A Data Mining Approach for Prediction of Heart Disease Using Neural Networks

Article · November 2012

2 authors:

Chaitrali S. Dangare
Bharat Institute of Engineering and Technology
6 PUBLICATIONS **84** CITATIONS
SEE PROFILE

Sulabha Apte
Solapur University
1 PUBLICATION **30** CITATIONS
SEE PROFILE

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET)

ISSN 0976 – 6367(Print)
ISSN 0976 – 6375(Online)
Volume 3, Issue 3, October - December (2012), pp. 30-40
© IAEME: www.iaeme.com/ijcet.asp
Journal Impact Factor (2012): 3.9580 (Calculated by GISI)
www.jifactor.com

IJCET
© I A E M E

A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS

Miss. Chaitrali S. Dangare[1], Dr. Mrs. Sulabha S. Apte[2]

[1] Student M.E. (CSE) , Walchand Institute of Technology, Computer Science and Engineering Department, Solapur, Maharashtra- 413006, India
chaitrali.dangare@gmail.com

[2] Professor, Walchand Institute of Technology, Computer Science and Engineering Department, Solapur, Maharashtra- 413006, India
headcsewit@gmail.com

ABSTRACT

Heart disease diagnosis is a complex task which requires much experience and knowledge. Traditional way of predicting Heart disease is doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Nowadays, Health care industry contains huge amount of heath care data, which contains hidden information. This hidden information is useful for making effective decisions. Computer based information along with advanced Data mining techniques are used for appropriate results. Neural network is widely used tool for predicting Heart disease diagnosis. In this research paper, a Heart Disease Prediction system (HDPS) is developed using Neural network. The HDPS system predicts the likelihood of patient getting a Heart disease. For prediction, the system uses sex, blood pressure, cholesterol like 13 medical parameters. Here two more parameters are added i.e. obesity and smoking for better accuracy. From the results, it has been seen that neural network predict heart disease with nearly 100% accuracy.

Keywords: Backpropagation, Data mining, Heart disease, Multilayer perceptron neural network, Neural Network.

1. INTRODUCTION

Heart is the important part of our body. Life is itself dependent on efficient working of heart. If operation of heart is not proper, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that affects on the operation of heart. There are number of factors which increases risk of Heart disease.

Some of them are listed below:

- Family history of heart disease
- Smoking
- Cholesterol
- high blood pressure
- Obesity
- Lack of physical exercise

Nowadays, in the world Heart disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. In 2008, 17.3 million people died due to Heart Disease. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease.

Predication should be done to reduce risk of Heart disease. Diagnosis is usually based on signs, symptoms and physical examination of a patient. Almost all the doctors are predicting heart disease by learning and experience. The diagnosis of disease is a difficult and tedious task in medical field. Predicting Heart disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. 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. Only human intelligence alone is not enough for proper diagnosis. A number of difficulties will arrive during diagnosis, such as less accurate results, less experience, time dependent performance, knowledge up gradation is difficult etc listed in Fig.1.

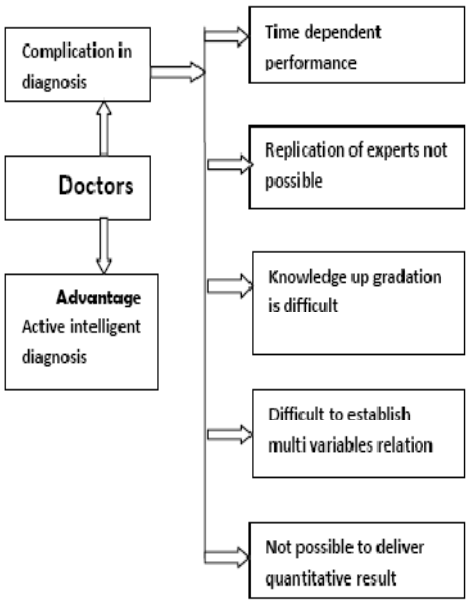

Figure 1: Complexity in diagnosis with doctor [4].

In order to improve the accuracy of diagnosis and to reduce the diagnosis time, we have developed an efficient and reliable Decision Support System for Heart Disease using Neural networks.

