International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 3, Issue 4, April 2014 ISSN 2319 - 4847

Artificial Neural Network Approach for Classification of Heart Disease Dataset

Manjusha B. Wadhonkar [1], Prof. P.A. Tijare[2] and Prof. S.N.Sawalkar[3]

[1]M.E Computer Engineering (Second Year)., Computer Science and Engineering Department.
Sipna college of Engineering, Amravati

[2]Associate Professor, Computer Science and Engineering Department,
Sipna college of Engineering, Amravati

[3]Assistant. Professor, Computer Science and Engineering department,
Sipna College of engineering, Amravati

Abstract

Artificial Neural Networks (ANNs) play an important role in the field of medical science in solving health problems and diagnosing various diseases .To accurately diagnose the persons' disease condition it is important to use appropriate methods that minimize the errors in diagnosis. So in this paper, for the classification of heart disease dataset, multilayer feed forward network with backpropogation algorithm is proposed. Artificial Neural Network (ANN) is widely used data mining method to extract patterns. Data mining is the process of automating information discovery. Main aim of data mining is to find relationships in data data and to predict outcomes. Classification is one of the important data mining techniques for classifying given set of input data. Many real world problems in various fields can be solved by using classification approach such as business, science, industry and medicine. For classification, medical data related to Heart disease is considered and analyzed using artificial neural network (ANN) [14]. To perform classification task of heart disease data, the neural network is trained using back propagation algorithm with momentum. For the classification, four databases such as Cleveland, Hungarian, Switzerland and long-Beach V.A are considered.

Keywords: Heart disease Dataset, Artificial neural network, MLP, Backpropogation Algorithm, Knowledge data discovery.

1. INTRODUCTION

Data mining also called knowledge data discovery is the process of analyzing data from different perspective and summarizing it into useful information. Knowledge Discovery in database is concerned with the development of methods and techniques for making sense of data. Main aim of data mining is to uncover relationship in data and predict the outcome [1]. Data mining extract the patterns in the process of knowledge discovery in the database. As the dataset has grown in size and complexity, new emerging field of data mining provides new techniques and methods which help to analyse and understand large bodies of data. Classification is one of the important techniques of data mining. Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts [2].

In classification, inputs are given a set of data, called a training set, where each record consists of several fields or attributes. These attributes are continuous, coming from an ordered domain, or categorical, coming from an unordered domain. One of the attributes, called the classifying attribute, indicates the class to which each dataset belongs. The objective of classification is the method to build a model of the classifying attribute based upon the other attributes which are not from the training data set [3].

Data mining is an interdisciplinary field. For the implementation of classification, techniques from other disciplines may also be applied, such as neural networks, fuzzy and/or rough set theory, knowledge representation, inductive logic programming, or high performance computing which will lead to intelligent, low cost solution [3].

Artificial neural network is one of the widely used techniques for extraction of patterns in data mining. Artificial neural networks has some advantages such as it automatically allow arbitrary nonlinear relations between the independent and dependent variables, and allows all possible interactions between the dependent variables. Due to above said advantages of ANN the use of neural network technique is adopted for the classification of dataset [14]. Neural network is trained using backpropogation algorithm. To increase the efficiency of classification process, parallel processing is implemented at each node in the network. After training the neural network performance of the network is analysed with various set of test data.

2. LITERATURE SURVEY

A major challenge, facing healthcare organizations (hospitals, medical centre) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective [4]

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. Global burden of disease estimates for 2001 by World Bank Country Groups shows severity statistics indicated in year 2001 is 25.2 % for India and from literature survey now it has increased to 46% [5]. In spite of the rapid development of pathological research and clinical technologies, more than 60,000 people die suddenly each year in India due to arrhythmias and heart diseases [6].

