International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 06 | June -2020 ISSN: 2582-3930

HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES & ARTIFICIAL INTELLIGENCE

Abhishek Verma, Ashutosh Shukla, Ashwani Kumar

Department of Electrical and Electronics Engineering
College of Engineering Roorkee, Roorkee

Abstract

In modern world, heart diseases are rapidly increased due to the food habit, stress, genetic reason and also lack of exercise. The prediction of the heart disease helps the people to take care of their health. Nowadays, Health care institutes contains huge amount of information, which contains hidden information of patients health. This hidden information is useful for making effective decisions. The factors such as high cholesterol, high blood pressure, blood sugar level, obesity are the major risk causes heart diseases. We designed the system using MATLAB. Computer based information along with advanced Data mining techniques like fuzzy logic,Neural Network and Machine Learning is used for predicting Heart disease diagnosis. Here, artificial intelligence played an important role in diagnosis of heart disease with improved effectiveness. Based on this perspective, several researches have been conducted in the literature recently. So, analysing those diagnosis techniques can lead to new development in this area. Accordingly, we present a detailed survey of 47 articles published in the standard journals from the year 2005 to 2013.From the survey the finding is that neural network based techniques contribute more effectiveness and some techniques have obtained more than 90% accuracy.

I. Introduction

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Healthcare industry today generates large amount 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 analysed for knowledge extraction that enables support for cost-savings and decision making. Data mining provides a set of tools and techniques that can be applied to this processed data to discover hidden patterns and also provides healthcare professionals an additional source of knowledge for making decisions.Coronary heart disease (CHD) can be caused due to risk factors like high blood pressure, high blood cholesterol, tobacco use, obesity, unhealthy diet, physical inactivity, diabetes, advancing age, and inherited disposition. Coronary heart disease (CHD) is the narrowing or blockage of the coronary arteries, usually caused by atherosclerosis. Atherosclerosis (sometimes called “hardening” or “clogging” of the arteries) is the build-up of cholesterol and fatty deposits (called plaques) on the inner walls of the arteries. These plaques can restrict blood flow to the heart muscle by physically clogging the artery or by causing abnormal artery tone and function.Without an adequate blood supply, the heart becomes starved of oxygen and the vital nutrients it needs to work properly. This can cause chest pain called angina. If blood supply to a portion of the heart muscle is cut off entirely, or if the energy demands of the

heart become much greater than its blood supply, a heart attack (injury to the heart muscle) may occur. Coronary heart disease (CHD) is the leading cause of death for both men and women and accounts for approximately 600,000 deaths in the United States every year.

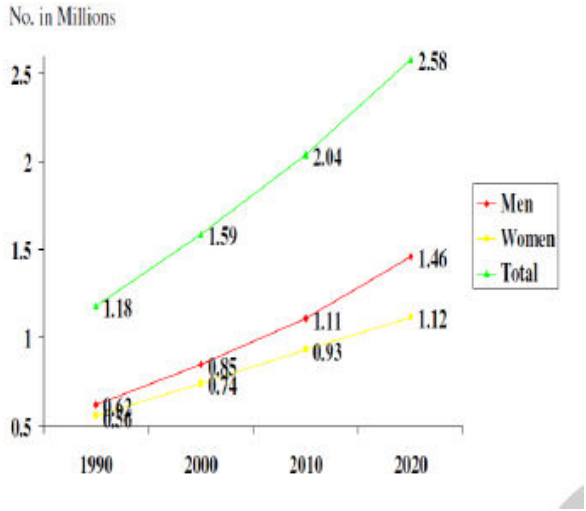

Fig:-1 Projections for Coronary Heart Disease (CHD) Mortality in India.

3. LITERATURE SURVEY

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

1. An Intelligent Heart Disease Prediction System (IHDPS) is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by SellappanPalaniappan et al . Each method has its own strength to get appropriate results. To build this system hidden patterns and relationship between them is used. It is web-based, user friendly & expandable.

