www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671

A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level

Ashish Kumar Sen[1], Shamsher Bahadur Patel[2], Dr. D. P. Shukla[3]

[1]Faculty of Math/Computer Science,
Govt. PG Science College Rewa (M.P.)-India
ashishsen1983@gmail.com[1]

[2]Faculty of Math/Computer Science,
Govt. PG Science College Rewa (M.P.)-India
sstpatel12@gmail.com[2]

[3]Head of Department Math/ Computer Science
Govt. PG Science College Rewa (M.P.)-India
shukladpmp@gmail.com[2]

Abstract- ***Cardiovascular disease remains the biggest cause of deaths worldwide and the Heart Disease Prediction at the early stage is importance. As large amount of data is generated in medical organizations (hospitals, medical centers) but as this data is not properly used. There is a wealth of hidden information present in the datasets. This unused data can be converted into useful data. For this purpose we can use different data mining techniques. In this paper, we have defined a two layered approach for identifying the disease possibility. The critical factors that are mandatory for occurrence of coronary heart disease are taken at first level and the rest one are taken at second level. This two level approach increases the performance of our work as it helps in predicting disease chances accurately. The heart disease dataset is taken from UCI machine learning repository to train the neural network and then fuzzy rules are applied to predict the chances of coronary heart disease as low, medium or critical.***

Keywords- Data mining, Neuro-fuzzy, Heart disease and Coronary Heart Disease, Back-propagation Algorithm.

1.0 Introduction

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. With the ever-growing complexity in recent years, huge amounts of information in the area of medicine have been saved every day in different electronic forms such as Electronic Health Records (EHRs) and registers. These data are collected and used for different purposes. Data stored in registers are used mainly for monitoring and analyzing health and social conditions in the population. The unique personal identification number of every inhabitant enables linkage of exposure and outcome data spanning several decades and obtained from different sources. The existence of accurate epidemiological registers a basic prerequisite for monitoring and analyzing health and social conditions in the population. Some registers are state-wide, cover the whole collieries population, and have been collecting data for decades. They are frequently used for research, evaluation, planning and other purposes by a variety of users in terms of analyzing and predicting the health status of individuals. Regardless of the research activities to prevent Coronary Heart Disease (CHD), it vestiges one of the most imperative reason of death in India. Coronary Heart Disease (CHD) is

frequently deadly and most of the people, who die because of it, have experienced different symptoms that were not taken into account. The facts illustrate that, each year, almost 100000 people die since of Coronary Heart Disease.

Description focuses on finding human-interpretable patterns describing the data. Data Mining has potential applications in several fields, not the least of which is Health Care. The primary purpose of the Data Mining is to help determine trends in patient records to improve Health Care. Healthcare industry today generates large amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc. The large amount of data is a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining brings a set of tools and techniques that can be applied to this processed data to discover hidden patterns that provide healthcare professionals an additional source of knowledge for making decisions. The decisions rest with health care professionals.

Neuro-Fuzzy [12] refers to combinations of artificial neural networks and fuzzy logic. Neuro-fuzzy was proposed by J. S. R. Jang. Neuro-fuzzy hybridization results in a hybrid intelligent system that synergizes these two techniques by combining the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks. In the present work, we have designed an automated tool which will analyze the chances of occurrence of coronary heart disease in a patient. For this analysis, we have proposed a layered neuro-fuzzy based approach. The present system is simple one that can accept the input in terms of patient's basic information and medical tests. The heart disease dataset taken from UCI machine learning repository [14] is used to train the proposed network. Then this training data is used to test the output with unseen data and error is obtained. 80% of data is used to train the network and 20% of data is used to test the accuracy of neural network.

The main objective of the work is to define a new, two-layered coronary heart disease prediction system. The work will depend on the multi parameters. It has been tried to minimize the error in order to obtain an accurate Neuro-Fuzzy integrated system which can correctly predict the chances of coronary heart disease (CHD) in a person. Coronary Heart Disease [12] (CHD) occurs when plaque builds up inside the coronary arteries. These arteries supply heart muscle with oxygen-rich blood. Plaque is made up of fat, cholesterol, calcium, and other substances found in the blood. Over time, plaque hardens and narrows the arteries, reducing blood flow to your heart muscle. Eventually, an area of plaque can rupture, causing a blood clot to form on the surface of the plaque. If the clot becomes large enough, it can mostly or completely block the flow of oxygen-rich blood to the part of the heart muscle fed by the artery. This can lead to angina or a heart attack. In order to facilitate systematic understanding of the proposed work, section II discusses related work, section III discusses present work and experimental results, and finally, section V presents conclusion and scope for future work.

The two "high-level" [16] primary goals of data mining are *prediction* and *description*. **Prediction** involves using some variables or fields in the database to predict unknown or future values of other variables of interest.

2.0 Literature Survey

In year 2000, Shusaku Tsumoto [1] performed a work,**"Problems with Mining Medical Data".** Thus, it is highly expected that data mining methods will find interesting patterns from databases as reuse of stored data and be important for medical research and practice because human beings cannot deal with such a huge amount of data. In this paper, we focus on the characteristics of medical data and discuss how data miners deal with medical data.

In year 2004, Y. Alp Aslandogan et. al. [2] performed a work,**"Evidence Combination in Medical Data Mining".** We combine the beliefs of three classifiers: k-Nearest Neighbor (kNN), Naïve Bayesian and Decision Tree. Dempster's rule of combination combines three beliefs to arrive at one final decision. Our experiments with k-fold cross validation show that the nature of the data set has a bigger impact on some classifiers than others and the classification based on combined belief shows better overall accuracy than any individual classifier. We compare the performance of Dempster's combination (with differentiation-based uncertainty assignment) with those of performance-based linear and majority vote combination models. We study the circumstances under which the evidence combination approach improves classification.

In year 2007, Constantinos Koutsojannis et. al. [3] performed a work, "**Using a Neurofuzzy Approach in Medical Application"**. Today hybrid computing is a popular framework for solving complex problems such as in medical domain. Hybrid intelligent systems are systems that combine two or more intelligent techniques. Medicine and health care are closely related domains, where the types of problems faced are suitable for application of hybrid intelligent techniques. In this paper we present the initial evaluation of Fuzzy NEUrule System which is a Neuro Fuzzy approach based on fuzzy Adaline neurons and uses Differential Evolution for optimization of membership functions. According to our previous Neuro-fuzzy approaches and a well-defined hybrid system HYMES, FUNEUS is an attempt to the direction for integration of neural and fuzzy components with Differential evolution. Despite the fact that it remains difficult to compare neurofuzzy systems conceptually and evaluate their performance, early experimental results in a medical database proved a promising performance and the need for further evaluation in other medical applications.

In year 2008, Hai Wang et. al. [4] performed a work, **"Medical Knowledge Acquisition through Data Mining".** Data mining has been widely considered as an effective tool for knowledge discovery. This paper discusses the important role of medical experts for medical data mining, and presents a model of medical knowledge acquisition through data mining.

In year 2008, Markos G. Tsipouras et. al. [5] performed a work **"Automated Diagnosis of Coronary Artery Disease**

Based on Data Mining and Fuzzy Modeling". A fuzzy rule-based decision support system (DSS) is presented for the diagnosis of coronary artery disease (CAD). The system is automatically generated from an initial annotated dataset, using a four stage methodology: 1) induction of a decision tree from the data; 2) extraction of a set of rules from the decision tree, in disjunctive normal form and formulation of a crisp model; 3) transformation of the crisp set of rules into a fuzzy model; and 4) optimization of the parameters of the fuzzy model. The dataset used for the DSS generation and evaluation consists of 199 subjects, each one characterized by 19 features, including demographic and history data, as well as laboratory examinations. Tenfold cross validation is employed, and the average sensitivity and specificity obtained is 62% and 54%, respectively, using the set of rules extracted from the decision tree (first and second stages), while the average sensitivity and specificity increase to 80% and 65%, respectively, when the fuzzification and optimization stages are used. The system offers several advantages since it is automatically generated, it provides CAD diagnosis based on easily and noninvasively acquired features, and is able to provide interpretation for the decisions made.

