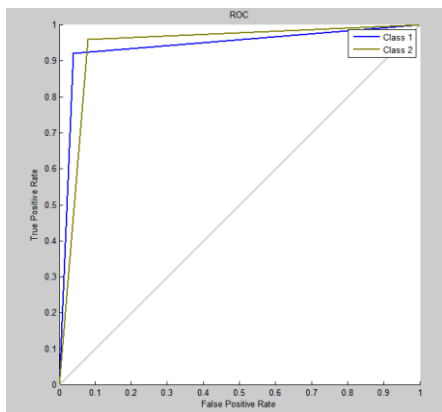


Figure 10(c): Receiver operating characteristic

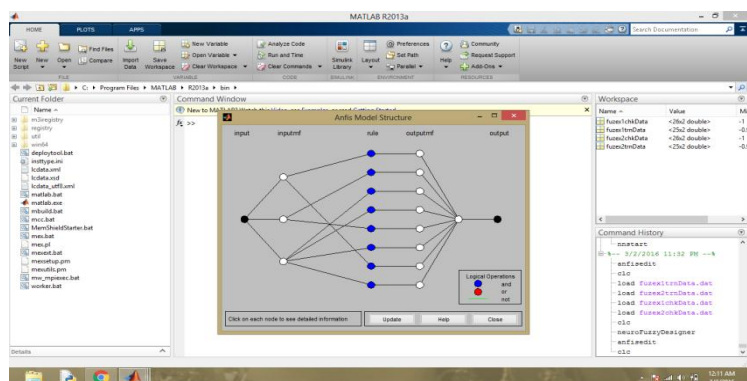


Figure 10(d): Classification Using ANFIS Toolcurve to detect performance

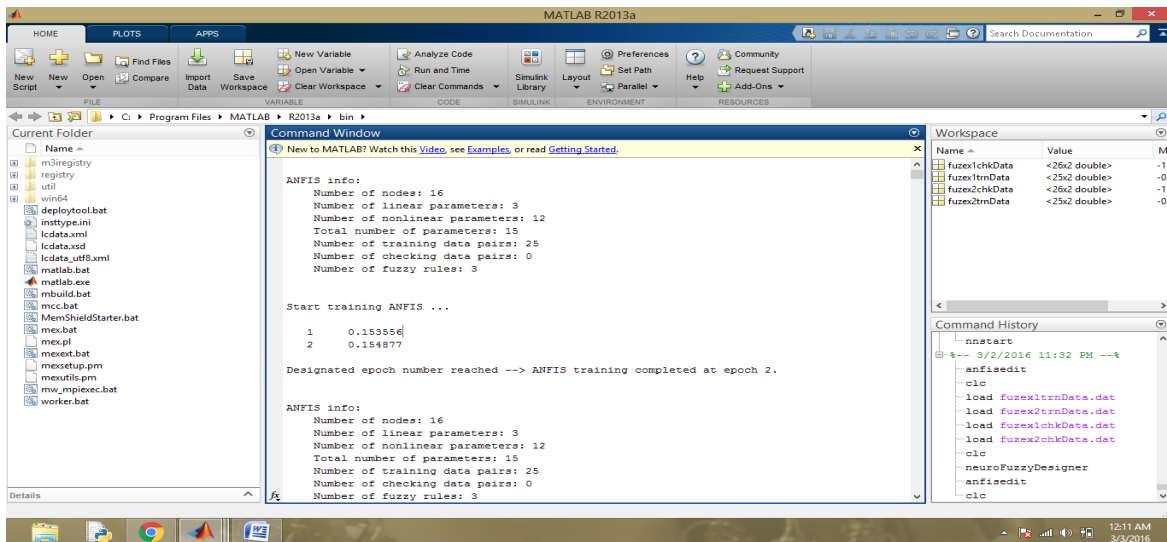


Figure 10(e): Number of nodes and training of data is done

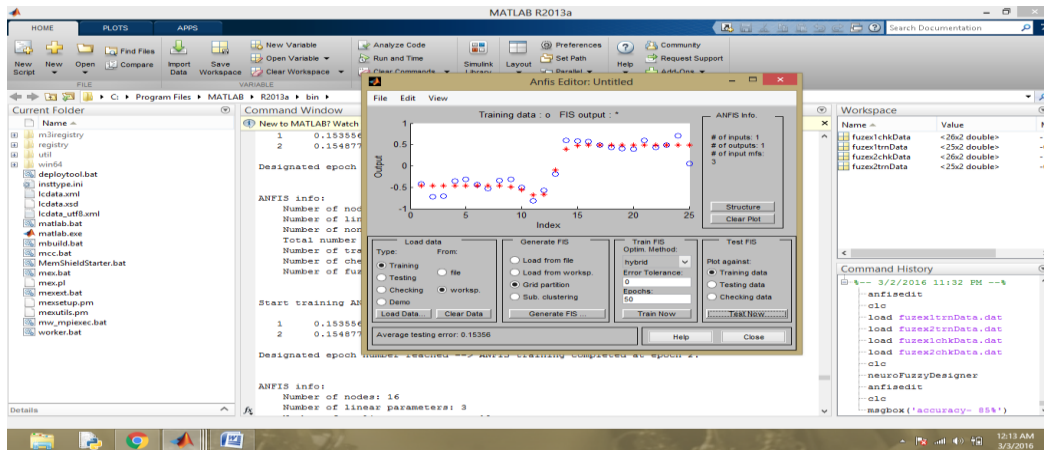


Figure 10(f): Training of data via ANFIS tool



Figure 10(g): Checking of data for the Classification

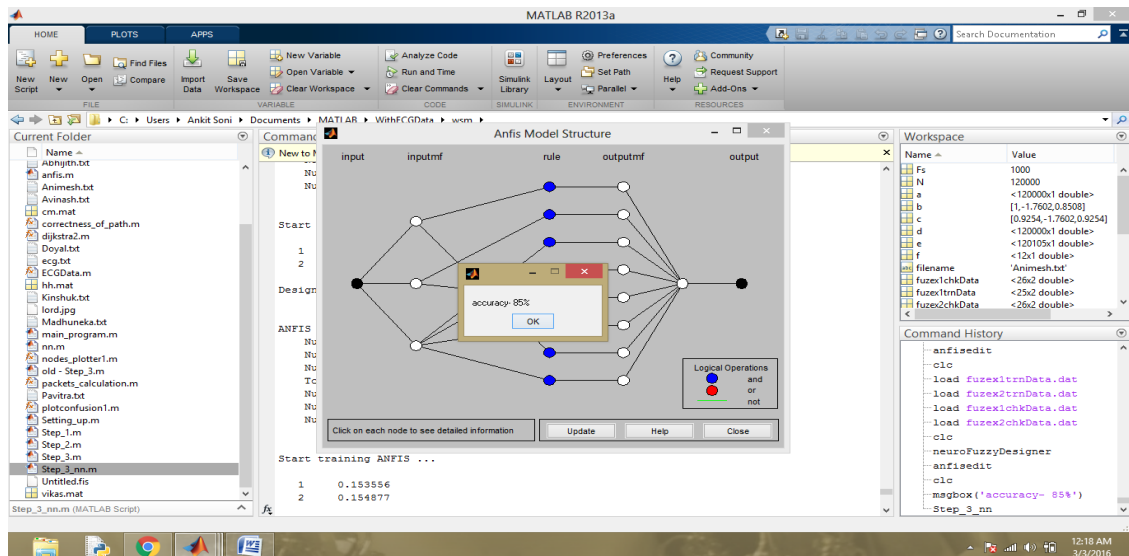


Figure 10(h): Performance by ANFIS model

It has been proved that the hybrid algorithm of Neural Networks and Fuzzy logic accuracy is more than the Algorithm of Neural networks.

5. CONCLUSION

The research has provided a solution utilizing the hybrid algorithms. A new way to know the accurate ECG classification scheme has been designed. It is a big achievement for medical science as we can identify the accurate heart disease. The whole analysis is divided into basic three stages. In first stage, based on the pre-processing stage analysis is subject to gather the database for patients from MITBH database. Then each of the element extracted from data in responsible to classify the ECG signal into category of heart attack or not. Then the data is processed at MATLAB software. Second stage comprises of extracting Features from the pre-processing stage which includes standard deviation, NN50, SSDN, Mean HRV and RMSSD. The third stage is applying the neural network algorithm to determine the data. By this, We can avoid the delay in treatments by updating information about patients if the technology is developed.

6. REFERENCES

- [1] Periyanga. J1 , Preethi. B2 , Priya. M3 , Ramakrishanan,' Instant Answering For Health Seekers Using Machine Learning', IJSRSET,2015.
- [2] http://www.emedicinehealth.com/electrocardiogram_ecg/article_em.htm.
- [3] Graja, S. and Boucher, J.M. 2005, 'Hidden Markov Tree Model Applied to ECG Delineation', IEEE Transactions on Instrumentation and Measurement, vol. 54, no. 6.
- [4] Jager F. 2002, 'Feature Extraction and Shape Representation of Ambulatory Electrocardiogram Using the Karhunen-Lo`eve Transform', Electrotechnical Review, Ljubljana, Slovenija.
- [5] Gao, D., Madden, M., Schukat, M., Chambers, D. and Lyons, G. 2004, 'Arrhythmia Identification from ECG Signals with a Neural Network Classifier Based on a Bayesian Framework', Department of Information Technology, National University of IrelandTinati, M.A. and Mozaffary B. 2006, 'A Wavelet Packets Approach to Electrocardiograph Baseline Drift Cancellation', International Journal of Biomedical Imaging, vol. 10, pp. 1-9.
- [6] Nobuo Ezaki, Marius Bulacu Lambert , Schomaker , "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", Proc. of 17th Int. Conf. on Pattern Recognition (ICPR), IEEE Computer Society, pp. 683-686, vol. II, 2004.
- [7] Herrero, G.G., Krasteva, V., Christov, I., Jekova, I., Gotchev, A. and Egjazarian, K. 2006, 'Comparative Study Of Morphological And Time-Frequency ECG Descriptors For Heartbeat Classification', Medical Engineering and Physics, vol. 28, issue 9, pp. 876-887.
- [8] Ahmadian, A., Karimifard, S. , Sadoughi, H. and Abdoli, M. 2007, 'An Efficient Piecewise Modeling of ECG Signals Based on Hermitian Basis Functions', Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, France, pp. 3180-3183.