Original Research Paper

Volume-9 | Issue-11 | November - 2019 | PRINT ISSN No. 2249 - 555X | DOI : 10.36106/ijar

Computer Science

PREDICTION OF TEST FOR HEART DISEASE DIAGNOSIS USING ARTIFICIAL NEURAL NETWORK

Rakshit Bhat — Department Of Computer Engineering Thadomal Shahani Engineering College (of Mumbai University) Mumbai, India

Saahil Chawande* — Department Of Computer Engineering Thadomal Shahani Engineering College (of Mumbai University) Mumbai, India *Corresponding Author

Saniya Chadda — Department Of Computer Engineering Thadomal Shahani Engineering College (of Mumbai University) Mumbai, India

ABSTRACT Recently, several software tools and various algorithms have been proposed by the researchers for developing effective medical decision support systems. Moreover, new algorithms and new tools are continued to develop and represent day by day. Diagnosing of heart disease and predicting test for diagnosis is one of the important issues and many researchers investigated to develop intelligent medical decision support systems to improve the ability of the physicians. Such an automated system for medical diagnosis would enhance medical care and reduce costs. However, accurate diagnosis at an early stage followed by proper subsequent treatment can result in significant life-saving. Test diagnosis can be solved by classification which is one the important techniques of Data mining. Neural Network has emerged as an important tool for classification. The advantages of Neural Network help for efficient classification of given data. This research work will be the implementation of Prediction of test for heart disease diagnosis system. For diagnosis of heart disease and their tests, significantly 13 attributes will be used in the proposed system as per the medical literature. The Multi-layer Perceptron Neural Network with Back-propagation as the training algorithm will be employed. The results thus obtained will have illustrated that the designed predictive system is capable of predicting the test for of heart disease effectively. This system will act as a promising tool for diagnosis of heart disease.

KEYWORDS : Heart Disease, Multilayer Perceptron, Machine Learning.

I. INTRODUCTION

The heart is the organ that pumps blood from the blood vessels to different body parts, with adequate proportion of oxygen and other essential nutritional components that are required. If there is some problem in the pumping action of the heart, the main organs of the body like the brain and kidneys are adversely affected. If the working of the heart stops then the death of the person occurs within few minutes [1]. The term heart disease means different

problems that affect the normal functioning of circulatory system.There are different categories of heart diseases like cardiovascular disease in which the heart and blood vessels are affected and as a result of which the blood is not pumped and circulated properly throughout the body parts. In coronary heart disease, the heart does not get sufficient amount of blood that it requires because of cholesterol and fat that is deposited inside the walls of the arteries that supply the blood to heart. In myocardial infarctions, which is also known as a heart attack, the path in the coronary artery is blocked due to the clotting of blood on the wall of the artery that supplies blood to the heart. In angina, chest pain occurs due to the supply of blood that is inadequate to the heart as a result of which it does not function properly. There are also other forms of heart disease include coronary artery disease, valvular heart disease, stroke, high blood pressure, etc [2].

There exist multiple tests that are conducted for the diagnosis of these heart diseases. One needs to get an angiogram done for the evaluation of coronary artery disease. In case of blockages, one has to get a Cardiac Magnetic Resonance Imaging (MRI) done. The MRI uses magnetic fields and radio waves to produce two- or three-dimensional images of blood vessels**.** These images provide doctors a more precise assessment of the severity and location of any blockages in the arteries. An echocardiogram is used to detect an enlarged heart that could be because of leaky heart valves, heart failure, heart valve disease or congenital heart defects. On the other hand, an electrocardiogram, which records the electrical activity of your heart at rest, is used to either detect high blood pressure or evidence of a previous heart attack. However, it does not show asymptotic blockages in the heart, and it also does not predict the possibility of a future heart attack. Holter monitoring is a test similar to electrocardiogram, except that the heart's rate and rhythm is monitored for longer spans of time such as 12-24 hours.

It can pick up not only your heart's rate and rhythm but also when you feel chest pains or exhibit symptoms of an irregular heartbeat, or arrhythmia [3].

The disease diagnosis process in the medical field can be considered as a decision-making process in which the diagnosis of a new and unknown case is made by medical practitioner from the information that is available from clinical data and from his/her experience in the clinical field. In order to make this decision-making process less costly, easier, faster, more accurate and efficient, the process can be automated.