In Year 2010, Harsh Vazirani et. al. [6] performed a work**," Use of Modular Neural Network for Heart Disease"**. In this paper Author mostly concern about the diagnosis of the heart disease. Mainly two types of the diagnosis method are used one is manual and other is automatic diagnosis which consists of diagnosis of disease with the help of intelligent expert system. In this paper the modular neural network is used to diagnosis the heart disease. The attributes are divided and given to the two neural network models Backpropagation Neural Network (BPNN) and Radial Basis Function Neural Network (RBFNN) for training and testing. In year 2010, Ali.Adeli et. al. [7] performed a work, **"A Fuzzy Expert System for Heart Disease Diagnosis"**. The aim of this study is to design a Fuzzy Expert System for heart disease diagnosis. The designed system based on the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation data base. The system has 13 input fields and one output field. Input fields are chest pain type, blood pressure, cholesterol, resting blood sugar, maximum heart rate, resting electrocardiography (ECG), exercise, old peak (ST depression induced by exercise relative to rest), thallium scan, sex and age.

In Year 2011, A.Q. Ansari et. al. [8] performed a work**, "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System"**. Computational intelligence combines fuzzy systems, neural network and evolutionary computing. In this paper, Neurofuzzy integrated system for coronary heart disease is presented. In order to show the effectiveness of the proposed system, Simulation for automated diagnosis is performed by using the realistic causes of coronary heart disease. The results suggest that this kind of hybrid system is suitable for the identification of patients with high/low cardiac risk.

In year 2012, Nidhi Bhatla et. al. [9] performed a work, **"A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic"**. Cardiovascular disease is a term used to describe a variety of heart diseases, illnesses, and events that impact the heart and circulatory system. A clinician uses several sources of data and tests to make a diagnostic impression but it is not necessary that all the tests are useful for the diagnosis of a heart disease. The objective of our work is to reduce the number of attributes used in heart disease diagnosis that will automatically reduce the number of tests which are required to be taken by a patient. Our work also aims at increasing the efficiency of the proposed system. The observations illustrated that Decision Tree and Naive Bayes using fuzzy logic has outplayed over other data mining techniques.

In year 2012, Chaitrali S. Dangare et. al. [10] performed a work, **"Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques"**. The Healthcare industry is generally "information rich", but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making. Advanced data mining techniques are used to discover knowledge in database and for medical research, particularly in Heart disease prediction. This paper has analyzed prediction systems for Heart disease using more number of input attributes. The system uses medical terms such as sex, blood pressure, cholesterol like 13 attributes to predict the likelihood of patient getting a Heart disease. Until now, 13 attributes are used for prediction. This research paper added two more attributes i.e. obesity and smoking. The data mining classification techniques, namely Decision Trees, Naive Bayes, and Neural Networks are analyzed on Heartdisease database. The performance of these techniques is compared, based on accuracy. As per our results accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively. Our analysis shows that out of these three classification models Neural Networks predicts Heart disease with highest accuracy.

In year 2013, S. Vijiyarani et. al. [11] performed a work, **"An Efficient Classification Tree Technique for Heart Disease Prediction"**. The data mining can be defined as discovery of relationships in large databases automatically and in some cases it is used for predicting relationships based on the results discovered. Data mining plays a vital role in various applications such as business organizations, e-commerce, health care industry, scientific and engineering. In the health care industry, the data mining is mainly used for predicting the diseases from the datasets. Various data mining techniques are available for predicting diseases namely Classification, Clustering, Association rules and Regressions. This paper analyzes the classification tree techniques in data mining. The aim of this paper is to investigate the experimental results of the performance of different classification techniques for a heart disease dataset.

The classification tree algorithms used and tested in this work are Decision Stump, Random Forest, and LMT Tree algorithm. Comparative analysis is done by using Waikato Environment for Knowledge Analysis or in short, WEKA. It is open source software which consists of a collection of machine learning algorithms for data mining tasks.

3.0 Present Work

In the present work, we have designed an automated tool which will analyze the chances of occurrence of coronary

heart disease in a patient. For this analysis, we have proposed a layered neuro-fuzzy based approach. The present system is simple one that can accept the input in terms of patient's basic information and medical tests. The training set is defined by all the parameters related to coronary heart disease *viz.* Age, Sex, Blood Sugar, Cholesterol, Blood Pressure, Max Heart rate, Exercise induced angina, ECG and
Chest pain type. The present system will generate fuzzy based rules under multiple parameters that are divided in two layers. The system generates a fuzzy based database and the query engine to process on database. The system processes on this database and return the chances occurrence of heart disease graphically.

