International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

Neural Network Based Intelligent System for Predicting Heart Disease

K. Subhadra, Vikas B

Abstract*: Heart disease diagnosis has become a difficult task in the field of medicine. This diagnosis depends on a thorough and accurate study of the patient's clinical tests data on the health history of an individual. The tremendous improvement in the field of machine learning aim at developing intelligent automated systems which helps the medical practitioners in predicting as well as making decisions about the disease. Such an automated system for medical diagnosis would enhance timely medical care followed by proper subsequent treatment thereby resulting in significant life saving. Incorporating the techniques of classification in these intelligent systems achieve at accurate diagnosis. Neural Networks has emerged as an important method of classification. Multi-layer Perceptron Neural Network with Back-propagation has been employed as the training algorithm in this work. This paper proposes a diagnostic system for predicting heart disease. For diagnosis of heart disease 14 significant attributes are used in proposed system as per the medical literature. The results tabulated evidently prove that the designed diagnostic system is capable of predicting the risk level of heart disease effectively when compared to other approaches.

Keyword: Neural Network; Perception; Back-Propagation.

I. INTRODUCTION

The most complicated and complex task in the field of medical sciences is the prediction of heart disease. Heart is considered to be the most vital organ of the human body [1]. There is an intense need in predicting the level and seriousness of heart disease that provide an accurate treatment to the patients. Heart disease can be referred to various conditions that lead to abnormal functioning of heart, which may involve blood vessels, arteries etc.

Effective diagnosis of heart disease results in an appropriate treatment to a patient. This requires a deep study of cardiovascular analysis of the patient that includes symptoms such as chest pain chest tightness, chest pressure, and discomfort in breathing, numbness etc [2]. The cardiovascular diagnosis involves certain decisions to be taken based upon the health history and the clinical test results of a person [11]. The process of decision making is a challenging task to the medical practitioners which has to be done accurately and efficiently where a mere negligence may lead to the life risk of a patient.

Manuscript published on 30 March 2019.
*****Correspondence Author(s)
Dr. K. Subhadra, Assistant Professor, GITAM Institute of Technology, GITAM, Visakhapatnam, Andhra Pradesh, India
Vikas B, Assistant Professor,, GITAM Institute of Technology, GITAM, Visakhapatnam, Andhra Pradesh, India

For proper and accurate diagnosis there must be an intelligent automated system that should assist medical practitioners in making decisions based on the current symptoms and medical history of a patient [3, 10].

This paper proposed an intelligent automated system incorporating the techniques of data mining with machine learning in order to make decisions. Medical practitioners are being assisted by the automated systems for providing effective treatment [18]. Data mining techniques involves a combination of statistical methods with machine learning algorithms .Data mining techniques help the system in analyzing the symptoms and machine learning methods help in predicting the disease based on the analysis performed [13,20]. The advantage of this automated system is that it predicts the disease in a less amount of time as well in less cost. Therefore, more research is carried out in the field of machine intelligence to improvise the system for an effective prediction. This paper proposed an intelligent system developed using the concept of Multilayer Perceptron Neural Network with Back propagation algorithm, as a practitioner needs to make a decision from multiple inputs such as current and previous medical history of a patient. Neural networks are proved to be effective in making decisions by predicting the data. As the inputs used in predicting the disease are more in number and diagnosis has to be performed at different stages, Multilayer Perceptron based neural networks are used in this proposed system. Neural Network extends its predictive capability at different hierarchical levels in a multi-layered structure of networks. This multi-layered structure helps in selecting features from the dataset at different scales in order to refine them into more specific features. To facilitate this, the concept of Multi-layer Perceptron Neural Network has been introduced through the implementation of Back-propagation algorithm for efficient diagnosis of heart disease. In this paper, 14 attributes are used as inputs for training the system of neural networks for diagnosing heart disease risk level using multi-layered network.

Traditional diagnosing approaches have no proper automated tools use for the purpose of heart disease diagnostic system. The commonly used data mining algorithms for predicting diseases are:

- Genetic algorithm
- K-means algorithm
- MAFIA algorithm

Several methods proposed the implementation of classification algorithms in diagnosis of heart disease and resulted with an accuracy of 88.33%.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

They used algorithms such as Naive Bayes algorithm, Decision list algorithm and KNN algorithm with the ECG attributes and clinical symptoms to detect the heart disease.
Classification is a two-step process. The first step is a learning step where a classification model is constructed. And a final model is created in the second step for the prediction of class labels for the given data [6].