In today's world, a large majority of the population is suffering from different types of heart problems and the count of patients suffering and dying from these diseases is increasing every day. Thus, there is a need for accurate and early detection of heart disease with proper and adequate treatment which can save the life of many patients. But unfortunately, due to the complicated processes of diagnosis, the different symptoms and pathological tests needed to detect the disease, the correct diagnosis of heart diseases is a difficult task which causes delay in the proper treatment. Hence, there is a need to develop the prediction systems for the tests of heart diseases which can directly help the medical experts in the early and accurate diagnosis of heart disease.

The Artificial Neural Network can be divided into two types based on the training method:

1. Supervised training and
2. Unsupervised training.

Networks that are supervised and require the actual desired output for each input whereas unsupervised networks do not require the desired output for each input. Neural network is an iterative learning process. In this process data cases are presented to the network one at a time, and the weights associated with the input values are adjusted each time. During this learning phase, the network learns by adjusting the weights therefore it can predict the correct class label of input samples. Once a network has been structured for a particular application, it is ready to be trained. To start this process, the initial weights are randomly selected. Then the training or learning, begins [4].

II. LITERATURE SURVEY

The number of systems for prediction of different diseases are proposed and implemented by using different techniques and methods. George et al. have proposed decision support system for dementia patients using support vector machines to define and detect agitation transition. In this system two new SVM architectures are presented, which were applied to the detection of agitation and agitation transition. This approach gives the accuracy of 91.4%, which is higher as compared to 90.9% for the traditional SVM (Support Vector Machines) [5].

Genetic algorithm is also used in another approach by Anbarasi et al. where the number of tests conducted by patient is reduced by determining the attributes that involved in the prediction of heart disease. In this approach 3 classifiers were used and these classifies were fed with reduced attributes, but the system takes more time for model construction.[6]

The automated recognition of obstructive sleep apnea syndrome by using SVM classifier is proposed by Haitham and Alan. In this approach, for the detection of the syndrome they computed the features from the phase and magnitude of the signals obtained from thoracic and abdominal respiratory effort and evaluated the classification of whole night normal and apneic epochs. The highest accuracy of this system is 95% [7].

Gudadhe et al. proposed a system to classify heart disease by using Support Vector Machine and multilayer perceptron neural network architecture. By using Support Vector Machine, the database of heart disease is divided into two classes which shows presence of heart disease or absence of heart disease with an accuracy of 80.41%, whereas the artificial neural network classifies the heart disease data into 5 with an accuracy of 97.5% [8].

Rajkumar and Sophia have proposed the use of Data mining algorithm in diagnosis of heart disease with an accuracy of 52.33%. They have combined the ECG attributes and clinical symptoms to detect the heart disease. The algorithms used by this system are Naive Bayes algorithm, Decision list algorithm and KNN algorithm [9].

Due to the higher accuracy and learning rate the artificial neural network k(ANN) algorithms can also be used in the prediction of heart disease [8]. Kumaravel et al. have proposed automatic diagnosis system for heart diseases using neural networks. In this system ECG data of the patients is used to extract features and 38 input parameters are used to classify 5 major types of heart diseases with accuracy of 63.6 - 82.9% [10].

III. PROPOSED SYSTEM

The fundamental steps that should be followed to apply ANN for the purposes of heart disease diagnosis with sufficient confidence are shown in Fig 1.

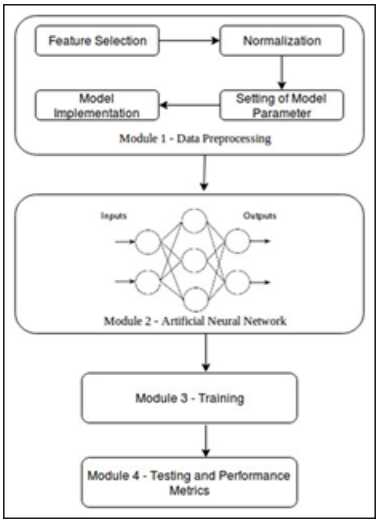

Fig. 1. Modules for Prediction of Test for Heart Disease using Artificial Neural Network

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit the use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. DATA PREPROCESSING

The dataset that has been created has 200 entries and 13 attributes. The attributes that will be used are shown in table 1.

The dataset that has been used is the Cleveland dataset with 13 attributes that have been used for the network.

