DIAGNOSIS OF HEART DISEASE USING NEURAL NETWORK APPROACH

[1]A. T. SAYAD, [2]P. P. HALKARNIKAR

[1]M. Tech, Fourth Semester, Department of Technology, Shivaji University, Kolhapur, India
[2]Department of Computer Science and Engineering, D. Y. Patil College of Engineering, Kolhapur, Shivaji University, Kolhapur, India
Email: [1]asma_t_sayad@yahoo.com, [2]pp_halkarnikar@rediffmail.com

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 is one of the important issue 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. Disease 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 helps for efficient classification of given data. This research work is the implementation of heart disease diagnostic system. For diagnosis of heart disease significantly 13 attributes are used in proposed system as per the medical literature.I have employed the Multi-layer Perceptron Neural Network with Back-propagation as the training algorithm. The results thus obtained have illustrated that the designed diagnostic system is capable of predicting the risk level of heart disease effectively. This system acts as promising tool for diagnosis of heart disease.

Keywords— Data mining, Disease Diagnosis, Neural Network, Multilayer Perceptron Neural Network and back propagation algorithm

I. INTRODUCTION

A major challenge facing healthcare organizations (hospitals, medical centres) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and decision support systems.

Hospitals and clinics accumulate a huge amount of patient data over the years. These data provide a basis for the analysis of risk factors for many diseases. For example, we can predict the level of heart attack to find patterns associated with heart disease. One of the major topics in data mining research is the discovery of interesting patterns in data. From the introduction frequency thresholds only common knowledge is revealed, while at low thresholds prohibitively many patterns are returned. A majority of areas related to medical services such as prediction of effectiveness of surgical procedures, medical tests, medication and the discovery of relationships among clinical and diagnosis data effectiveness of medical treatments can be estimated by developing the data mining applications.

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. Diagnosis of Heart disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. 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 as shown in Fig. I.

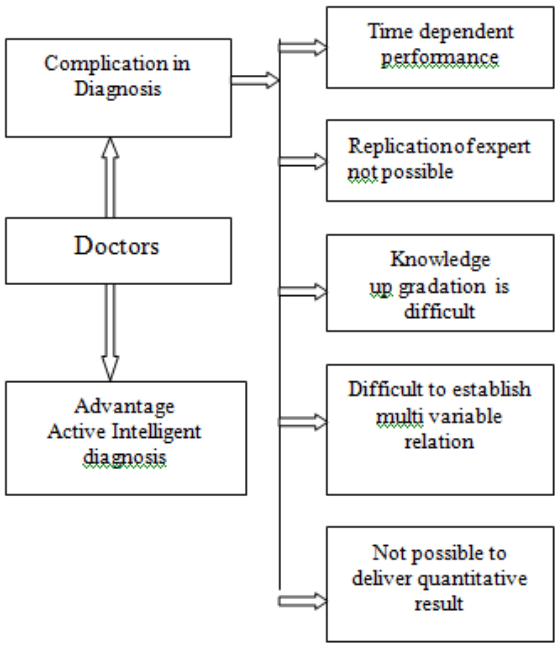

Fig. I Complexity in diagnosis with doctors.

Data mining is capable of delivering an analysis of which courses of action prove effective, achieved by comparing and contrasting causes, symptoms, and courses of treatments. In this paper, I have presented diagnosis of heart disease from the heart disease database. The heart disease database consists of mixed attributes containing both the numerical and categorical data. These records are cleaned and filtered with the intention that the irrelevant data from the database is processed before further processing. This system can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. Results show proposed system has its unique strength in realizing the objectives of the defined mining goals.