Decision Tree Induction: A tree like structure is a decision tree, the root node is the topmost node an attribute is tested on internal node, outcome of the test is represented on branch, and a class label is denoted by leaf node [14].

Naïve Bayes Algorithm: This is based on the Bayes theorem. Even though the datasets are larger in size this method can produce very high classification accuracy. This algorithm uses class conditional independence and has ability to learn quickly.

Random Forests: Here each classifier is a decision tree and uses ensemble learning algorithm. The group of classifiers is a forest and individual decision trees are generated by randomly selecting the attributes at each node to split the tree [8,9].

Artificial Neural networks: The idea of ANN is based on the human brain which consists of neurons. An artificial neural network is constructed with the input, output and hidden layers. A transition in a layer is associated with a weight. Input is fed to the input layer and the output of the output layer is compared with the input which gives the error rate of the weights in the neural network [4].

Support Vector Machines: SVM is one of the techniques of classification which uses a supervised learning mechanism. The main aim of SVM is to use a determined hyper plane to separate the two classes. When compared to other algorithms this has a high training speed [7].

Linear Regression: It is a method of fitting the best line to the attributes present. Therefore to predict other attributes one can use every attribute in the dataset.

K-NN: K Nearest Neighbor algorithm is used for classification as well as for regression. Here 'k' is the data set items considered for the classification. Now the required distance is calculated using Euclidian distance or Manhattan distance or any metric as per the user's choice [12].

Deep learning: Data Mining is implemented by Deep learning which is a sub field of machine learning using the artificial neural networks. It works by constructing architectures such as deep neural networks and deep belief networks.

Many works resulted in predicting heart disease effectively but the accuracy is below 90% that specifies that the systems developed so far cannot be used for the accurate heart disease prediction. The proposed system aims at improving the accuracy by implementing multi layered neural networks. The accuracy of the system is presented in the experimental results.

II. METHODOLOGY

The proposed system implements the concept of multi-layered neural networks as neural networks are proved to be effective for practical applications. This system is processed in two phases: in the first phase 14 clinical attributes are fed as input and then the network is trained with training data by back-propagation learning algorithm as mentioned in the Figure 1 below:

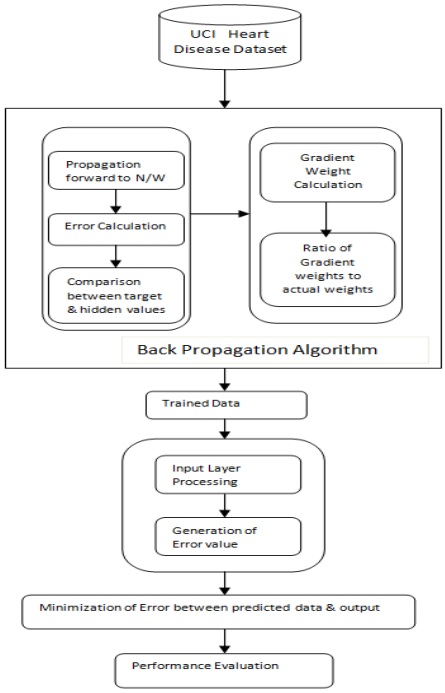


Figure1: Flowchart of the Proposed Prediction system

The performance of the developed system can be evaluated by the following measures.
For each algorithm the sensitivity, specificity, precision [15,16,17,19] and accuracy are observed which are described as follows

Sensitivity:
Sensitivity is the true positive rate and is defined as the number of positive tuples which are correctly classified.

$$\text{Sensitivity} = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ negatives}$$

Specificity:
Specificity is the negative rate is the number of negative tuples that are correctly classified.

$$\text{Specificity} = \frac{\#\ of\ true\ negatives}{\#\ of\ false\ positives + \#\ of\ true\ negatives}$$

Precision:
This is the fraction of true positives in contrast to the overall correct results is calculated.

$$\text{Precision} = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ positives}$$

Accuracy
It is the percentage of the test tuples that are classified properly by any algorithm.