TABLE I. ATTRIBUTE TABLE

ID	Attribute
1	Age
2	Sex
3	Chest Pain
4	BP in mmhg
5	Serum Cholesterol in mg/dl
6	Fasting blood sugar > 120 mg/dl
7	Rest ECG
8	thalach: maximum heart rate achieved
9	exang: exercise induced angina
10	oldpeak = ST depression induced by exercise relative to rest
11	slope: the slope of the peak exercise ST segment
12	ca: number of major vessels (0-3) colored by fluoroscopy
13	thal: 3 = normal; 6 = fixed defect; 7 = reversible defect

Each of the attributes mentioned above affects the output of the system. The meaning of, and the ranges considered for the attributes are as follows:

- **Chest Pain** records the intensity of pain a person is undergoing. Angina is a type of chest pain caused by reduced blood flow to the heart.
- **BP** stands for Blood Pressure.
- The normal value for it is 120/80 mmhg.
- **Serum Cholesterol** level is a measurement of
- certain elements in the blood including the amount of high and low – density lipoprotein cholesterol. A reading between 200 and 239 mg/dL is considered borderline high and a reading of 240 mg/dL and above is considered high.
- **Fasting blood sugar** indicates the level of sugar in the person's body after fasting. A fasting blood sugar level less than 100 mg/dL (5.6 mmol/L) is normal. A fasting blood sugar level from 100 to 125 mg/dL (5.6 to 6.9 mmol/L) is considered prediabetes. If it's 126 mg/dL (7 mmol/L) or higher on two separate tests, you have diabetes.
- **Rest ECG** records the electrical activity of your heart at rest. For adults 18 and older, a normal resting heart rate is between 60 and 100 beats per minute (bpm), depending on the person's physical condition and age. For children ages 6 to 15, the normal resting heart rate is between 70 and 100 bpm, according to the AHA.
- **Thalach** is the maximum heart rate achieved.
- **Exang** stands for exercise induced angina, indicating the amount of chest pain one has after exercise.
- **Thal** indicates the Thallium Stress Test Result..

B. ARTIFICIAL NEURAL NETWORK

MULTILAYER PERCEPTRON NEURAL NETWORK:

The multilayer perceptron neural network, as its name indicates that it is made up of multiple layers. The single layer perceptron solves only linearly separable problems but many of the complex problems are not linearly separable so to solve such problems one or more layers are added in single layer perceptron hence it is known as multilayer perceptron. Multilayer perceptron network is known as feed-forward neural network having one or more hidden layers as shown. They are generally used for pattern recognition, classification of input patterns, prediction based on the input information and approximation [8].

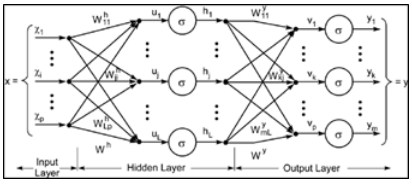

Fig 2. Multilayer Perceptron Network.

The network as shown in Fig 2 has an input layer with three neurons, one hidden layer (in the middle) with three neurons and an output layer with three neurons.

INPUT LAYER - The input layer accepts the input vector (x1...xp) and standardizes the values of each variable in the range of -1 to 1. Then the distribution of these standardized values along with constant input called bias of value 1 is given to each hidden layer neuron by

input layer. This bias value is then multiplied by a weight and added to the sum that is going into the neuron.

HIDDEN LAYER - At each neuron in the hidden layer, a weight (wji) is multiplied to the value from each input neuron. Then a combined value uj is produced by adding the resulting weighted values from each hidden layer neuron. This weighted sum (uj) is then given to the transfer function σ, producing the outputs of value hj . The combined outputs obtained from the hidden layer neurons are then given to the neurons in output layer.

OUTPUT LAYER – At each output layer neuron weight (wkj) is multiplied to the value that is obtained from each hidden layer neuron, and then a combined value vj is produced by adding the resulting weighted values. The weighted sum (vj) is then given to the transfer function σ, which outputs a value yk. The y values are the outputs of the network.

C. TRAINING USING BACK PROPAGATION ALGORITHM

The back-propagation algorithm is a popular algorithm for the training of the neural network. This algorithm is generally used to train multilayer perceptron and many other neural networks. In back-propagation algorithm, the output obtained is compared with the target or expected output and the error is computed. This computed error is then again given to the neural network (fed back or back propagated) and weights are adjusted using this error so that the resulting output will get closer to the target or expected output. This process is repeated for a number of times such that at each iteration the error value gets reduced and the output gets more and more closer to the target or expected output. This process is known as "training" of neural network. 80% of the given dataset is used for training purposes [8].

Rule Based Test Prediction is used for the purpose of predicting the test that the patient should undergo, after the heart disease is predicted. This gives the final output.

D. TESTING

Testing of the accuracy for the system is done using different performance metrics. 20% of the data set is used for testing purposes.

In predictive analytics, a table of confusion, also known as a confusion matrix, is very useful for evaluating classifiers, as they provide an efficient snapshot of its performance displaying the distribution of correct and incorrect instances [4].