3.1 Coronary Heart Disease

Coronary Heart Disease [12] (CHD) occurs when plaque builds up inside the coronary arteries. These arteries supply heart muscle with oxygen-rich blood. Plaque is made up of fat, cholesterol, calcium, and other substances found in the blood. Over time, plaque hardens and narrows the arteries, reducing blood flow to your heart muscle. Eventually, an area of plaque can rupture, causing a blood clot to form on the surface of the plaque. If the clot becomes large enough, it can mostly or completely block the flow of oxygen-rich blood to the part of the heart muscle fed by the artery. This can lead to angina or a heart attack. Angina is chest pain or discomfort that occurs when not enough oxygen-rich blood is flowing to an area of heart muscle. Angina may feel like pressure or squeezing in the chest. The pain also may occur in your shoulders, arms, neck, jaw, or back. A heart attack occurs when blood flow to an area of your heart muscle is completely blocked. This prevents oxygen-rich blood from reaching that area of heart muscle, causing it to die. Without quick treatment, a heart attack can lead to serious problems or death. Figure 1 shows the information of Plaque inside coronary arteries.

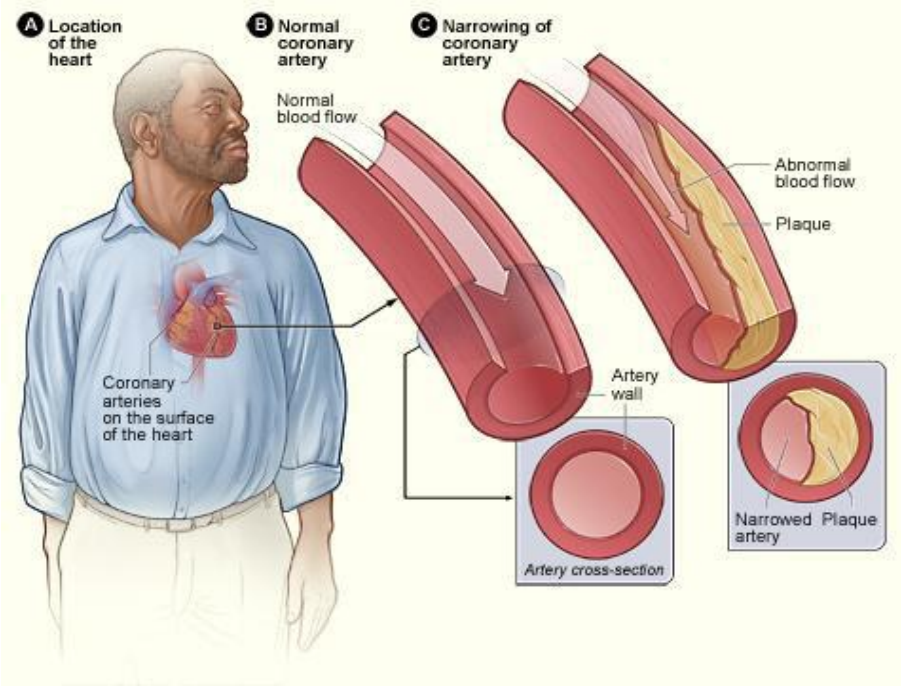

Fig. 1 a shows location of heart, B shows a normal artery with normal blood flow and C shows an artery with plaque build up [14].

3.2 Dataset

Data used for present work is obtained from UCI (University of California, Irvine C.A) Centre for machine
learning and intelligent systems. The data have been collected from 303 patients are used for proposed work. This database contains 76 attributes, 13 of them are available for experiments but we have used only 9 of them which are required for coronary heart disease. The attributes that we have considered in this proposed work are:

1) Age
2) Sex
3) Blood Pressure
4) Cholesterol
5) Blood Sugar
6) Electrocardiogram (ECG)
7) Maximum Heart Rate
8) Chest Pain Type
9) Exercise induced angina

Out of these 9 attributes, only 4 are considered as critical parameters for coronary heart disease. These are: Age, Cholesterol, Blood Pressure and Maximum heart rate.

3.3 Objective of Present Work

1) The main objective of the work is to define new two-layered coronary heart disease prediction system. The work will depend on the multi parameters.