II. RELATED W ORK

Tremendous works in literature related with heart disease diagnosis using data mining techniques have motivated our work. The researchers in the medical field diagnose and predict the diseases in addition to providing effective care for patients by employing the data mining techniques. The data mining techniques have been employed by numerous works in the literature to diagnose diverse diseases, for instance: Diabetes, Hepatitis, Cancer, Heart diseases and more [4]. A model Intelligent Heart Disease Prediction System (IHDPS) built with the aid of data mining techniques like Decision Trees, Naïve Bayes and Neural Network was proposed by Sellappan Palaniappan, Rafiah Awang. The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez. The assessed data set encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing be removed before mining process occurs. Association rule mining is a major data mining technique, and is a most commonly used pattern discovery method. It retrieves all frequent patterns in a data set and forms interesting rules among frequent patterns. Most frequently used association rule mining methods are Apriority and Growth. Frequent Itemset Mining (FIM) is considered to be one of the elemental data mining problems that intends to discover groups of items or values or patterns that co-occur frequently in a dataset. The term Heart disease encompasses the diverse diseases that affect the heart. Heart disease was then major cause of casualties in the United States, England, Canada and Wales as in 2007. Heart disease kills one person every 34 seconds in the United States. Coronary heart disease, Cardiomyopathy and Cardiovascular disease are some categories of heart diseases. The term “cardiovascular disease” includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in severe illness, disability, and death [10].

III. ANN IN MEDICAL FIELD

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.

The advantages of Neural Networks for classification are:

1) Neural Networks are more robust because of the weights

2) The Neural Networks improves its performance by learning. This may continue even after the training set has been applied.

3) The use of Neural Networks can be parallelized as specified above for better performance.

4) There is a low error rate and thus a high degree of accuracy once the appropriate training has been performed.

5) Neural Networks are more robust in noisy environment

IV. PROBLEM STATEMENT

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like “What is the average age of patients who have heart disease?”, “How many surgeries had resulted in hospital stays longer than 10 days?”, “Identify the female patients who are single, above 30 years old, and who have been treated for cancer.” However, they cannot answer

complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease." Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Hian Chye Koh and Gerald Tan [2] proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation and improve patient outcome. This suggestion is promising as data modelling and analysis tools e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

V. OBJECTIVES

The main objective of this project is to develop a system for diagnosis of Heart disease using data mining modelling technique. This system can discover and extract hidden knowledge associated with heart disease from historical heart disease database. It assists healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. Here, system can classify the patients based on risk level using data mining technique into two classes as Low, High.

So, these diagnosis helps healthcare practitioners in taking positive and effective decision in following condition:

1) Patient at high risk requires lowering treatment and special attention.

2) Patient at low risk and no risk may be encouraged by their doctor to follow Health Recommendation for prevention of heart disease.

So, the proposed system significantly acts as ***decision support system*** for doctors in taking clinical decisions.

VI. PROBLEM FORMULATION

A. *Data Collection*

The data is collected from Cleveland Clinic Foundation This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them as shown in Table I.

Table1. Attribute Description

ID	Attribute Name	Attribute Domain Values
1	Age(In Years)	Age of patient in years. 1: >=30 & <70 2: >=70
2	Sex	1: male, 0: female.
3	CP(Chest Pain)	chest pain type Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic
4	Trestbps (resting blood pressure)	Resting blood pressure (in mm/ Hg on admission to the hospital).
5	Chol(Cholesterol)	Serum cholesterol in mg/dl (vh: very high; h: high)
6	Fbs(Fasting Blood Sugar)	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
7	Restecg(Resting ECG result)	Resting electrocardiographic results. 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	Thalach(Maximum heart rate)	Maximum Heart Rate achieved. (ab: abnormal, norm: normal)
9	Exang(Exercise induced angina)	Exercise induced angina (1 = yes; 0 = no).
10	Slope(The slope of ST segment)	The slope of the peak exercise ST segment. Value 1: upsloping. Value 2: flat. Value 3: downsloping.
11	CA(Number of vessels colored)	Number of major vessels (0-3) coloured by fluoroscopy.
12	Thal (Obtained defect)	3 = normal; 6 = fixed defect; 7 = reversible defect.
13	Heredity	Family history of coronary artery disease (1 = yes; 0 = no).
14	Class	Class is either healthy (0) or with heart-disease (1)

B. *Data Preprocessing*

The collected data were used to create a structured database system. The fields were identified, duplications were extracted, missing values were filled and the data were coded according to attribute domain value.