$$\text{Accuracy} = \frac{\#\ of\ true\ positives + \#of\ true\ negatives}{\#\ of\ true\ positives + \ false\ negatives + \ false\ positives + true\ negatives}$$

Data Source
The heart disease data set has been referenced from UCI Repository, and it consists of 303 records, 297 are complete and 6 with missing/unknown values[5].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 1: Heart Disease Dataset

S No	Attribute Name	Attribute Information
1.	#3 (age)	Age of the patient in years.
2.	#4 (sex)	Represented as a binary number. 1 = male. 0 = female.
3.	#9 (cp)	Chest pain type. Values range from 1 to 4. Value 1: typical angina. Value 2: atypical angina. Value 3: non-anginal pain. Value 4: asymptomatic.
4.	#10 (trestbps)	Resting blood pressure measured in mm Hg on admission to the hospital.
5.	#12 (chol)	Serum cholesterol of the patient measured in mg/dl.
6.	#16 (fbs)	Fasting blood sugar of the patient. If greater than 120 mg/dl the attribute value is 1 (true), else the attribute value is 0 (false). Value 1 = true. Value 0 = false.
7.	#19 (restecg)	Resting electrocardiographic results for the patient. This attribute can take 3 integer values 0, 1, or 2. Value 0: normal. Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV). Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria.
8.	#32 (thalach)	Maximum heart rate achieved of the patient.
9.	#38 (exang)	Exercise induced angina. Values can be 0 or 1. Value 1 = yes. Value 0 = no.
10.	#40 (oldpeak)	ST depression induced by exercise relative to rest.
11.	#41 (slope)	Measure of slope for peak exercise. Values can be 1, 2, or 3. Value 1: up sloping. Value 2: flat. Value 3: down sloping.
12.	#44 (ca)	Number of major vessels (0-3) colored by fluoroscopy. Attribute values can be 0 to 3.
13.	#51 (thal)	Represents heart rate of the patient. It can take values 3, 6, or 7. Value 3 = normal. Value 6 = fixed defect. Value 7 = reversable defect
14.	#58 (num)	Contains a numeric value between 0 and 4. Each value represents a heart disease or absence of all of them. Value 0 : < 50% diameter narrowing. (Absence of heart disease). Value 1 to 4: > 50% diameter narrowing. (Presence of different heart diseases).

III. RESULT ANALYSIS

This proposed system classifies the heart disease data into 5 categories diseases with 97.5% accuracy using back-propagation algorithm. In this paper, the prediction system gives the improved result with highest accuracy of 98.58% for 20 neurons in hidden layer with same Cleveland heart disease database. The multi-layered perceptron decision making system embedded with improved algorithm proved to be effective by dividing its training dataset on multiple subsets with 82.8% accuracy with running time of 5.97seconds. The prediction system in this paper gives higher accuracy of 93.39% for 5 neurons in hidden layer with running time of 3.86seconds. The performance of the proposed system is compared with other classification techniques and is tabulated as shown in the result analysis. This shows that the proposed prediction system shows higher performance.

Table 2: Performance analysis after applying various and proposed algorithms on dataset

Performance Metrics	Sensitivity	Specificity	Precision	Accuracy
Decision Tree	75	90.9	87.5	83.6
Logistic Regression	85.7	84.8	82.8	85.2
Naïve Bayes Algorithm	92.9	87.9	86.7	90.2
Random Forests	82.1	87.9	85.2	85.2
Support Vector Machines	74.47	0	76.57	76.57
Generalized Linear Model	85.7	84.8	82.8	85.2
Gradient Boosted Trees	89.3	87.9	86.2	85.2
Deep Learning	92.9	84.8	83.9	88.5
MLPNN-Proposed Algorithm	92	92.5	90	94

The above table, Table 2 depicts the various performance metrics of the classification algorithms Decision tree, Logistic Regression, Naïve Bayes, Random forests, Support Vector Machines, Generalized Liner Model, Gradient Boosted Trees, Deep Learning and MLPNN models on the Heart Disease dataset.

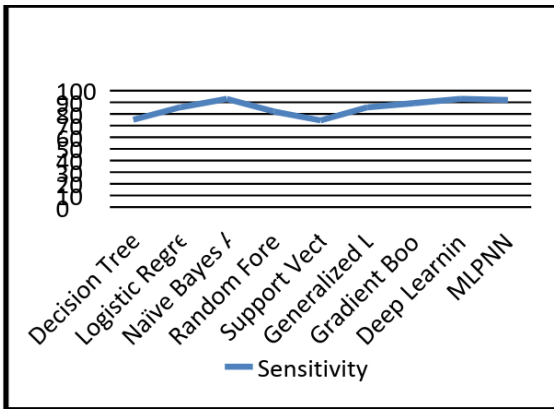
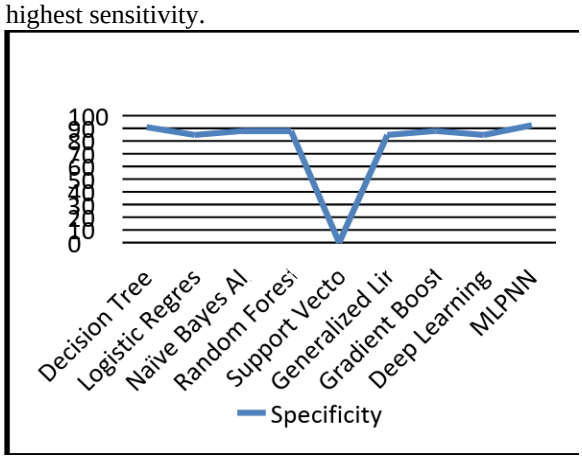
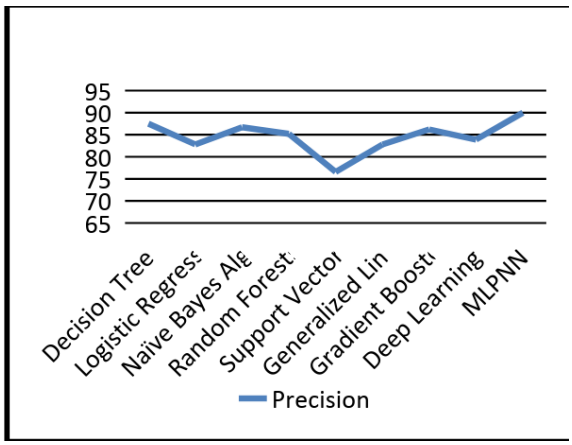


Figure 2: Graphical representation of Sensitivity

The above graph, Figure 2 depicts that Naïve Bayes, Random forests Deep Learning and MLPNN models have the highest sensitivity.



The above graph, Figure 3 depicts that the MLPNN has the highest specificity when compared to other classifiers.



Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The above graph, Figure 4 depicts that the algorithms Decision tree, Naïve Bayes, Gradient Boosted Trees and MLPNN models classifiers have the highest precision.

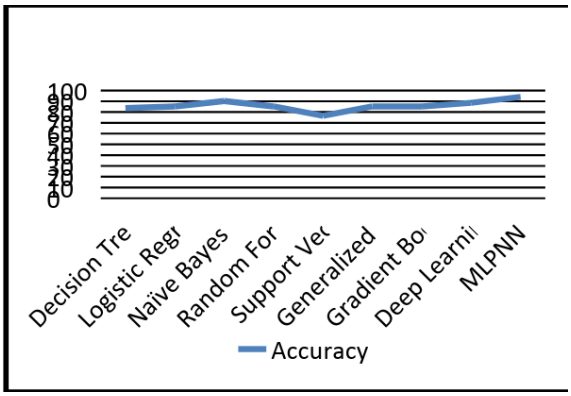


Figure 5: Graphical representation of Accuracy

The above graph, Figure 5 depicts that the Naive Bayes and MLPNN have the highest accuracy when compared to the other algorithms.

IV. CONCLUSION

The proposed system of heart disease prediction with appropriate diagnosis has been framed up using Multilayer Perceptron Neural Network. For effective prediction, back propagation algorithm was applied to train the data and compare the parameters iteratively. The propagation algorithm has been repeated until minimum error rate was observed. And it is quite evident from the results presented in the previous section that the accuracy rate is maximized. It is proven from the results that the proposed method effectively predicts the heart disease through the 14 attributes when compared to the other approaches. This work can be extended to predict and analyze the level of the disease by considering more attributes.

REFERENCES

1. Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888)Volume 47–No.10, June 2012.
2. K.Srinivas, Dr.G.Raghavendra Rao, Dr. A.Govardhan,"Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010.
3. R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Syst. Appl., vol.36, no. 4, pp. 7675–7680, 2009.
4. Yanwei Xing, Jie Wang and Zhihong Zhao, "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease", 2007 International Conference on Convergence Information Technology.
5. UCI Machine Learning Repository [homepage on the Internet]. Arlington: The Association; 2006 [updated 1996 Dec 3; cited 2011 Feb 2]. Available from:http://archive.ics.uci.edu/ml/datasets/Heart+Disease
6. Jiawei Han, Micheline Kamber & Jian Pei-Data Mining: Concepts and Techniques; 3rd ed; 2011.
7. Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System", World Journal of Engineering and Technology, 2018, 6, 854-873.
8. Hazra, A., Mandal, S., Gupta, A. and Mukherjee, A. (2017) Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. Advances in Computational Sciences and Technology, 10, 2137-2159.
9. Patel, J., Upadhyay, P. and Patel, D. (2016) Heart Disease Prediction Using Machine learning and Data Mining Technique. Journals of Computer Science & Electronics, 7, 129-137.
10. Chavan Patil, A.B. and Sonawane, P. (2017) To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients. International Journal on Emerging Trends in Technology (IJETT), 4, 8274-8281
11. Weng, S.F., Reps, J., Kai, J., Garibaldi, J.M. and Qureshi, N. (2017) Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data? PLoS ONE, 12, e0174944. https://doi.org/10.1371/journal.pone.0174944.
12. Soni, J., Ansari, U. and Sharma, D. (2011) Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. International Journal on Computer Science and Engineering (IJCSE), 3, 2385-2392.
13. Shouman, M., Turner, T. and Stocker, R. (2012) Using Data Mining Techniques in Heart Disease Diagnosis and Treatment. Electronics, Communications, and Computers, Alexandria, 173-177.
14. Masethe, H. and Masethe, M. (2014) Prediction of Heart Disease Using Classification Algorithms. Proceedings of the World Congress on Engineering and Computer Science, San Francisco, 809-812
15. K.Subhadra, M.Shashi,'Effectiveness of Ant Colony Optimization with Hybrid distance for document Retrieval', IJARCS,Volume 8,2017.
16. K.Subhadra, M.Shashi 'An Efficient Relevance-Based Document Ranking Clustering System for Conjunctive Queries', International Journal of Advanced Computing, ISSN: 2051-0845, Vol.47, Issue.1 1446.
17. K.Subhadra, M.Shashi]'Hybrid Distance Based Document Clustering with Keyword and Phrase Indexing', IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
18. Vikas B, B.S.Anuhya, K Santosh Bhargav, Sipra Sarangi, Manaswini Chilla. (2017, June). Application of the Apriori Algorithm for Prediction of Polycystic Ovarian Syndrome (PCOS). 4th International Conference on Information System Design And Intelligent Applications, 2017
19. K. Santosh Bhargav, Dola Sai Siva Bhaskar Thota, T. Divya Kumari, Vikas B. (2018). Application of Machine Learning Classification Algorithms on Hepatitis Dataset.International Journal of Applied Engineering Research,13(16), 12732-12737
20. Vikas B, Yaswanth D.V.S., Vinay W., Sridhar Reddy B., Saranyu A.V.H. (2017). Classification of Hepatitis C Virus Using Case-Based Reasoning (CBR) with Correlation Lift Metric. 4th International Conference on Information System Design And Intelligent Applications, 2017

AUTHORS PROFILE

Dr. K. Subhadra has received Doctorate from JNTUniversity, Hyderabad in 2017 and Masters from Andhra University. She is actively engaged in research in the area of Big Data Analytics and Algorithms. She has a teaching experience of about 13 years.

Vikas B received the Bachelor in IT degree from the JNTUH, Hyderabad, in 2010 and the Master in Bioinformatics degree from the JNTUH, Hyderabad, in 2012. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, GITAM (Deemed to be University), Visakhapatnam. He is member of IAENG and IEEE-CS and his research interests include Data mining, Bioinformatics, Information Security, and Data Sciences. He has a teaching experience of about 9 years.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)