Heart Disease Anticipation Through Genetic Algorithm And BP-Neural Network

Jaishri[1], Santosh Biradar[2]
[1, 2] Dept of Computer Engineering
[1, 2] D.Y. Patil College of Engineering, Ambi

Abstract- *Heart disease impacts few individual's health issues. One such state is Cardiovascular Disease (CVD) which causes the risk for life if not recognized and cured at an early stage. There is not enough focus on diagnostic devices to find networks and information patterns, especially in the medical department. The medical industry collects a large number of heart statistics that make it difficult to find unknown material to make important decisions. The human services industry now a days generates a many medical statistics about the assets of patients and other treatment facilities. Information mining schemes used to review such statistics from a different perspective and by extracting valuable data. The plan is to explore and create a framework for the discovery and expectations of heart disease through sensible mining. Number of decisions were made to focus on the execution of various statistics extraction methods, including neural network and genetic algorithms. In the planned work, the attribute organized clinical database of information is used as the basic information. The neural network provides useful information of the given statistics*

Keywords- BP Neural Network, Data mining, Genetic Algorithm, Heart Disease prediction

I. INTRODUCTION

Cardiac disease is a brutal disease. There are several types of cardiac disease, coronary artery disease, congestive heart death, by multiple heart attacks, if not projected early. But if this is anticipated early, it can be cured. There are similar signs of heart disease, which can be predicted based on heart disease. Symptoms include chest pain, difficulty breathing, swelling, and cyanosis. If heart disease is projected, one can be cured in hospitals. Predicting heart disease in the early stages is beneficial for patients in terms of cost and time. Early predictions can decrease the risk of heart stroke with the right medication. The most common cause of death in India or other Asian countries is heart disease. In 2003, an estimated 17.3 million people died worldwide, and 10 million of these were due to coronary heart disease alone. Without lifestyle changes, smoking, alcohol, obesity, high blood pressure, diabetes, etc., can lead to heart problems. However, recent studies have shown that artificial intelligence and medical science can really help prevent such illnesses. To make a good decision, machine learning helps to extract relevant data from the large database available at the hospital.

II. LITERATURE SURVEY

Few of the methods used to recognize heart disease in the literature are Waveform analysis, time-frequency analysis, Neuro Fuzzy RBF ANN, and Total Last Square-based Prony modeling algorithm. However, in a research by Marshall (Marshall in 1991), classification accuracy was not good with this practice (up to 79%) and the series of enhancements to select the suitable model was still adequate. They verified the effectiveness of neural networks in diagnosing heart attacks (acute myocardial infarction) by comparing multiple neural network classifiers, the multilayer perceptron and the Boltzmann perceptron classifier. Most of these methods are not related to the understanding of basic knowledge but relate to the diagnosis.

Attribute selection for high-level comprehensible knowledge can be made prior to genetic programming, this is concluded by (Celia et al. 2000). Following this study, a multiplayer perceptron-based decision support system that supported the diagnosis of heart disease is developed by Hongmei Yan (Hongmei Yan 2006).

The subsequent job in this arena was by Huang Jianyong (2004). To solve the problem of multi-classification using binary classifiers is the aim of this study. Tree design with Binary classification is done according to the class group from one node is divided into two different subgroups. Node adopts class module method to improve binary classification capability. Partitioning is created as an optimization issue, and genetic algorithms are given to solve the optimization issues. In pediatric and vascular diseases, heart palpitations were a major problem, and high heart rate (reported 77 - 95% was reported in this group) resulting in congestive heart failure. A reliable test device to identify palpitations in pediatrics was developed by Sanjay (Sanjay 2005).

Carlos Ordonez [2] been through the prognosis of heart disease using rules of the association. Simple mapping algorithm is used by them. These algorithms treat properties as numerical or hierarchical. It is used to renovate medicinal histories into transactional form. Improved procedures are used to reduce limited association guidelines. A plotting table is created and its attribute value is plotted. For data mining Decision trees are used as, they divide spontaneously numerical data. Selected split points are rarely used in decision trees. Clustering is used to make a holistic understanding of data.

Pruning Classification Association Rules (PCAR) is studied by Deepika [3] to draw up the rules of association. PCAR are dependent on study of Apriori algorithm. PCAR consists of a set of least occurrence entries, that extracts and eliminates irregular items from a set of items. Based on the frequency of the set of items it then sorts the set of items and finds the set of repeated items. Ordonez uses association rules in data mining to achieve better predictive outcomes for heart disease. The limitations of the association rules that set the whole data without verifying individual samples was studied by the author.