2.To develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) a novel technique was proposed by HeonGyu Lee et al. To achieve this, they have used several classifiers e.g. Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), (Decision Tree) and SVM (Support Vector Machine)

3. The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. The dataset contains records with 13 attributes in each record. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data.

4. The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez. The resultant dataset contains records of patients having heart disease. Three constraints were introduced to decrease the number of patterns. They are as follows:

1. The attributes have to appear on only one side of the rule.
2. Separate the attributes into groups. i.e. uninteresting groups.
3. In a rule, there should be limited number of attributes.

The result of this is two groups of rules, the presence or absence of heart disease.

5. Franck Le Duff et al. [9] builds a decision tree with database of patient for a medical problem.

6. LathaParthiban et al. projected an approach on basis of coactive neuro-fuzzy inference system (CANFIS) for prediction of heart disease. The CANFIS model uses neural network capabilities with the fuzzy logic and genetic algorithm.

7. Kiyong Noh et al. uses a classification method for the extraction of multiparametric features by assessing HRV (Heart Rate Variability) from ECG, data pre-processing and heart disease pattern. The dataset

consisting of 670 peoples, distributed into two groups, namely normal people and patients with heart disease, were employed to carry out the experiment for the associative classifier.

4. PROPOSED PREDICTION SYSTEM

Today, many hospitals manage healthcare data using healthcare information system; as the system contains huge amount of data, used to extract hidden information for making intelligent medical diagnosis. The main objective of this research is to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. To develop this system, medical terms such as sex, blood pressure, and cholesterol like 13 input attributes are used. To get more appropriate results, two more attributes i.e. obesity and smoking are used, as these attributes are considered as important attributes for heart disease. The data mining classification techniques viz. Neural Networks, Decision Trees, and Naive Bayes are used.

5. DATA SOURCE

The publicly available heart disease database is used. The Cleveland Heart Disease database consists of 303 records &Statlog Heart Disease database consists of 270 records. The data set consists of 3 types of attributes: Input, Key & Predictable attribute which are listed below.

5.1 INPUT ATTRIBUTES

Sr. No.	Attributes	Description	Value
1	age	Age in years	Continuous
2	sex	Male or female	1 = male 0 = female
3	cp	Chest pain type	1 = typical type 1 2 = typical type agina 3 = non-agina pain 4 = asymptomatic
4	thestbps	Resting blood pressure	Continuous value in mm hg
5	chol	Serum cholesterol	Continuous value in mm/dl
6	Restecg	Resting electrographic results	0 = normal 1 = having_ST_T wave abnormal 2 = left ventricular hypertrophy
7	fbs	Fasting blood sugar	1 ≥ 120 mg/dl 0 ≤ 120 mg/dl
8	thalach	Maximum heart rate achieved	Continuous value
9	exang	Exercise induced agina	0= no 1 = yes
10	oldpeak	ST depression induced by exercise relative to rest	Continuous value

11	solpe	Slope of the peak exercise ST segment	1 = unsloping 2 = flat 3 = downslopin g
12	ca	Number of major vessels colored by floursopy	0-3 value
13	thal	Defect type	3 = normal 6 = fixed 7 = reversible defect

Table 1. Description of 13 input attributes

5.3. Predictable attribute

Diagnosis: Value 1 = < 50 % (no heart disease)

Value 0 = > 50 % (has heart disease)

6. DATA MINING TECHNIQUES USED FOR PREDICTIONS

The three different data mining classification techniques, i.e. Neural Networks, Decision Trees, and Machine Learning.

6.1. Neural Networks

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural network. In other words, it is an emulation of biological neural system. A Multi-layer Perceptron Neural Networks (MLPNN) is used. The structure of MLPNN is as shown in Figure 1.

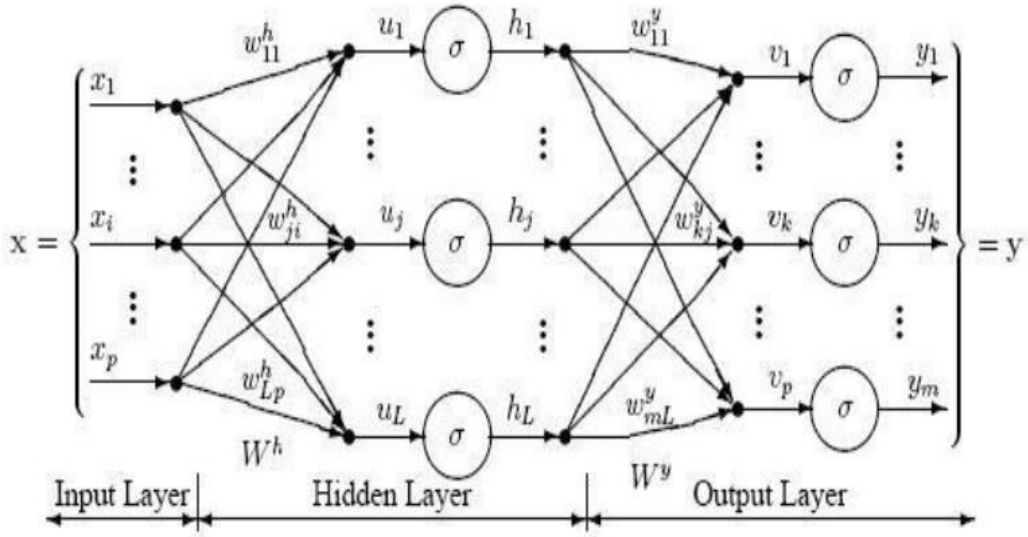

Figure 2: Structure of Multi-Layer Perceptron Neural Network

It maps a set of input data onto a set of appropriate output data.It consists of 3 layers input layer, hidden layer & output layer. There is connection between each layer & weights are assigned to each connection. The primary function of neurons of input layer is to divide input xi into neurons in hidden layer. Neuron of hidden layer adds input signal xi with weights wji of respective connections from input layer. The output Yj is function of Yj = f

(Σ wji xi) Where f is a simple threshold function such as sigmoid or hyperbolic tangent function.

6.2. Decision Trees

The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, C4.5, CHAID, and J48. From these J48 algorithm is used for this system. J48 algorithm uses pruning method to build a tree. Pruning is a technique that reduces size of tree by removing overfitting data, which leads to poor accuracy in predications. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible. This technique gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy.

6.4 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed**.** Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Some machine learning methods

Machine learning algorithms are often categorized as supervised or unsupervised.

- Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
- In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.
- Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise,

acquiringunlabeled data generally doesn't require additional resources.

- Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

The dataset consists of total 270 records in Heart disease database. The total records are divided into two data sets one is used for training consists of 222 records & another for testing consists of 48 records.
Sampled data for the propsed system is shown below in the figure. It has total 13 different attributes in 13 different coulmn and other 1 coulmn shows the result for the Heart disease ie 0 for absence and 1 for present.

7. RESULTS

45	0	2	112	160	0	0	138	0	0	2	0	3	1
55	0	4	180	327	0	1	117	1	3.4	2	0	3	2
41	1	2	110	235	0	0	153	0	0	1	0	3	1
60	0	4	158	305	0	2	161	0	0	1	0	3	2
54	0	3	135	304	1	0	170	0	0	1	0	3	1
42	1	2	120	295	0	0	162	0	0	1	0	3	1
49	0	2	134	271	0	0	162	0	0	2	0	3	1
46	1	4	120	249	0	2	144	0	0.8	1	0	7	2
56	0	4	200	288	1	2	133	1	4	3	2	7	2
66	0	1	150	226	0	0	114	0	2.6	3	0	3	1
56	1	4	130	283	1	2	103	1	1.6	3	0	7	2
49	1	3	120	188	0	0	139	0	2	2	3	7	2
54	1	4	122	286	0	2	116	1	3.2	2	2	3	2
57	1	4	152	274	0	0	88	1	1.2	2	1	7	2
65	0	3	160	360	0	2	151	0	0.8	1	0	3	1
54	1	3	125	273	0	2	152	0	0.5	3	1	3	1