After data cleaning the number of cases was reduced mainly due to unavailability of clinical result. Attribute domain values are provided by practicing cardiologist.

VII. CLASSIFICATION ALGORITHM

1. Multi-Layer Perceptron Neural Network (MLPNN)

Literature analysis unveils a persistent application of feed forward neural networks, from amidst the various categories of connections for artificial neurons. A kind of feedforward neural

network mechanism is the Multi-layer Perceptron Neural Networks (MLPNN). The structure of MLPNN is shown in Fig. II.

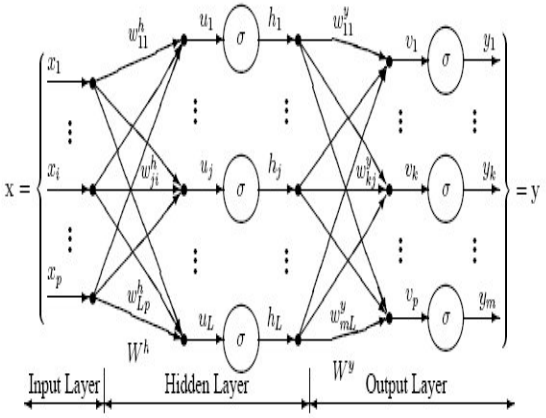

Fig. II Structure of MLPNN.

In MLPNN the lone and primary task of the neurons in the input layer is the division of the input signal x_i among neurons in the hidden layer. Every neuron j in the hidden layer adds up its input signals x_i once it weights them with the strengths of the respective connections w_{ji} from the input layer and determines its output y_j as a function f of the sum, given as

$$y_i = f(\sum w_{ji} x_i) \qquad (1)$$

At this instant it is possible for f to be a simple threshold function such as a sigmoid, or a hyperbolic tangent function. The output of neurons in the output layer is determined in an identical fashion.

The working of Multi-Layer Perceptron Neural Network is summarized in steps as mentioned below:

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 Back-Propagation 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.

2. *Back-Propagation Training*

The back-propagation algorithm can be employed effectively to train neural networks; it is widely recognized for applications to layered feed-forward networks, or multi-layer perceptrons. The back-propagation learning algorithm can be divided into two phases: propagation and weight update [14].

Phase 1: Propagation

1) Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.

2) Back propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

Phase 2: Weight update

For each weight-synapse:

1) Multiply its output delta and input activation to get the gradient of the weight.

2) Bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight.

Repeat the phase 1 and 2 until the performance of the network is good enough.