R. Setthukkarasi [5] established a neural obscure method to detect the evidences in patient reports for the disease. The general database was configured to make decisions from a low set of properties, which is a result of genetic algorithms. Four-layer opaque neural networks considered for effective modeling and for time-dependent things in ambiguity. A system for predicting heart disease was implemented by Chaitrali Dangare [7] and has used three techniques of data mining classification: Neural Networks, Decision trees & Naive Bayes. The outcome demonstrates that neural networks are better than Naive Bayes and decision trees.

The feature subgroup selection to make class association rules for detecting heart disease is used by P. Chandra and M. Jabbar. The association between attribute principles and arrangement to forecast classification in patient data sets is determined by the rules of association. Functional selection, such as hereditary detection, finds the characteristics that help forecast heart disease. A system is projected by Usha Rani for forecasting disease related to heart with grouping of backpropagation and Feed forward algorithms with the help of artificial neural networks.

M. Akhil Jabbar [6], B L Deekshatulu and Priti Chandra have proposed a new algorithm for efficient classification with the combination of genetic algorithms and KNN. To provide genetic algorithms Ideal Solution Execute a universal exploration for complex bigger and several modal data sets. It has also been noticed from the results that the hybrid has worked well with KNN and offers accurateness. An effective hereditary algorithm mix is proposed by Ankita Diwan using back propagation to predict heart disease. They concluded that the neural network for nonlinear data is the finest among all other techniques. The BP algorithm is a common range method of artificial neural networks [4]. Whereas the primary system result is compared to the anticipated result and the system is adjusted until the difference is reduced among the two. However, there is a lack of stuck in the local minima

To develop a decision base in the cardiovascular prognosis system, Asma Parveen with Shadab Adam Pattekari [8] used the Navy Bayesian classification technique. The system detects unseen information from previous databases of heart disease. This is the greatest operative model for forecasting heart disease. This model can answer composite questions and the model definition has its individual strengths in terms of effortlessness, access to more evidence and correctness.

III. PROPOSED FRAMEWORK

Two algorithms are used here, which are Back Propagation Neural Network (BPNN) and Genetic Algorithm (GA). Using this algorithm on the opti, reducing the optimal properties along with the properties on the dataset can predict heart disease. The Dataset is used from UCI (Machine Learning Repository) repository which includes attributes such as sex, cholesterol, Blood pressure and additional more attributes such as obesity and smoking. The data set having the attributes which is using the inputs to Neural network, then the neural network is using input as BPNN and genetic algorithm these two is using as an input to reduced optimal attribute for predicted result as shown in fig 1 below.

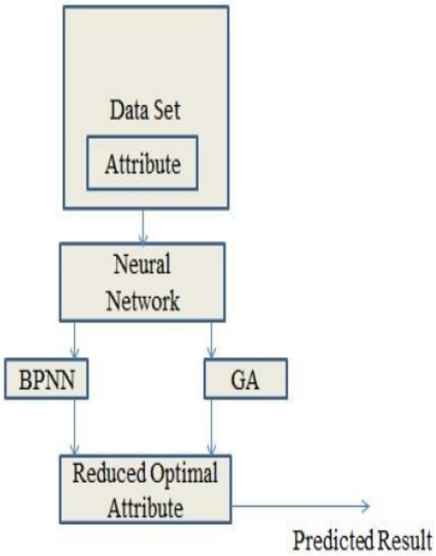

Fig 1: Architecture Diagram

IV. PROPOSED ALGORITHMS

A. Back-propagation Neural Networks -(BPNN)

On error correction the [1] backpropagation technique is used in neural networks. Forward pass and backward pass are the two main passes in Back-propagation Neural Network. This is shown in fig 2

Forward Pass: It works layer by layer. This takes the input node and affects the network layer to which the activity outline is used. The definite response produced is set of outputs. The weight is static throughout this technique.

Backward pass: The weight is adjusted according to certain rules. An error signal is generated by subtracting the actual response from the desired response. This error signal is traversed by the network.