Figure 3: Sample data used in Neural Network

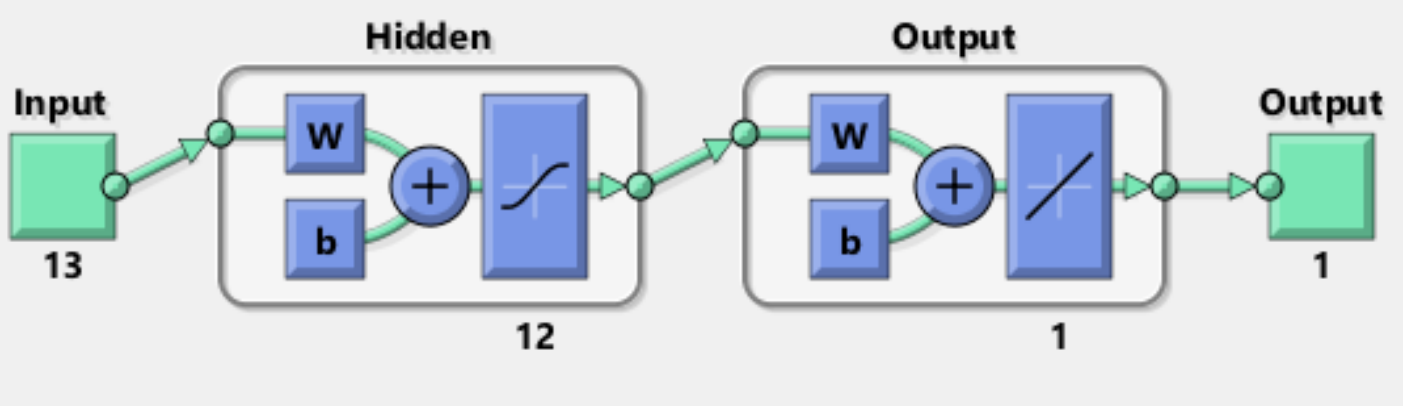

Figure 4: Structure of Multi-Layer Perceptron Neural Network

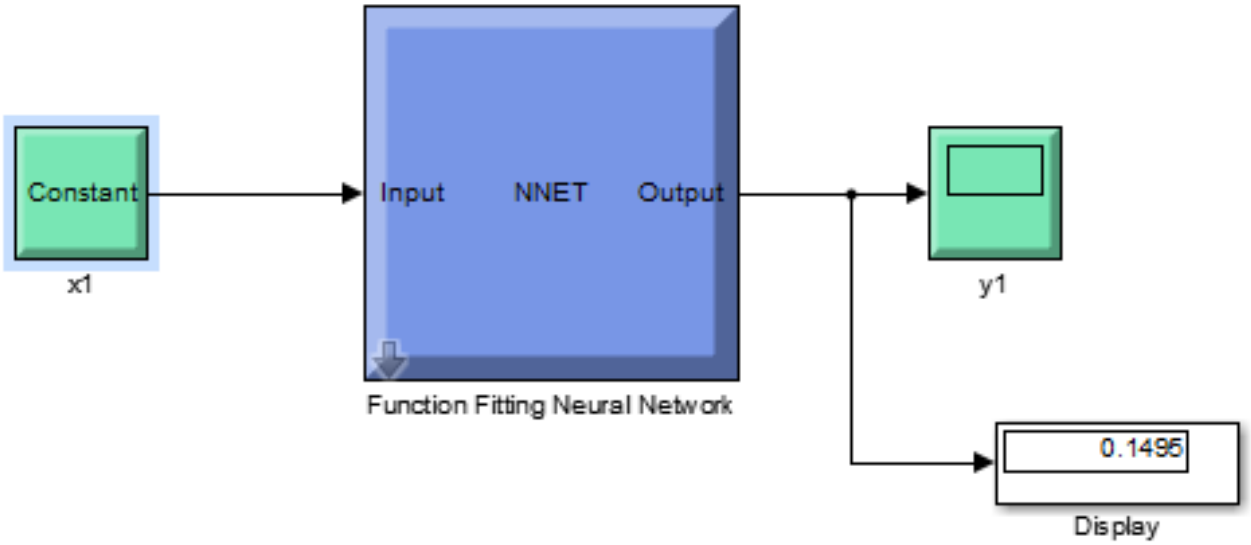

Figure 5: Simulation Diagram of Neural Network

7.1 Result of predictor using machine learning method

```
enter the details of pateint age45

m1 =

    45

enter the details of pateint sex0

m2 =

     0

enter the details of pateint chestpain2

m3 =

     2

enter the details of pateint restbp112

m4 =

   112

enter the details of pateint cholestrol160
```

```
enter the details of pateint blood sugar0

m6 =

     0

enter the details of pateint rest ecg0

m7 =

     0

enter the details of pateint mhr138

m8 =

   138

enter the details of pateint exercise0
```

Figure 6:--output using machine learning

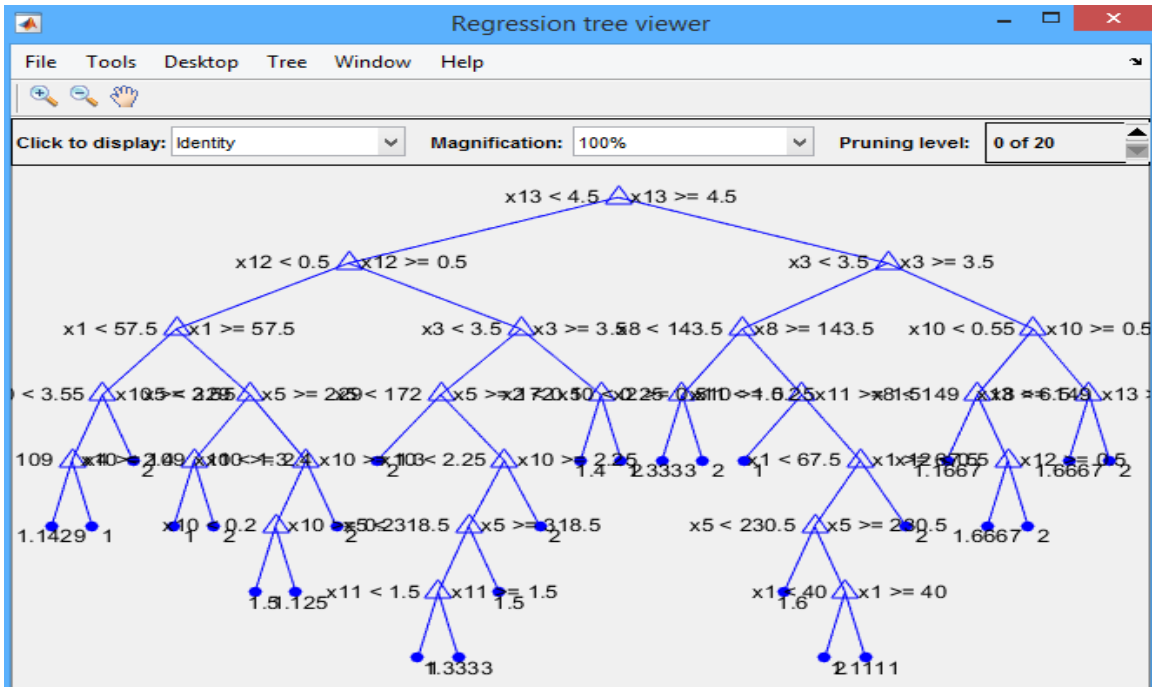

Figure 7:-- Regression Tree Diagram

The source code for the above regression tree plotting and for heart disease prediction by machine learning method written in editor window.

```
x = [age sex cprestbp chol1 blSugrestecgmhr exercise oldpeakslopeofpeakexSTsegnoOfmajorvesseslscolByflorthal]
t= RegressionTree.fit(x,heartdis)
m1=input('enter the details of pateint age')
m2=input('enter the details of pateint sex')
m3=input('enter the details of pateintchestpain')
m4=input('enter the details of pateintrestbp')
m5=input('enter the details of pateintcholestrol')
m6=input('enter the details of pateint blood sugar')
m7=input('enter the details of pateint rest ecg')
m8=input('enter the details of pateintmhr')
m9=input('enter the details of pateint exercise')
m10=input('enter the details of pateintoldpeak')
m11=input('enter the details of pateint slope of peak stseg')
m12=input('enter the details of pateint no of major vessels col by flour')
m13=input('enter the details of pateintthal')
a=predict(t,[m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12,m13])
view(t,'mode','graph')
if a <=1;
disp('heartdisease absence');
end
```

```
if a >1;
disp('heart disease presence');
end
```

The output of the above code is given by heart disease is absence or present and a regression tree.

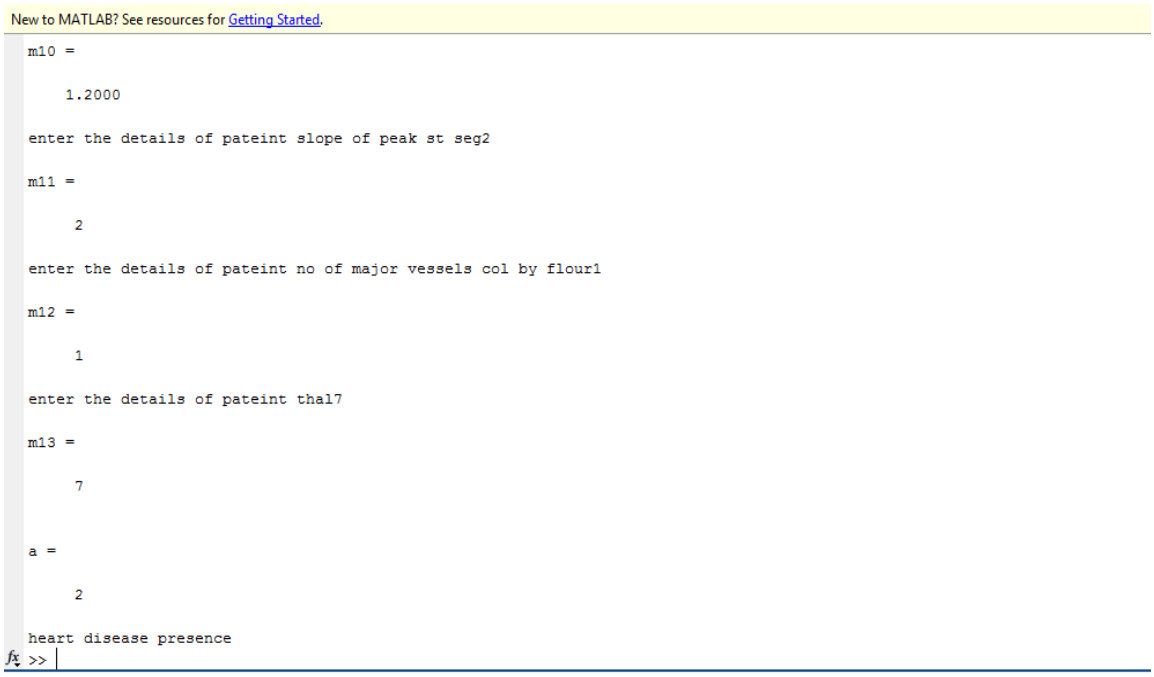

Figure 8:-- prediction of heart disease

8. CONCLUSION

The overall objective of our work is to predict more accurately the presence of heart disease. In this paper, 13 input attributes are used to get more accurate results. Three data mining classification techniques were applied namely Decision trees, Machine Learnig&Neural Networks. From results it has been seen that Neural Networks provides accurate results as compare to Decision trees & Machine Learning. This system can be further expanded. It can use more number of input attributes listed above in table 1 and 2. Other data mining techniques can also be used for predication e.g. Clustering, Time series, Association rules. The text mining can be used to mine huge amount of unstructured data available in healthcare industry database.