In case of uncertainty of heart disease symptoms even experienced cardiologists need complimentary assistance from intelligent decision system to arrive at precise diagnosis of cardiac disease. A number of techniques have been used for identification of heart diseases including waveform analysis, time frequency analysis, complexity measures, Neuro Fuzzy RBF NN and a total least square based prony modeling algorithm. But it has been observed that classification accuracies were not good (only up to 79 %) , Classification of artificial neural network using ANN with feature selection gives only 80% result, with these techniques and still enough scope in improving by choosing appropriate NN model [6].

The researchers in the medical field identify and predict the diseases besides proffering effective care for patients with the aid of data mining techniques. The data mining techniques have been utilized by a wide variety of works in the literature to diagnose various diseases including: Diabetes, Hepatitis, Cancer, Heart diseases and the like. Information associated with the disease, prevailing in the form of electronic clinical records, treatment information, gene expressions, images and more; were employed in all these works. In the recent past, the data mining techniques were utilized by several researchers to present diagnosis approaches for diverse types of heart diseases [7]-[14].

Analysis of different data mining techniques that can be employed in automated heart disease prediction systems. Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that Neural Network with 8 and 13 attributes has shown the approximate accuracy of 81% so far. Moreover, in combination with Genetic Algorithm and 6 attributes, Decision Tree has shown 99.2% efficiency [7].

By the analysis of above results, it is observe that more careful and efficient methods of cardiac diseases and periodic examination are of high importance. Thus for efficient classification of dataset ANN is proposed as a method of classification in data mining.

3. PROPOSED WORK

For the classification of Heart disease dataset using 13 input attributes and one output testing results gives maximum 90.6% accuracy for single layer and 94% for multilayer feed forward network [10]. To increase the accuracy of classification of heart disease dataset in a proposed system two other input attributes as Smoke and Fam_History are used which increases the risk of cardiovascular disease [11]. Thus this proposed system is an attempt to introduce a classification approach using multilayer Perceptron (MLP) with backpropogation which includes 15 input attributes and an output attribute. Attribute values for smoke and family-history are tabulated in below table 1.

Table 1: Proposed attribute values

Name	Description	Range
Smoke	1=True,0=False	1,0
Family history	1=True,0=False	1,0

3.1 Multilayer Perceptron Neural network

For more complex decision function the inputs are fed into a number of perceptions nodes, each with its own set of weights and threshold [8].The output of these input nodes are given as an output to another layer of nodes. Output of final layer of nodes is the output of the network. Such type of network is termed as MLP [6]. Learning algorithm is used to compute the connection weights. There are different variants of backpropogation algorithm in the literature [9]-[13].

3.2 Dataset for classification of heart disease

According to the World Health Organization statistics, heart disease is the most common cause of death among other diseases [15]. One of the most important parts for the classification of heart disease dataset using ANN is selecting data. Data is obtained from four different datasets of UCI, centre for machine learning and intelligent system [12].This database contains total 76 attributes , but for classification ,a subset of 16 of them namely Age(in years),Sex, Chest Pain type, Resting blood Pressure, Serum cholesterol, fasting blood sugar, Resting ECG, Maximum heart rate achieved, exercise induced angina, ST depression induced by exercise relative to rest, The slope of the peak exercise ST segment, Number of major vessels ,Smoke and fam_hisory and last feature is output based on classification of heart disease. Below table shows name of all four dataset and their number of instances [12].

Table 2: Database Names and their Instances

Name of Database	No of instances
Cleveland	303
Hungarian	294
Switzerland	123
Long Beach V.A	200

The goal or output field is a two bit value which will represent four different classes as class 0-normal person, 1-first stroke, 2- second stroke and 3- end of life.

3.3 Framework for the classification of heart disease dataset

The workflow of ANN analysis for the classification of dataset is as shown in figure 1(a) and 1(b) which provides brief description of fundamental steps that should be followed to apply ANNs for the classification of heart disease dataset.

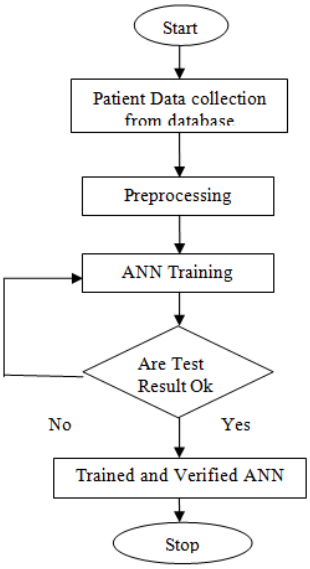

Figure 1(a) Training procedure for ANN based classification of heart disease dataset

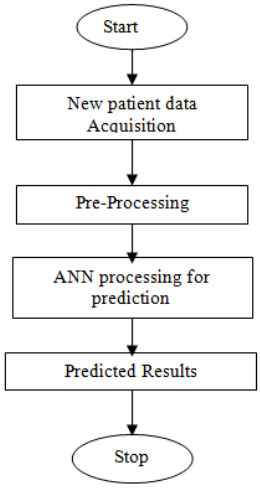

Figure 1(b) Testing of ANN based classification of heart disease dataset.

3.3.1 Data Collection

Neural network is trained using suitable dataset of example cases. This dataset is nothing but records of patient's stores in a database. Database should contain sufficient number of reliable examples to be given as an input to the training network.

3.3.2 Pre-processing

Data in the training dataset must be pre-processed before the evaluation by the neural network. Normally data are scaled to be within the interval [0, 1] because the interference function used is logistic one. During pre-processing some data are missing should be removed from the dataset to improve the classification performance.

3.3.3 Training & verification using ANN

The neural network is trained with Heart Diseases database by using feed forward neural network model and backpropogation learning algorithm with momentum and variable learning rate. The input layer of the network consists of 15 neurons to represent each attribute as the database consists of 15 attributes. The number of classes is four: 0 – normal person, 1- first stroke, 2- second stroke and 3- end of life. The output layer consists of two neurons to represent these four classes. The backpropogation algorithm with momentum is used to train the neural network during the training process. Several neural networks are constructed with and without hidden layers, i.e., single and multi layer networks and trained with heart disease dataset. A selection of maximum number of epochs is provided prior to training within which the training is expected to converge. The convergence is said to be achieved when the error between the output generated by the trained network and the actual output from the database matches within a certain error limit preset before the training. If a convergence is not achieved then training with new network configuration (i.e. hidden neuron count) is carried out.

3.3.4 Verification

Once a convergence is achieved the ANN is declared to be trained and its verification is initiated which normally is similar to the verification carried out during training by comparing the predicted outputs of the ANN with the actual ones, only difference being the dataset used this time is different from the one used in training. Once this verification results match then the ANN is declared as trained and verified for application purpose. Periodic verification of ANN and retraining if verification fails is a normal process with the ANNs.

3.3.5 Testing

Once an ANN is declared to be trained and verified it is usable for application to the classification problem. In this phase it is provided with new user's heart disease data and asked to classify. The results are used as correctly generated.

4. PROPOSED GUI FOR THE CLASSIFICATION OF HEART DISEASE DATASET

The architecture of the system proposed for the classification of heart disease dataset is as shown in figure 2. Initially selected database (76 attributes) and its subset of (15) have been acquired and a Database structure for the system is being set into place for the loading of the Database as well as Help presentation on the database. A scalable approach is used with the use of Database module which uses two scripts labeled as "Database Info" and "Database Load". The first one provides the information about the Database features/attributes and their naming, the second one is provided for loading the database in memory for processing.

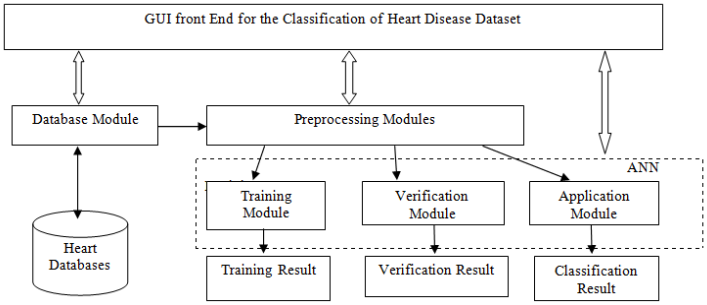

Figure 2 Architecture of the System for ANN based classification of Heart Disease Dataset

- Training Module:-ANN is trained by using MLP with Backpropogation learning algorithm.
- Training Result: - Are predicted results, obtained by summing the results of inputs with adjusted weights.
- Verification Module:-In this module predicted output of ANN is compared with actual output.
- Verification Result: - Once this verification results match then the ANN is declared as trained and verified for application purpose.
- Application Module: - Once an ANN is declared to be trained and verified it is usable for application to the classification problem. In this phase it is provided with new patient's heart disease data as an input and asked to classify.
- Classification Result: - For inputs of any new patient's heart disease dataset, it provides results such as whether the patient is a healthy person or if not then to which class it belongs.

5. CONCLUSION

For the classification of heart disease dataset ANN is proposed as a classification method. As per wide range of applicability of ANN and their ability to learn complex and nonlinear relationships including noisy or less precise information, neural networks are well suited to solve problems. From the design of neural networks, it is evident that MLP NNs required a compact architecture as compared to other NNs, in terms of number of hidden nodes required for the classification. The number parameters such as weights and biases required for the designing of MLP NN is sufficiently lower than other. Thus to improve computational efficiency of the system, Backpropogation algorithm with momentum and variable learning rate is used to train the neural networks for classification.

REFERENCES

[1] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, (1996) "From Data Mining to Knowledge discovery in Databases ", AI Magazine Volume 17 Number 3 (© AAAI) style)
[2] Jiawei Han, Micheline Kamber, (2006)"Data Mining: concepts and Techniques", ELSEVIER
[3] John Shafer, Rakesh Agarwal, and Manish Mehta, (1996)"SPRINT: A scalable parallel Classifier for data mining", In Proc. Of the VLDB Conference, Bombay, India.
[4] 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.
[5] R. Rojas, (1996) "Neural Networks: a systematic introduction", Springer-Verlag.
[6] R.P.Lippmann, "Pattern classification using neural networks, (1989)" IEEE Communication. Mag., pp.47–64
[7] Simon Haykin, (2001) "Neural Networks – A Comprehensive Foundation", Pearson Education.
[8] Bishop C. 1997. "Neural Networks for pattern Recognition", Oxford University Press, New York.
[9] Hagan, M.T., Demuth H.B., Beale M.H. 1997. "Neural Network Design, PWS publishing", Boston, M

[10] Dr. K Usha Rani, 2011 .” Analysis of Heart Disease dataset using Neural Network Approach”, International journal of data Mining & Knowledge Management (IJDKP) Vol .1

[11] Nidhi Bhatia and Kiran Jyoti, 2012,”An analysis of heart disease prediction using different data mining Techniques”, International Journal of Engineering Research and Technology (IJERT) Vol 1.

[12] Murphy, P. M. and Aha, D. W. 2004. UCI Machine Learning Databases Repository Irvine C. A.: University of California, Department of Information and Computer science,ftp://ftp.ics.uci.edu/pub/machinelearningdatabases/heart/.

[13] Ranjana Raut, Dr. S. V. Dudul 2010,”Intelligent Diagnosis of Heart Diseases using Neural Network Approach” International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 2.

[14] Manjusha B. Wadhonkar, P.A.Tijare, S.N.Sawalkar,April 2013,”Classification of Heart disease Dataset using Multilayer feed forward backpropogation algorithm”, International Journal of Application or Innovation in Engineering & management(IJAIEM),Vol 2,issue 4,pp 214-220.

[15] N. Elfadil, A. Hossen, December 2009. “Identification of Patients with Congestive Heart Failure Using Different Neural Networks Approaches”, Journal Technology and Health Core, vol. 17, No. 4.

AUTHOR

Manjusha B. Wadhonkar M.E(Computer engineering) Second Year. Computer Science and Engineering Department. Sipna college of Engineering, Amravati (M.S).