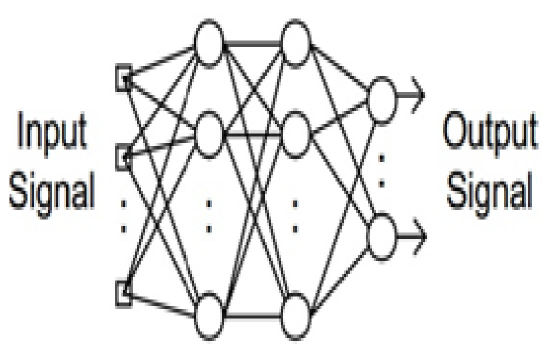

Fig 2. BPNN hidden layers

B. Data Mining

Data mining techniques help in extracting and verifying complex medical data using a variety of techniques [2]. In 2012 Chaitrali and Sulabha [7] executed a relative study of three different data mining classification techniques, that is: decision trees, Naive Bayes and Neural Network on heart Disease and K.Srinivas [9] in 2010 also made a relative analysis on these 3 Techniques of popular data mining.

C. Genetic Algorithm

The human DNA genome consists of 46 chromosomes, consisting of four different base string, abbreviated as A, T, G and C, summarizing about three billion nucleotides. It can be structured into genes, which give one or more information about a person's construction. However, it is estimated that only 3% of genes contain meaningful information, most genes - "junk" genes are not used.

Genetic information the genome itself is called the genetic makeup of that person. As a result, the person is called a phenotype. The same genotype can lead to different phenotypes. The basic GA operators are crossover, selection and mutation. Here are some of the parameters used in the genetic algorithm as shown in Table 1 below.

With this technique we are not able to get maximum profit. So, we can use an excellent one to solve this problem Optimizer is a genetic algorithm that uses mutations and crossover events across different generations. The weight used for BP can be optimized using genetic algorithm first and then given as input to give better results to our network. Weight adjustment using GA is done with ‘population size =20’.

Table I : Parameters used in GA

SOME PARAMETERS USED IN GA

Search Method	Genetic Algorithm
Population Size	20
Generations	100
Cross Over Fraction	0.8000
Migration Interval	20
Migration Fraction	0.2000
Elite Count	2
Tolfun	1.0000e-006

Basic structure of the genetic algorithm is explained in Fig 3. It starts with a random generation of early sets of early people

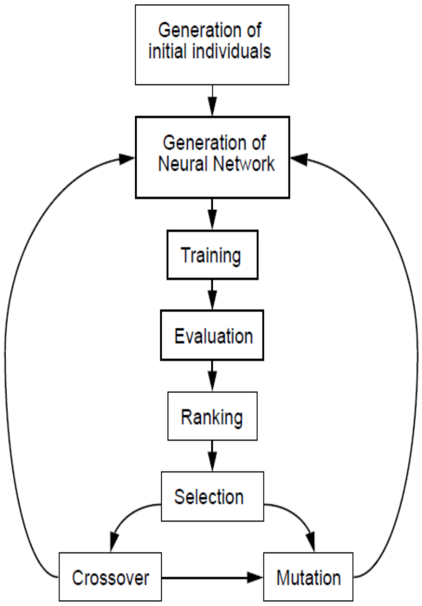

Fig 3. GANN System Principle structure

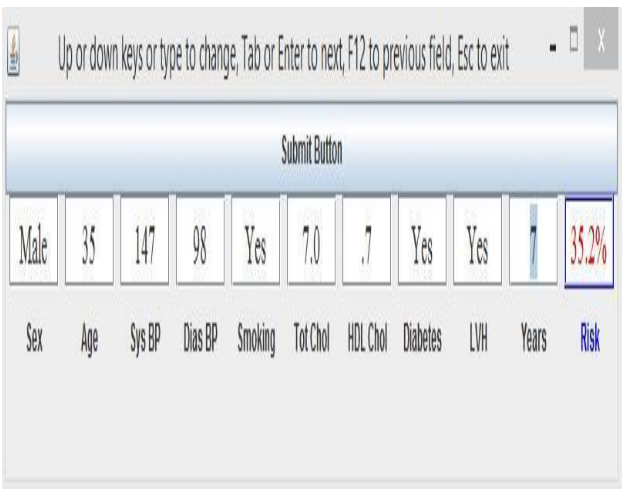

Fig 4: Parameters Used in Health profile

The parameters used in the health profile to check the patients risk factors is shown in fig 4.