2) A neuro-fuzzy approach will be implemented on two layers.
3) Each layer consists of different parameters with first layer having critical ones and second layer having the rest.
4) The work is about to identify the disease chances accurately.

3.4 Flowchart of Present Work

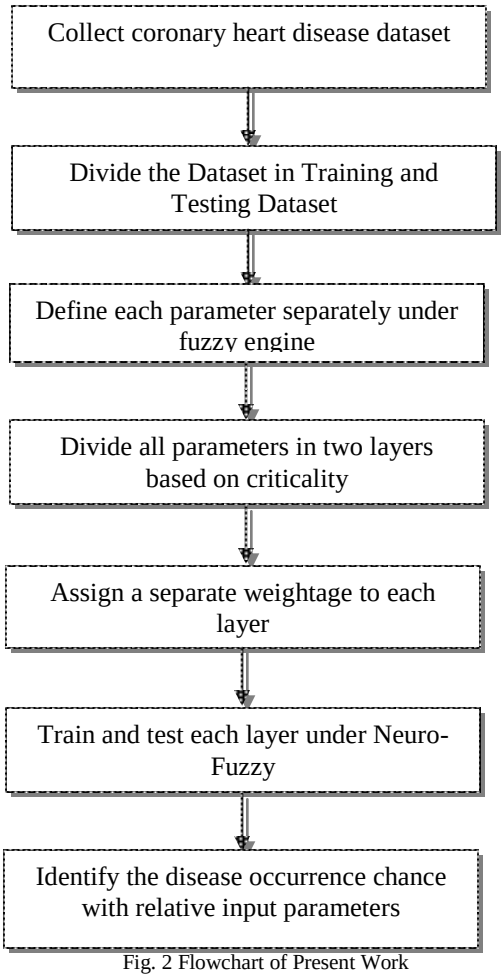

Fig. 2 Flowchart of Present Work

4.0 Results

4.1 Simulation Tool

1) MATLAB Editor is used for writing the code to implement our algorithm.
2) The result will be shown in the command window of MATLAB.

4.2 Training Input Data

In this figure, the input to the current system is perform is performed called the training input. The training data is taken under different medical parameters such as age, cholesterol, blood pressure etc. The available UCI dataset is divided in two parts called training dataset and the testing dataset. We have selected 80% data values as the training dataset and 20% dataset as the testing dataset. Backpropagation algorithm is used to train artificial neural network. In the figure 3, the patient analysis based on training dataset is defined. The first image in figure 1 is showing the increment respective to the training dataset in on different medical parameters and the second image is showing the error effectiveness.

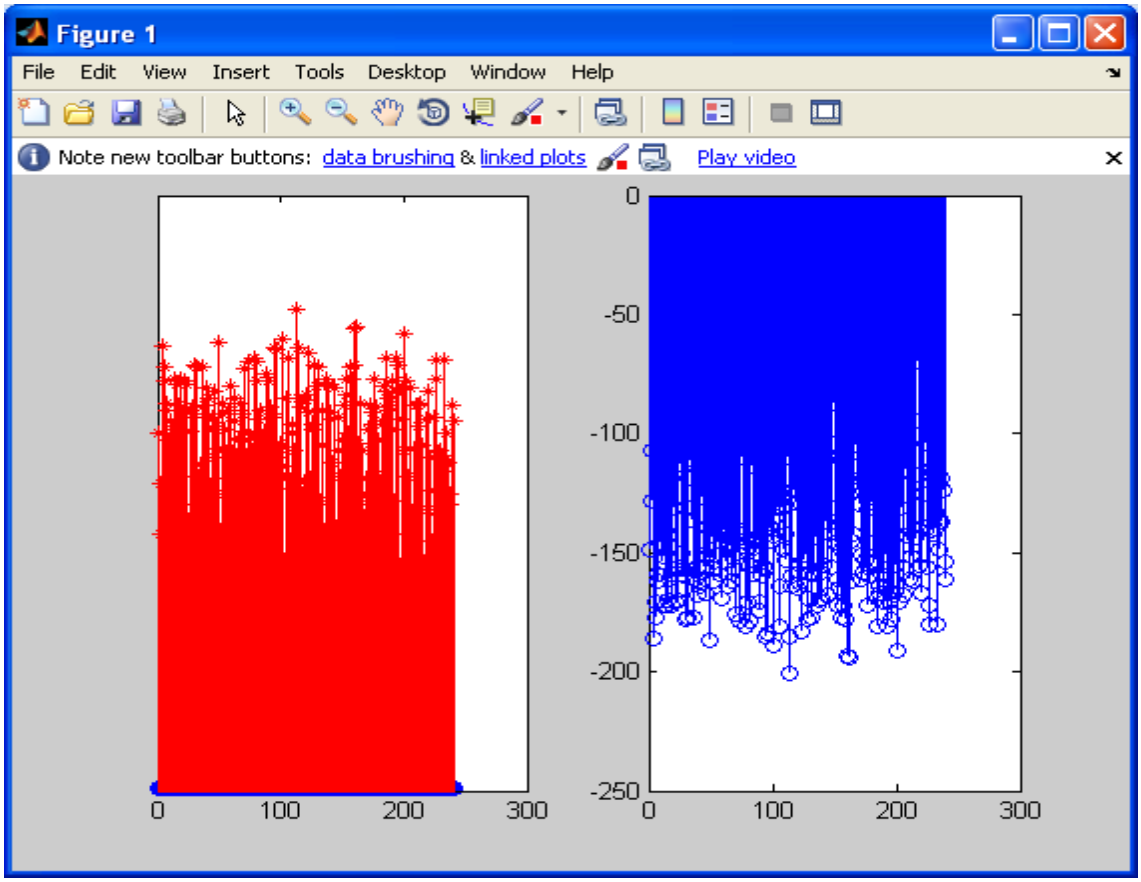

Fig. 3 Training input data

4.3 Predictive Result Driven

Figure 4 is showing the predictive results driven from the model. As we can see, the model has shown the incremental change in the patient information under different parameters.