9. REFERENCES

[1] NidhiBhatla, and KiranJyoti, "AnAnalysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 1, Oct.2012.

[2] Chaitrali S.Dangare, and Sulabha S.Apte, "Improved Study of Heart Disease Prediction System using Data Mining

Classification Techniques", International Journal of Computer Applications (0975 – 888), Vol. 47, No.10, June.2102.

[3] Dr. T. Karthikeyan, and V.A. Kanimozhi, "Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network", International Journal of Advanced Research in Science, Engineering and Technology, Vol. 4, Issue 1, January 2017.

[4] Hlaudi Daniel Masethe, and Mosima Anna Masethe, "Prediction of Heart Disease Using Classification Algorithms", in Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. II WCECS 2014, 22-24 Oct. 2014, San Francisco, USA.

[5] AshaRajkumar, and Mrs. G. Sophia Reena, "Diagnosis of Heart Disease Using Data Mining Algorithm", Global Journal of Computer Science and Technology, Vol. 10, pp. 38-43, Sept. 2010.

[6] SarangamKodati, and Dr. R Vivekanandam, "A Comparative Study on Open Source Data Mining Tool for Heart Disease", International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, March-2018.

[7] B.Venkatalakshmi, and M.V.Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.

[8] AbhishekTaneja, "Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal of Computer Science and Technology, Vol. 6, pp. 457-466, Dec.2013.

[9] Jaymin Patel, Prof.TejalUpadhyay, and Dr. Samir Patel, " Heart Disease Prediction Using Machine Learning and Data Mining Technique", IJCSC, Vol. 7, No. 1, pp.129-137, September-2015.

[10] K.Gomath, Dr. Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification",International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol.4, Issue 2,February-2016.

[11] G.Subbalakshmi, K. Ramesh, and M.C. Rao, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2, No. 2, Apr-May 2011.

[12] MA.Jabbar, Dr. Priti Chandra, and B.L. Deekshatulu, "Cluster Based Association Rule Mining For Heart Attack Prediction",Journal of Theoretical and Applied Information Technology, Vol. 32 No.2, October-2011.

[13] Nikita Shirwalkar, and TusharTak, "Human Heart Disease Prediction System Using Data Mining Techniques", International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, Mar. 2018.

[14] Navdeep Singh and Sonika Jindal, " Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms", International Journal of Advance Research, Ideas and Innovations in Technology, Vol.4, Issue 2, 2018.

[15] BeantKaur, and Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", International Journal on Recent and

Innovation Trends in Computing and Communication, Vol. 2 Issue 10, October-2014.

[16] AnimeshHazra, S.KumarMandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, Vol. 10, No.7, July-2017.

[17] Stuart J. Russell and Peter Norvig, Artificial Intelligence A Modern Approach, 2nded., New Jersey: Pearson Education Inc., 2003.

[18] Shadab A. Pattekari, and AsmaParveen, "Prediction System For Heart Disease Using Naive Bayes",International Journal of Advanced Computer and Mathematical Sciences, Vol.3, pp. 290-294, 2012.

[19] M.A. Jabbar, B. L. Deekshatulu, and Priti Chandra," Intelligent heart disease prediction system using random forest and evolutionary approach", Journal of Network and Innovative Computing, Vol. 4, pp.174-184, 2016.

[20] M. Akhiljabbar, B.L Deekshatulu and Priti Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm",International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.

[21] ZeinabArabasadi et al. , " Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm" , Computer Methods and Programs in Biomedicine-ELSEVIER, Vol. 141, pp.19- 26, April-2017.