2. LITERATURE REVIEW

Work done in heart disease diagnosis using data mining and artificial neural networks are discussed below:

- Sellappan Palaniappan et al. [7] developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. IHDPS can answer complex "what if" queries which traditional decision support systems cannot. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. IHDPS is Web-based, user-friendly, scalable, reliable and expandable. It is implemented on the .NET platform.
- Dilip Roy Chowdhury et al. [1] represent the use of artificial neural networks in predicting neonatal disease diagnosis. The proposed technique involves training a Multi Layer Perceptron with a BP learning algorithm to recognize a pattern for the diagnosing and prediction of neonatal diseases. The Backpropogation algorithm was used to train the ANN architecture and the same has been tested for the various categories of neonatal disease. About 94 cases of different sign and symptoms parameter have been tested in this model. This study exhibits ANN based prediction of neonatal disease and improves the diagnosis accuracy of 75% with higher stability.
- Milan Kumari et al. [3], proposed research contains data mining classification techniques RIPPER classifier, Decision Tree, Artificial neural networks (ANNs), and Support Vector Machine (SVM) are analyzed on cardiovascular disease dataset. Performance of these techniques is compared through sensitivity, specificity, accuracy, error rate, True Positive Rate and False Positive Rate. 10-fold cross validation method was used to measure the unbiased estimate of these prediction models. As per our results error rates for RIPPER, Decision Tree, ANN and SVM are 2.756, 0.2755, 0.2248 and 0.1588 respectively. Accuracy of RIPPER, Decision Tree, ANN and SVM are 81.08%, 79.05%, 80.06% and 84.12% respectively. The analysis shows that out of these four classification models SVM predicts cardiovascular disease with least error rate and highest accuracy.
- Vanisree K et al. [2] , has been proposed a Decision Support System for diagnosis of Congenital Heart Disease. The proposed system is designed and developed by using MATLAB's GUI feature with the implementation of Backpropagation Neural Network. The Backpropagation Neural Network used in this study is a multi layered Feed Forward Neural Network, which is trained by a supervised Delta Learning Rule. The dataset used in this study are the signs, symptoms and the results of physical evaluation of a patient. The proposed system achieved an accuracy of 90%.
- Niti Guru et al. [6] proposed a system that uses neural network for prediction of heart disease, blood pressure and sugar. A set of 78 records with 13 attributes are used for training and testing. He suggested supervised network for diagnosis of heart disease and trained it using back propagation algorithm. On the basis of unknown data is entered by

doctor the system will find that unknown data from training data and generate list of possible disease from which patient can suffer.

- A proficient methodology for the extraction of significant patterns from the heart disease warehouses for heart attack prediction has been presented by Shantakumar B.Patil et al [5]. Initially, the data warehouse is pre-processed in order to make it suitable for the mining process. Once the preprocessing gets over, the heart disease warehouse is clustered with the aid of the K-means clustering algorithm. Consequently the frequent patterns applicable to heart disease are mined with the aid of the MAFIA algorithm from the data extracted. In addition, the patterns vital to heart attack prediction are selected on basis of the computed significant weightage. The neural network is trained with the selected significant patterns for the effective prediction of heart attack.

3. NEURAL NETWORKS IN DATA MINING

A neural network (NN) is a parallel, distributed information processing structure consisting of multiple numbers of processing elements called node, they are interconnected via unidirectional signal channels called connections. Each processing element has a single output connection that branches into many connections; each carries the same signal i.e. the processing element output signal. The NN can be classified in two main groups according to the way they learn,

1. Supervised learning:

   It is a simple model, in which the networks compute a response to each input and then compare it with target value. If the computed response differs from target value, the weights of the network are adapted according to a learning rule.

e.g.: Single-layer perceptron, Multi-layer perceptron.

2. Unsupervised learning:

   These networks learn by identifying special features in the problems they are exposed to.

e.g.: Self-organizing feature maps.

Neural network has following properties:

- Nonlinearity
- Learning ability
- Input-output mapping
- Adaptivity
- Evidential response
- Fault tolerance
- Neurological analogy

In medical field, decision making is done by neural network because they provide more accurate results.

4. NEURAL NETWORKS BASED HEART DISEASE PREDICTION SYSTEM

In this research paper, decision support system is developed for predicting heart disease of a patient. The prediction is done based on historical heart disease database. The system uses 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 is used, as these attributes are considered as important attributes for heart disease. The technique used to develop system is Multilayer Perceptron Neural Network (MLPNN) with Backpropagation algorithm (BP).

4.1 Multilayer Perceptron Neural Network (MLPNN)

One of the most important models in Artificial Neural Network is Multilayer Perceptron (MLP). The type of architecture used to implement the system is Multilayer Perceptron Neural Network (MLPNN).