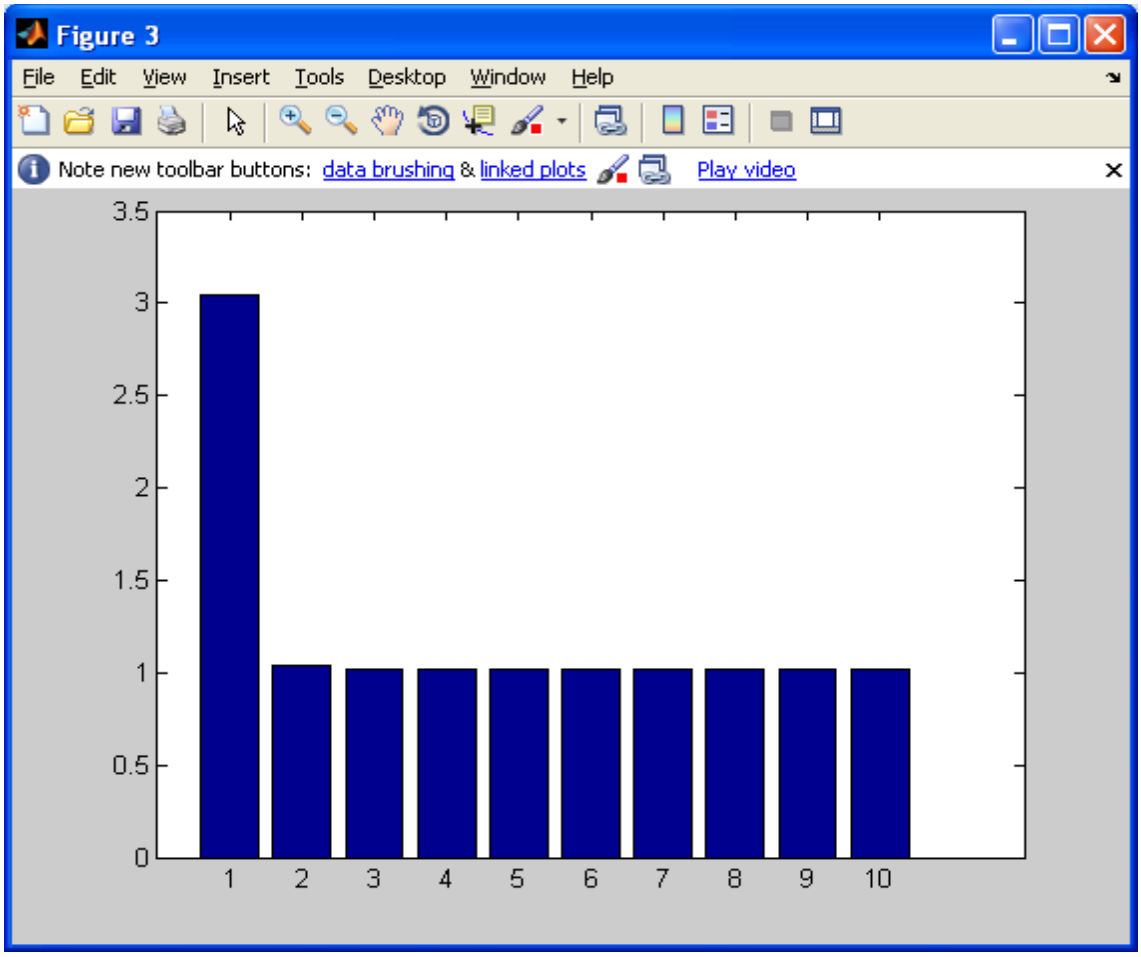

Fig. 4 Predictive Result Driven

4.5 Performance Plot

Figure 5 shows performance plot of the present system. The mean square error is very low i.e ≈ .04. So, the performance of our work is very high.

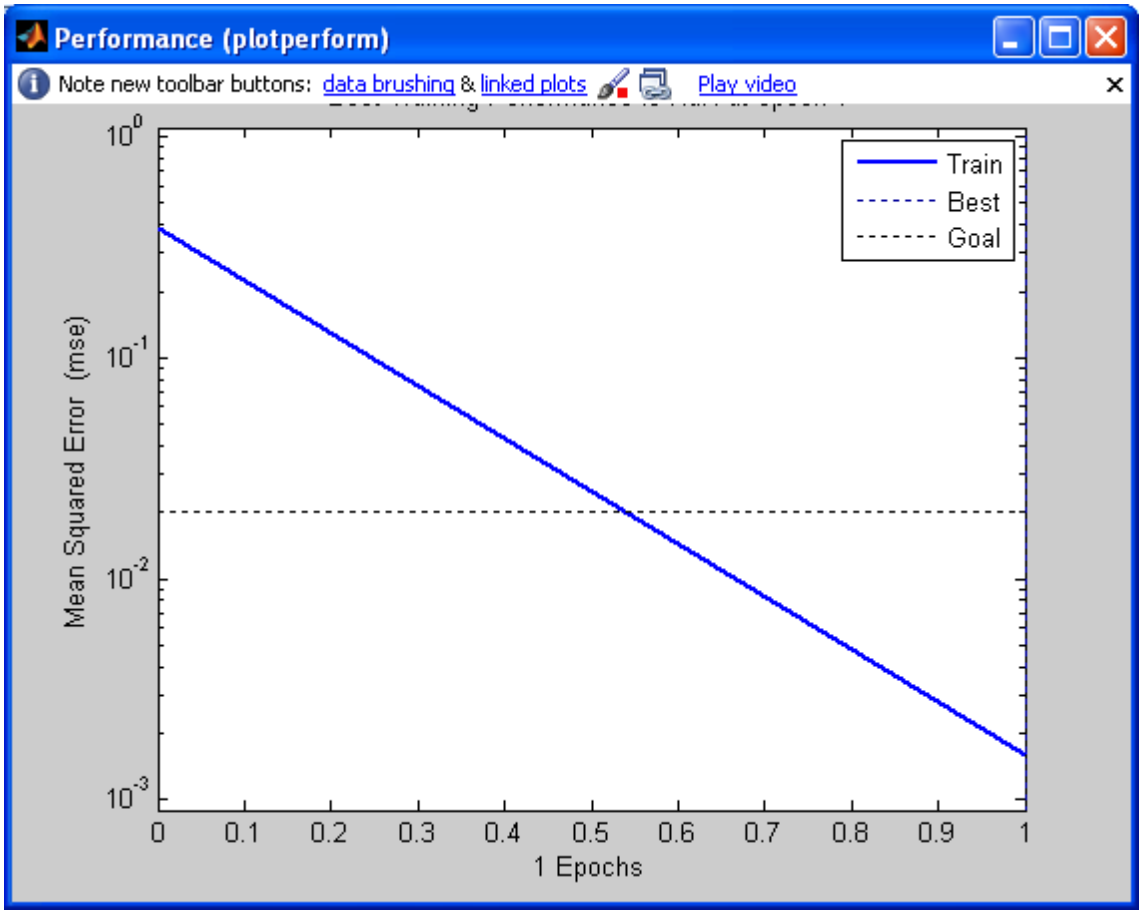

Fig. 5 Performance plot

5.0 Conclusion

Clinical medicine is one of the most interesting areas in which data mining may have an important practical impact. The widespread availability of large clinical data collections enables thorough retrospective analysis, which may give healthcare institutions an unprecedented opportunity to better understand the nature and peculiarity of the undergoing clinical processes. In present work, we have designed a system to identify the chances of a coronary heart disease. We have divided all parameters in two levels according to criticality and each level is assigned separate weightage. Finally both levels are considered to derive a final decision. We have implemented neuro-fuzzy integrated approach at two levels. So, error rate is very low and work efficiency is high. In this work, we have performed the analysis for coronary heart disease. In future, we can use the same neuro-fuzzy integrated approach to perform the analysis on some other disease.

References

[1] Shusaku Tsumoto," Problems with Mining Medical Data", 0-7695- 0792-1 I00@ 2000 IEEE.

[2] Y. Alp Aslandogan et. al.," Evidence Combination in Medical Data Mining", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.

[3] Constantinos Koutsojannis et. al., "Using a Neurofuzzy Approach in Medical Application", Springer-Verlag Berlin Heidelberg, 2007 (pp 477-484).

[4] Hai Wang et. al.," Medical Knowledge Acquisition through Data Mining", Proceedings of 2008 IEEE International Symposium on IT in Medicine and Education 978-1-4244-2511-2/08©2008 Crown.

[5] Markos G. Tsipouras et. al., "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 4, JULY 2008.

[6] Harsh Vazirani et. al.," Use of Modular Neural Network for Heart Disease", Special Issue of IJCCT Vol.1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5 August 2010 (pp 88-93).

[7] Ali.Adeli et. al., "A Fuzzy Expert System for Heart Disease Diagnosis", Proceedings of the international MultiConference of Engineers and computer scientists 2010 Vol 1, March 17-19, 2010, Hong Kong, ISSN 2078-0966.

[8] A.Q. Ansari et. al.," Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System", 2011 World Congress on Information and Communication Technologies 978-1-4673-0125-1@ 2011 IEEE (pp 1383-1388).

[9] Nidhi Bhatla, Kiran Jyoti, "A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic", International Journal of Computer Applications, Volume 54– No.17, (pp 16-21), September 2012, ISSN 0975 – 8887.

[10] Chaitrali S. Dangare et. al., "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (IJCA) (0975 – 8887), Vol. 47, No. 10, June 2012 (pp 44-48).

[11] S. Vijiyarani et. al., "An Efficient Classification Tree Technique for Heart Disease Prediction",International Conference on Research Trends in Computer Technologies (ICRTCT - 2013) Proceedings published in International Journal of Computer Applications (IJCA) (0975 – 8887), 2013 (pp 6-9).

[12] Priyanka Palod,"EFFICIENT MODEL FOR CHD USING ASSOCIATION RULE WITH FDS", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 6, June 2013, ISSN: 2319-8753.

[13] The Wikipedia website [online]. Available at: http://www.wikipedia.org/

[14] The UCI machine learning repository [online]. Available at: http://www.ics.uci.edu/~mlearn/

[15] The National Heart, Lung and Blood Institute official website [online]. Available at: http://nhlbi.nih.gov/.

[16] Fayyad et al., the Primary Tasks of Data Mining, 1996 [online]. Available at: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/2_tasks.html

Author Profile

Ashish Kumar Sen, received the M.Phil degree in Computer Science and MCA from Guru Jambheshwar University Hisar, Haryana, India. He is currently working towards the Department of Computer Science and Mathematics, Govt. PG Science College, Rewa (M.P.)-India. His research interests include datamining, intelligent techniques and Bioinformatics.
E-mail: ashishsen1983@gmail.com