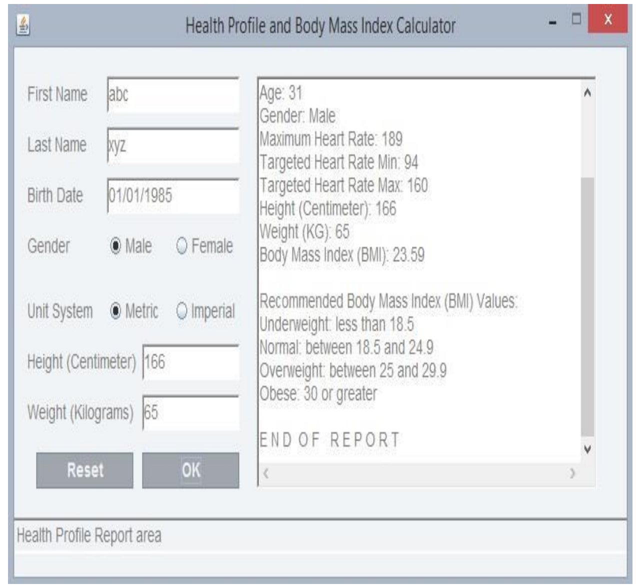

Fig 5: Health Profile and body mass calculator

The health profile and body calculator for the patients is shown in fig 5. Here we are using the patient first name, last name, birth date, gender (male or female), unit system (metric or imperial), height (centimetre) and weight of the patient to calculate patient health profile to get end of report.

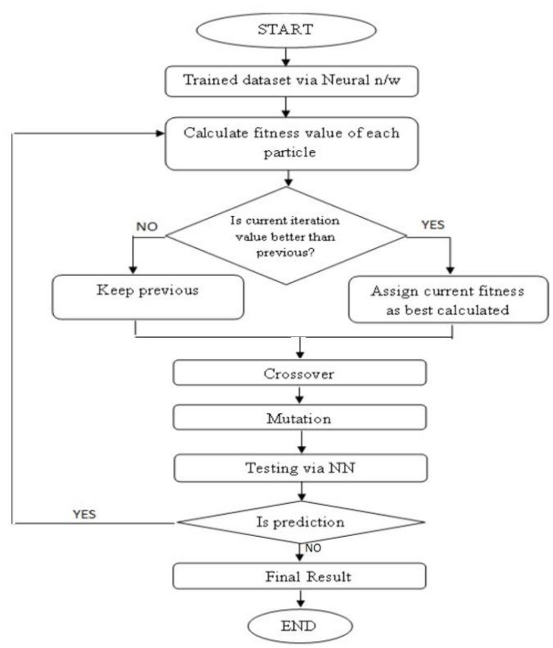

Fig. 6. Proposed Methodology algorithm

It turns out that the best classification technique that can use in our domain is nothing but back propagation as it is the only technique used for nonlinear relationships. But it lacks due to being stuck in the locale minima and not able to get us maximum profit from this technique.

So to solve this problem we can use an excellent optimizer i.e. genetic algorithm that uses mutations and crossover events in different generations. The weight used for BP can be optimized first and then given as input to give better results to our network as shown in above fig 6

Table II : Case Study Data of patient

No	Sex	Age	Blood Cholesterol	Blood Pressure	Hereditary	Smoking	Alcohol Intake	Physical Activity	Diabetes	Diet	Obesity	Stress	Heart Disease
1	Female	35	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
2	Male	70	Low	Low	No	No	Yes	High	Yes	Normal	No	No	No
3	Female	60	High	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
4	Female	36	Low	Normal	No	No	No	Normal	No	Good	No	No	No
5	Male	30	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
6	Female	39	Low	Normal	Yes	No	Yes	High	Yes	Normal	No	Yes	No
7	Female	41	High	Normal	No	No	No	Low	No	Poor	Yes	No	No
8	Male	70	High	Normal	No	No	Yes	Low	No	Poor	Yes	No	Yes
9	Male	65	Normal	High	Yes	Yes	Yes	Normal	Yes	Poor	Yes	No	Yes
10	Male	30	Normal	High	No	Yes	No	Normal	No	Good	No	Yes	No
11	Female	31	Low	Normal	No	No	No	High	No	Normal	No	No	No
12	Female	29	Low	Normal	No	No	Yes	High	No	Good	No	No	No
13	Male	30	Low	Normal	No	No	Yes	Normal	No	Normal	No	No	No
14	Female	45	Normal	High	Yes	Yes	No	Normal	Yes	Normal	Yes	Yes	No
15	Male	25	High	Normal	Yes	Yes	Yes	Low	Yes	Normal	No	No	Yes
16	Female	37	Normal	Normal	No	No	No	Normal	Yes	Poor	No	Yes	No
17	Female	37	Normal	High	No	Yes	Yes	High	No	Poor	No	Yes	No
18	Male	53	High	Low	No	Yes	No	Normal	Yes	Normal	No	Yes	No
19	Male	57	High	Normal	No	Yes	No	Low	No	Poor	Yes	Yes	Yes
20	Male	52	High	Low	No	No	No	Normal	Yes	Poor	Yes	No	No
21	Male	48	Normal	Normal	Yes	Yes	Yes	Normal	No	Normal	No	No	Yes
22	Male	62	High	High	No	Yes	Yes	Normal	Yes	Normal	No	No	Yes
23	Male	56	Normal	High	No	Yes	Yes	Low	No	Poor	Yes	Yes	Yes
24	Female	27	Low	Normal	No	No	No	High	No	Good	No	No	No
25	Male	33	Normal	Normal	No	No	No	Normal	Yes	Good	No	No	No
26	Female	33	Normal	Normal	No	No	Yes	Low	Yes	Poor	No	Yes	No
27	Male	37	High	Normal	No	No	Yes	Normal	No	Normal	No	Yes	No
28	Male	43	Normal	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
29	Male	46	Low	Normal	No	No	No	Normal	Yes	Poor	Yes	Yes	No
30	Female	36	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
31	Female	29	Low	Normal	No	No	No	Normal	No	Good	No	No	No
32	Female	47	Normal	Normal	No	No	Yes	High	Yes	Normal	No	Yes	No
33	Male	58	High	High	No	Yes	Yes	Normal	Yes	Normal	No	Yes	Yes
34	Male	44	High	Normal	Yes	Yes	Yes	Normal	No	Normal	Yes	Yes	Yes
35	Female	36	Normal	High	No	No	No	Normal	No	Good	Yes	No	Yes
36	Male	42	Low	Normal	Yes	No	Yes	Low	No	Poor	No	Yes	No
37	Female	25	Low	Normal	No	No	No	High	No	Poor	No	No	No
38	Female	28	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
39	Female	26	Low	Normal	Yes	No	No	Normal	No	Normal	Yes	No	Yes
40	Male	28	Low	Normal	No	No	No	Normal	No	Poor	No	No	No
41	Female	45	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
42	Male	63	Low	Low	No	No	Yes	High	Yes	Good	No	No	No
43	Female	55	High	High	No	No	No	Normal	Yes	Normal	Yes	Yes	Yes
44	Female	44	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
45	Male	35	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
46	Female	42	Normal	Normal	No	No	Yes	High	Yes	Good	No	No	No
47	Female	43	Normal	Normal	No	No	No	Low	No	Poor	Yes	No	No
48	Male	65	Normal	Normal	No	No	Yes	Low	No	Normal	Yes	Yes	Yes
49	Male	74	Normal	High	No	Yes	Yes	Normal	Yes	Normal	Yes	Yes	Yes
50	Male	36	Normal	High	No	Yes	No	Normal	No	Poor	No	No	No

V. RESULT AND DISCUSSIONS

MATLAB R2012a is used for the development of the system. The Global Optimization and the Neural Network Toolbox were used to implement the algorithm [11]. On the American Heart Association's [10] website, the input data with risk factors collected from 50 people through a case study provided in Table 2. The information was determined as shown in Table 2. 15% data was used for testing, 15% for validation and 70% for training. Using Matlab a confusion matrix is produced and accuracy is determined (shown in Table 3) as

Accuracy = (TP + TN) / (TP + FP + TN + FN);

where FP, FN, TP and TN denote False positives, false negatives, true positives and true negatives, respectively. The accuracy of prediction of heart disease on the training data was calculated as 96.2% and accuracy on validation data was 89%.