VIII. FUNDAMENTAL STEPS IN HEART DISEASE DIAGNOSIS SYSTEM USING ARTIFICIAL NEURAL NETWORK

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

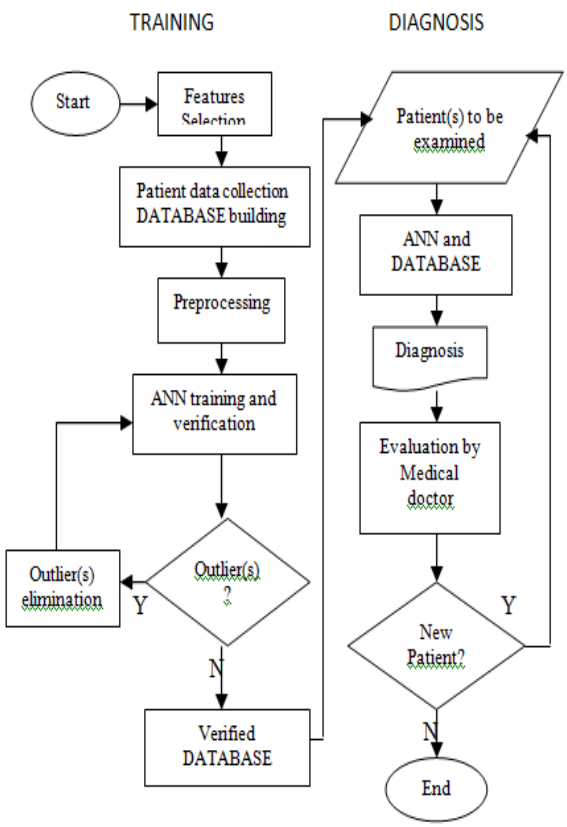

Fig. III Fundamental Steps in Heart Disease Diagnosis using ANN

IX. EXPERIMENTAL RESULT

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.

Table II. Confusion Matrix

		Actual Value		
Prediction		*p*	*n*	Total
	P'	True Positive	False Positive	P'
	n'	False Negative	True Negative	N'
	Total	P	N	

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

$$SE = \frac{TP}{(TP + FN)} * 100 \quad (2)$$

$$SP = \frac{TN}{(TN + FN)} * 100 \quad (3)$$

$$AC = \frac{(\mathrm{TP} + \mathrm{TN})}{(TN + TP + FN + FP)} * 100 \quad (4)$$

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.

After data preprocessing, obtained data set is divided into two parts. On the first part, made of 70 patient records, the training algorithm is built and on remaining 100 records testing is performed. After completion, following results are obtained as shown in Table III.

Table III Analysis Result

Proposed Algorithm	Sensitivity	Specificity	Accuracy
MLPNN with Back Propagation Algorithm	92%	92.5%	94%

From Table III it is clear that the proposed Heart Disease Diagnostic System using MLPNN (Multi-layer Perceptron Neural Network) with Back Propagation algorithm has better accuracy, sensitivity and specificity compared to other approaches.
During training of system, as the numbers of input data sets are increased designed network leads to more accuracy.

CONCLUSION

In this research paper, I have presented diagnostic system for Heart disease using data mining and artificial neural network (ANN) techniques. From the ANN, a Multi-layer 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 efficiently.

REFERENCES

[1] Tzung-I Tang, Gang Zheng, Yalou Huang, Guangfu Shu, Pengtao Wang, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and System Reconstruction Analysis", IEMS Vol. 4, No. 1, pp. 102-108, June 2005.

[2] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare", Journal of healthcare information management, Vol. 19, Issue 2, Pages 64-72, 2005.

[3] S. Stilou, P. D. Bamidis, N. Maglaveras, C. Pappas, "Mining association rules from clinical databases: an intelligent diagnostic process in healthcare".

[4] A. Bellaachia and Erhan Guven, "Predicting Breast Cancer Survivability using Data Mining Techniques", the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006.

[5] 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.

[6] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules", Seminar Presentation at University of Tokyo, 2004.

[7] Agrawal R., Imielinski T. and Swami A., 'Mining association rules between sets of items in large databases'.

[8] Douglas Burdick, Manuel Calimlim, Johannes Gehrke, "MAFIA: A Maximal Frequent Itemset Algorithm for Transactional Databases", Proceedings of the 17th International Conference on Data Engineering.

[9] http://en.wikipedia.org/wiki/Heart_disease.

[10] "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.

[11] Cornell D., "A Vertical Partitioning Algorithm for Relational databases", Proceedings of the Third International Conference on Data Engineering, Los Angeles, CA, February 1987.

[12] O. Galip Saracoglu, "Artificial Neural Network Approach for Prediction of Absorption Measurements of an Evanescent Field Fiber Sensor", Sensors, Vol. 8, pp. 1585-1594, 2008.

[13] Savkovic-Stevanovic "Neural networks for process analysis and optimization: modeling and application" Computer and chemical engineering, Vol. 18, No 11-12 (14 ref.), pp. 1149-1155, 1994.

[14] http://en.wikipedia.org/wiki/Backpropagation.

[15] http://archive.ics.uci.edu/ml/datasets/Heart+Disease.