TABLE II. CONFUSION MATRIX

		Actual Value		
Prediction		P	N	Total
	p'	TP	FP	p'
	n'	FN	TN	n'
	Total	P	n	

The performance of Multilayer Perceptron Neural Network using back propagation algorithm is evaluated by computing the percentages of Sensitivity (SE), Specificity (SP) and Accuracy (AC), the respective definitions are as follows:

$$SE = TP / (TP + FN) * 100$$

$$SP = TN / (TN + FN) * 100$$

$$AC = (TP+TN) / (TN + TP+FN+FP) * 100$$

Where **TP** is the number of true positives,
TN is the number of true negatives,
FN is the number of false negatives,
FP is the number of false positives.

- True positives are the cases where the system rightly predicts the positive output i.e. when the patient actually has the disease, the system predicts so
- True negatives are the cases where the system rightly predicts the negative output i.e. when the patient has no disease and the system predicts so.
- False positives are the cases where the system wrongly predicts the positive output i.e. the patient has no disease but the system predicts that he/she does.
- False negatives are the cases where the system wrongly predicts the negative output i.e. the patient has a disease but the system predicts that he/she does not.

E. DESIGN CONSIDERATION

In this system, the heart disease database is used to feed the input to the neural network. The network is having 3 layers and feed forward neural network model. The back-propagation learning algorithm with the learning rate of momentum and adaptive learning is used to train the neural network. In the input layer of the network, there are multiple neurons that accept the values of the clinical information from the heart disease database. The hidden layer neurons can be varied in order to reduce error and increase accuracy and the output layer consists of a single neuron that indicates which test needs to be conducted for diagnosis of heart disease [11].

IV. CONCLUSIONS

Prediction of test for heart diseases is very important in the current world because of the number of people dying due to heart disease. Quick prediction of test and fast diagnosis can help save lives. This report concludes that artificial neural network algorithms can be used for predicting test for heart diseases. Certain attributes linked to heart diseases are used to train the system. The trained system then predicts what test the patient should undergo for proper diagnosis of disease.

For future work, the following tasks are under consideration:

- Improving the efficiency of the system by increasing the number of tuples used for training the network.
- Expanding the scope of the project from only heart diseases to multiple other diseases such as lung diseases, neurological disorders, etc.

REFERENCES

1. "Medline plus: Heart diseases,," http://www.nlm.nih.gov/medlineplus /heartdiseases.html
2. A. Rajkumar and G. S. Reena, "Diagnosis of heart disease using data mining algorithm", Global Journal of Computer Science and
3. Technology, vol. 10, pp. 38– 43, December 2010.
4. "A-to-Z Guide to Heart Tests", http://www.secondscount.org/tests/a-to-z-guide-to-heart-tests#.W7TiL2gzbIU.
5. V. Manikantan & S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering (IJACTE), Volume-2, Issue-2, 2013.
6. G. E. Sakr, I. Elhajj, and H. Huijer, "Support vector machines to define and detect agitation transition", IEEE Transactions On Affective Computing, vol. 1, pp. 98–108, December 2010.
7. A. Angari, and A. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier", IEEE Transactions On Information Technology In Biomedicine, vol. 16, pp. 463–468, May 2012.
8. M. Gudadhe, K. Wankhade, and S. Dongre, "Decision support system for heart disease based on support vector machine and artificial neural network", In proceedings of IEEE International Conference on Computer and Communication Technology
9. Jayshril S. Sonawane and D. R. Patil, "Prediction of Heart Disease Using Multilayer
10. Perceptron Neural Networks", ICICES2014 - S.A. Engineering College, Chennai, Tamil Nadu, India, February 2014.
11. N. Kumaravel, K. Sridhar, and N. Nithiyanandam, "Automatic diagnosis of heart diseases using neural networks", In Proceedings of the Fifteenth Biomedical Engineering Conference, pp. 319–322, March 1996.
12. C. L. Huang and C. J. Wang, "A ga-based feature selection and parameter optimization for support vector machines," Expert Systems with Applications, vol. 31, pp. 231–240, October 2006.
13. D. Y. Tsai and S. Watanabe, "Method for optimization of fuzzy reasoning by genetic algorithms and is application to discrimination of myocardial heart disease", In proceedings of IEEE Nuclear Science Symposium and Medica1 Imaging Conference (NSSMIC),pp. 2239–2246, December 1966.
14. S. Ghumbre, C. Patil, and A. Ghatol, "Heart disease diagnosis using support vector machine", In proceedings of International Conference on Computer Science and Information Technology (ICCSIT), pp. 84–88, December 2011.