The fitness function is calculated based on the mean square error for each chromosome. The function used is mean square error (mse) which is calculated as below:

$$mse = \sum_k (O_k - T_k)^2 / n$$

The least mean square error (MSE) achieved was 0.034683 after 12 epochs, as shown in Fig 7. Results show genetic algorithm and neural network approach [4] gives better average prediction accuracy than the traditional ANN

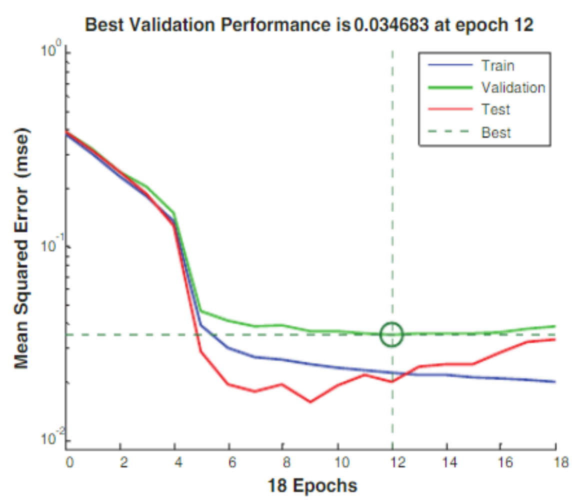

Figure 7: Performance Graph

Table 3 : Data Sets

Data Set	Number of data	Accuracy (%)
Training Set	34	96.2%
Test Set	8	92%
Validation Set	8	89%
Total instances	50	.

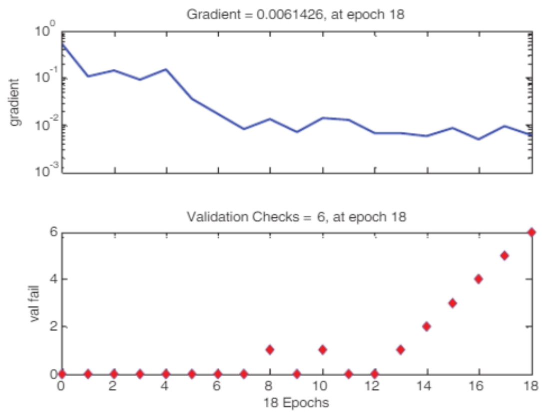

Fig 8: Training State Graph

VI. CONCLUSIONS

It can be concluded that among all the classification techniques the neural network is the best techniques when we talk about the prediction of non-linear data. Predicting heart disease will help both society and the medical industry. But we still require mechanisms that will forecast heart disease in the initial stages. In this study, a new hybrid model of neural networks and genetic algorithms used to optimize the weight of ANN connections to improve the efficiency of artificial neural networks. The system uses identified important risk factors for the prediction of heart disease and it does not require costly medical tests.

VII. ACKNOWLEDGEMENT

We would like to thank Prof. Santosh Biradar and Prof. Deepali. for making their assets accessible. We also thank the institution authorities for providing the appreciative base and support.

REFERENCES

[1] S.Florence , N.G.Bhuvaneswari Amma, G.Annapoorani, K.Malathi, "Predicting the Risk of Heart Attacks using Neural Network and Decision Tree", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 11, November 2014

[2] Carlos Ordonez, Edward Omincenski and Levien de Braal ,"Mining Constraint Association Rules to Predict Heart Disease", Proceeding of 2001, IEEE International Conference of Data Mining, IEEE Computer Society, ISBN-0-7695-1119-8, 2001, pp: 433- 440

[3] Deepika. N, "Association Rule for Classification of Heart Attack Patients ", IJAEST, Vol 11(2), pp 253-257, 2011.

[4] Usha. K Dr, "Analysis of Heart Disease Dataset using neural network approach", IJDKP, Vol 1(5), Sep 2011.

[5] R. Setthukkarase, Kannan, "An Intelligent System for mining Temporal rules in Clinical database using Fuzzy neural network",European Journal of Scientific Research, ISSN 1450-216, Vol 70(3), pp 386-395, 2012.

[6] M Akhil Jabbar, BL Deekshatulu, Priti Chandra," Heart disease classification using nearest neighbor classifier with feature subset selection", Anale. Seria Informatica, 11, 2013.

[7] Chaitrali S Dangare, Sulabha S Apte." Improved study of heart disease prediction system using data mining classification techniques", International Journal of Computer Applications, 47(10):44–48, 2012.

[8] Shadab Adam Pattekari and Asma Parveen," PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624, Vol 3, Issue 3, 2012, pp 290-294.

[9] K.Srinivas B.Kavihta Rani Dr. A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering (IJCSE) Vol. 02, No. 02, 2010, 250-255

[10] American Heart Association, http://www.heart.org/HEARTORG/Conditions

[11] J. Guo, and W. J. Sun, Theory of Neural Network and its Implementation with MatLab , Electronic Industry Press, Beijing, 2005.