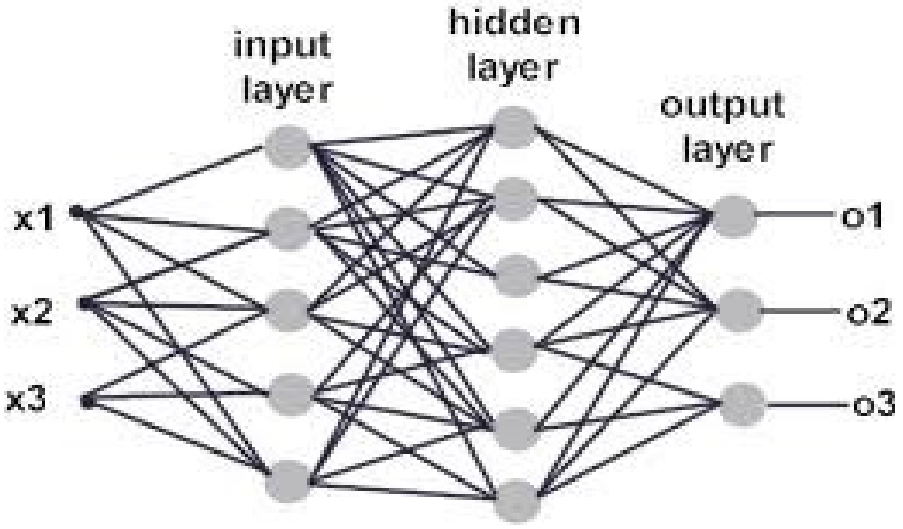

Figure 2: Multilayer Perceptron Neural Network (MLPNN)

The MLPNN consists of one input layer, one output layer and one or more hidden layers. Each layer consists of one or more nodes, represented by small circles. The lines between nodes indicate flow of information from one node to another node. The input layer receives signals from external nodes. The output of input layer is given to hidden layer, through weighted connection links. It performs computations and transmits the result to output layer through weighted links. The output of hidden layer is forwarded to output layer, it performs computations and produce final result.

The working of multilayer perceptron neural network is summarized in steps as mentioned below:

Steps:

1. Input data is provided to input layer for processing, which produces a predicted output.
2. The predicted output is subtracted from actual output and error value is calculated.
3. The network then uses a Backpropagation algorithm which adjusts the weights.
4. For weights adjusting it starts from weights between output layer nodes and last hidden layer nodes and works backwards through network.
5. When back propagation is finished, the forwarding process starts again.
6. The process is repeated until the error between predicted and actual output is minimized.

4.2 Backpropagation network

The most widely used training algorithm for multilayer and feed forward network is Backpropagation. The name given is back propagation because, it calculates the difference between actual and predicated values is propagated from output nodes backwards to nodes in previous layer. This is done to improve weights during processing.

The working of Backpropagation algorithm is summarized in steps as follows:

Steps:

1. Provide training data to network.
2. Compare the actual and desired output.
3. Calculate the error in each neuron.
4. Calculate what output should be for each neuron and how much lower or higher output must be adjusted for desired output.
5. Then adjust the weights.

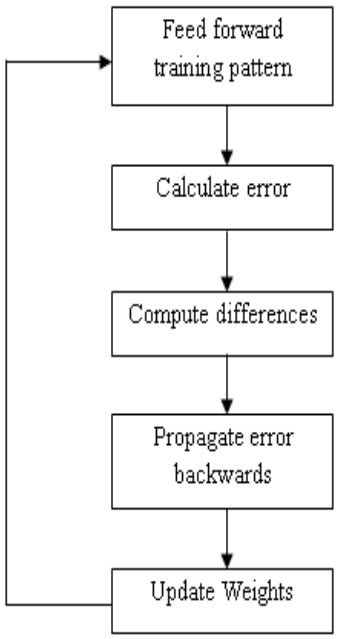

Figure 3: Back-propagation Neural network

5. Results

The experiment is carried out on a publicly available database for heart disease. The dataset contains total 573 records. The dataset is divided into 2 sets training (303 records) and testing set (270 records). A data mining tool Weka 3.6.6 is used for experiment. Parameters used for experiment are listed below.

PatientID: Patient Identification number.

Diagnosis: Value 1: =<50% (no heart disease)
Value 0: >50% (has heart disease)

The remaining parameters are listed out in tabular form as below.

Table 1: Description of 13 parameters used

Sr. no	Attribute	Description	Values
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 angina 3 = non-angina 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 angina	0= no 1 = yes
10	oldpeak	ST depression induced by exercise relative to rest	Continuous value
11	slope	Slope of the peak exercise ST segment	1 = unsloping 2 = flat 3 = downsloping
12	ca	Number of major vessels colored by floursopy	0-3 value
13	thal	Defect type	3 = normal 6 = fixed 7 = reversible defect

All the research papers referred above have used 13 input attributes for prediction of Heart disease. For getting more accurate results 2 more parameters are used i.e. obesity and smoking.

Table 2: Description of newly added parameters

Sr. no	Attribute	Description	Values
14	obes	obesity	1 = yes 0 = no
15	smoke	smoking	1= past 2 = current 3 = never

After applying neural networks on training dataset the results obtained is shown in the form of two dimensional matrixes, the matrix is known as confusion matrix. The confusion matrix is easy to understand as the correct and incorrect classification is displayed in the table. The confusion matrix for a two class classifier is shown in Table 3.

Class a: YES (has heart disease)

Class b: No (has no heart disease)

Table 3: A confusion matrix

	a (has heart disease)	b (no heart disease)
a (has heart disease)	TP	FN
b (no heart disease)	FP	TN

TP (True Positive): It denotes the number of records classified as true while they were actually true.

FN (False Negative): It denotes the number of records classified as false while they were actually true.

FP (False Positive): It denotes the number of records classified as true while they were actually false.

TN (True Negative): It denotes the number of records classified as false while they were actually false.

The following table shows results obtained with 13 and 15 parameters.

Table 4: Results for Neural networks

- **With 13 parameters**

	a	B
a	117	0
b	2	151

- **With 15 parameters**

	a	B
a	106	0
b	0	164

The following table shows comparison of accuracies obtained with 13 and 15 parameters.

Table 5: Comparison of accuracies

Classification Techniques	Accuracy with	
	13 attributes	15 attributes
Neural Networks	99.25	100

The results obtained in table 5 are plotted on a graph as shown below. The x-axis represents number of parameters used for prediction and y-axis represents accuracy in terms of percentage. There are two bars, which shows accuracy for 13 parameters and 15 parameters.

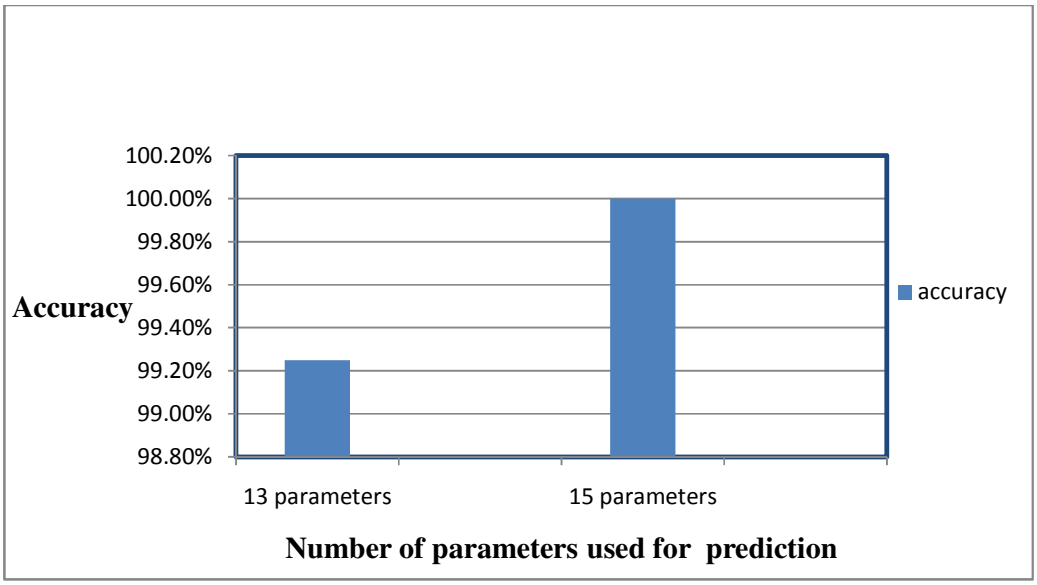

Figure 4: Graph shows accuracy for 13 and 15 parameters

6. CONCLUSION

In this research paper, we have presented Heart disease prediction system (HDPS) using data mining and artificial neural network (ANN) techniques. From the ANN, a multilayer perceptron neural network along with back propagation algorithm is used to develop the system. Because MLPNN model proves the better results and helps the domain experts and even person related with the field to plan for a better diagnose and provide the patient with early diagnosis results as it performs realistically well even without retraining. The experimental result shows that using neural networks the system predicts Heart disease with nearly 100% accuracy.

REFERENCES

[1] Dilip Roy Chowdhury, Mridula Chatterjee & R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011.

[2] Vanisree K, Jyothi Singaraju, Decision Support System for Congenital Heart Disease Diagnosis based on Signs and Symptoms using Neural Networks, International Journal of Computer Applications (0975 – 8887) Volume 19– No.6, April 2011.

[3] Milan Kumari, Sunila Godara, Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, IJCST Vol. 2, Iss ue 2, June 2011.

[4] K.S.Kavitha, K.V.Ramakrishnan, Manoj Kumar Singh, Modeling and design of evolutionary neural network for heart disease detection, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 ISSN (Online): 1694-0814

[5] Shantakumar B.Patil, Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research, ISSN 1450-216X, Vol.31 No.4 (2009), pp.642-656.

[6] Niti Guru, Anil Dahiya, Navin Rajpal, Decision Support System for Heart Disease Diagnosis Using Neural Network, Delhi Business Review, Vol. 8, No. 1, January-June 2007.

[7] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Technique, 978-1-4244-1968-5/08/$25.00 ©2008 IEEE.

[8] Cleveland database: http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[9] Statlog database: http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/