iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** IV **Month of publication:** April 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

A Review-Heart Disease Prediction using Data Mining Technique

Nidhi Sharma 1, Purushottam Sharma 2

1 Dept. of IT
Amity University
Uttar Pradesh

2 Dept. of IT (Faculty)
Amity University
Uttar Pradesh

Abstract: ***Heart disease diagnosis is a intricate task which requires much familiarity and knowledge. The old way of predicting Heart disease is doctor's assessment or number of medical tests such as Heart MRI ,ECG and Stress Test etc. The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still "information rich" but "knowledge poor". There is a bulk of data available within the healthcare systems. However, there is a lack of efficient analysis tools to discover unseen relationships and trends in data. 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 obtained, it has been seen that neural network predicts with nearly 100% accuracy.***

General Terms: Data mining, artificial neural networks, back propagation networks, multi layer neural network.

Keywords: Data Mining, Neural network,

INTRODUCTION

The heart is important organ of human body part. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it. A number of factors have been shown that increases the risk of Heart disease [1]:

- Family history
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

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

ISSN: 2321-9653

Heart diseases. In 2008, 17.3 million people died due to 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. 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.

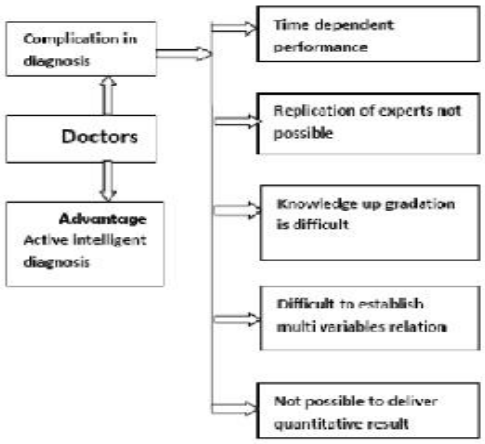


Figure 1: Complexity in diagnosis with doctor [2].

LITERATURE SURVEY

To develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) a novel technique was proposed by Heon Gyu Lee et al. [3]. To achieve this, they have used several classifiers e.g. Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine).

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [4]. 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.

An Intelligent Heart Disease Prediction System (IHDPS) is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by Sellappan Palaniappan et al .[5]. 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

Franck Le Duff et al. [6] builds a decision tree with database of patient for a medical problem.

Latha Parthiban et al. [7] 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.

Kiyong Noh et al. [8] 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.

NEURAL NETWORKS IN DATA MINING

Neural Networks are an information processing technique based on the way biological nervous systems, such as the brain, process information. They resemble the human brain in the following two ways:

A neural network acquires knowledge through learning.

A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights.

The NN can be classified in two main categories according to the way they learn:

1. Supervised learning:

Supervised learning networks have been the mainstream of neural model development. The training data consist of pairs of input/output training patterns. Therefore, the learning will benefit from the assistance of the teacher. e.g.: Single-layer perceptron.

2. Unsupervised learning:

For an unsupervised learning rule, the training set consist of input training patterns only. Therefore, the network is trained without any teacher. The network adapt based on the experiences collected through the previous training patterns.

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.

HEART DISEASE PREDICTION SYSTEM BASED ON NEURAL NETWORKS

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).

Multilayer Perceptron Neural Network

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable.[9]

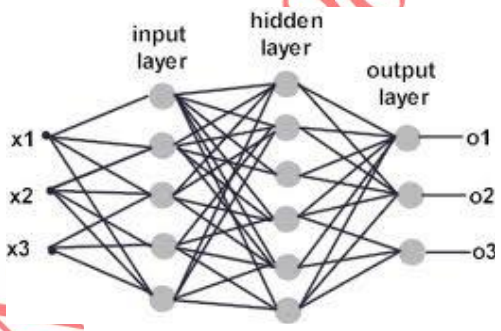


Figure 2: Multilayer Perceptron Neural Network

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

Input data is provided to input layer for processing, which produces a predicted output

The predicted output is subtracted from actual output and error value is calculated.

The network then uses a Back propagation algorithm which adjusts the weights.

For weights adjusting it starts from weights between output layer nodes and last hidden layer nodes and works backwards through network.

When back propagation is finished, the forwarding process starts again

The process is repeated until the error between predicted and actual output is minimized.

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 Back propagation 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

RESULTS OF THE EXPERIMENTS

CONDUCTED

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.

PID: Patient Identification number.

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

Value 0: >50% (has heart disease)

Table 1: Description of 13 parameters used

Sr.no	Attribute	Description	Values
1	Gender	M or F	0=Male 1=Female
2	Age	In Years	Continuos
3	chol	Serum Cholestrol	Continous value in mm/dl
4	RER	Resting Elecrtrograph Results	0=normal 1=Abnormal wave 2=left ventricular
5	CP	Chest pain Type	1 = typical type 1 2 = typical type angina
6	EIA	Exercise Induced Angina	0=No 1=yes
7	Thalach	Max heart Rate	Continous Value
8	Slope	Slope of the peak exercise ST segments	1 = unsloping 2 = flat
9	CA	Number of major vessels colored by	0-3 value
10	Thal	Defect type	3 = normal 6 = fixed
11	FBS	Fasting Blood Sugar	1 >= 120 mg/dl 0 <= 120 mg/dl
12	Old peak	ST depression induced by exercise	Continuous value
13	thestbps	Resting blood pressure	Continuous value in mm hg

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	Smoke	Smoking	1= past 2 = current 3 = never
15	Obes	Obesity	1 = yes 0 = no

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

Class	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[10]

With 13 parameters

Class	A	b
a	117	0
b	2	151

With 15 parameters

Class	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[10]

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.

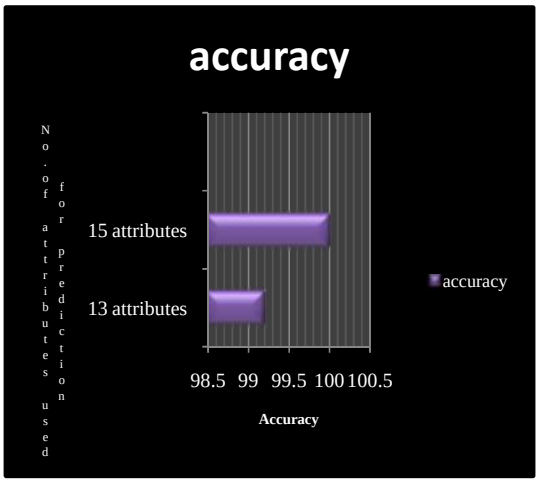


Figure 3: Graph shows accuracy for 13 and 15 parameters

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] Yanwei, X.; Wang, J.; Zhao, Z.; Gao, Y., "Combination data mining models with new medical data to predict outcome of coronary heart disease". Proceedings International Conference on Convergence Information Technology 2007, pp. 868 – 872

[2] 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

[3] Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: [1] [1]

[4] Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

[5] 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).

[6] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008

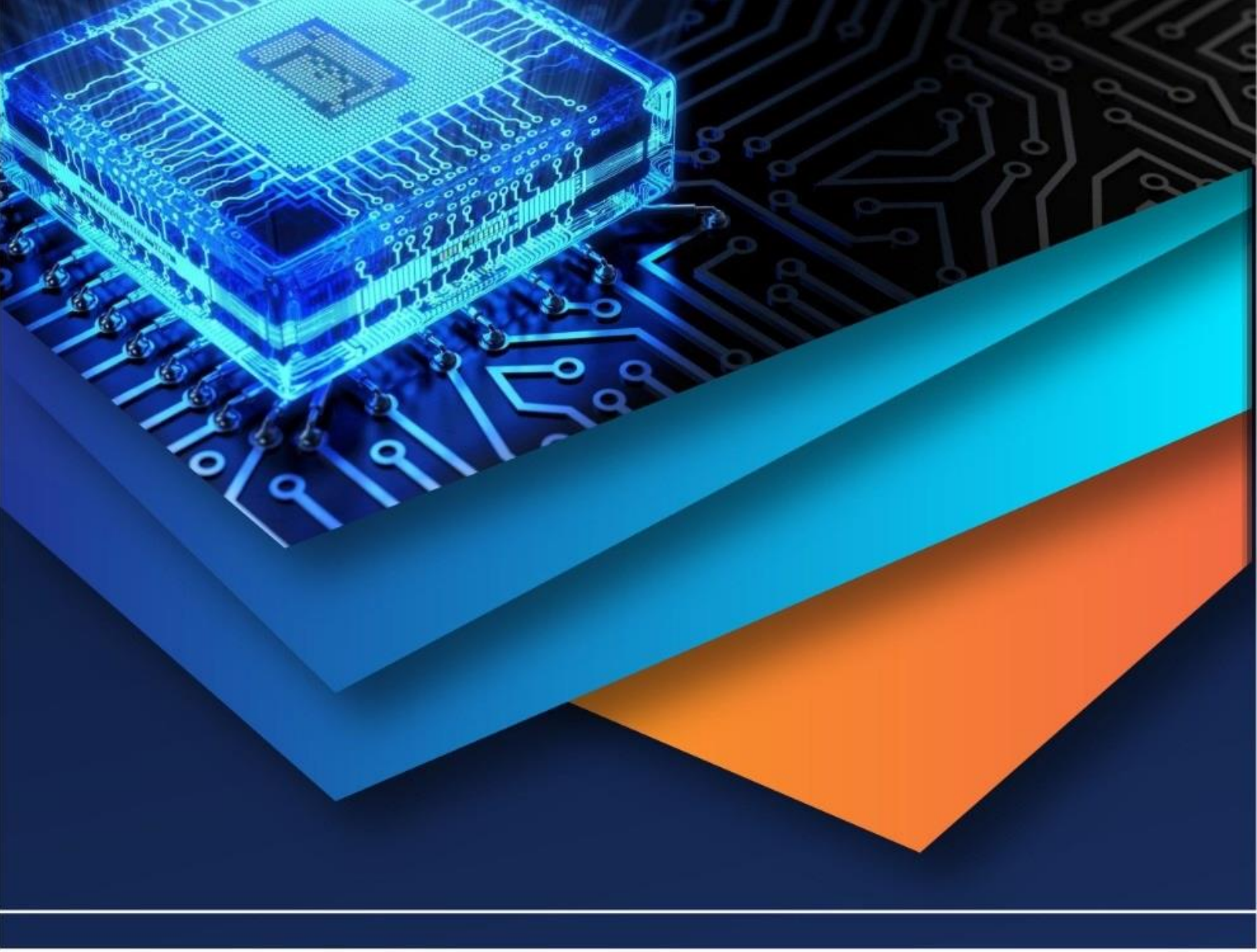
[7] Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 1256-9, 2004.

[8] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, 2008.

[9] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006.

[10] http://en.wikipedia.org/wiki/Multilayer_perceptron

[11] Chaitrali S. Dangare1, Dr. Mrs. Sulabha S. Apte, "Data mining approach for prediction of heart disease using neural networks" International Journal Of Computer Engineering & Technology, Volume 3, Issue 3, October - December (2012),



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY