IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

Diagnosis of Heart Disease using Neural Network

Shiva Shrivastava
shiva.shri16@gmail.com

Neeraj Mehta
neerajmehta@ipsacademy.org

Abstract – **Heart disease prediction is treated as most complicated task in the field of medical sciences. Thus there arises a need to develop a decision support system for detecting heart disease of a patient. Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. The main objective of this research work is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. This paper proposes back propagation Neural Network technique for heart disease prediction. Performance of proposed approach is evaluated using confusion matrix plot.**

Keywords – **BPNN, Confusion Matrix, Heart Diseases.**

I. Introduction

Heart diseases remain the biggest cause of deaths for the last two decades. Recently computer technology and machine learning techniques to develop software to assist doctors in making decision of heart disease in the early stage. The diagnosis of heart disease depends on clinical and pathological data. Heart disease prediction system can assist medical professionals in predicting heart disease status based on the clinical data of patients. In biomedical field data mining plays an essential role for prediction of diseases in biomedical diagnosis, the information provided by the patients may include redundant and interrelated symptoms and signs especially when the patients suffer from more than one type of disease of the same category. The physicians may not able to diagnose it correctly [1].

Making a diagnosis of heart disease includes taking a complete medical evaluation and history and physical examination and early diagnosis of heart disease can help reduce the rate of mortality. One of the best ways to diagnose a heart disease is by using echocardiography. Echocardiography, or echo, is a painless test that uses sound waves to create pictures of the heart. The test gives information about the size and shape of the heart and how well the heart chambers and valves are working. Echo also can be done to detect heart problems in infants and children.

Data mining provides methodology and technology to process and analyze huge amounts of data into useful information for decision making that is a fundamental part of healthcare management.

The main objective of this paper is to implement a framework for prediction of heart disease using major risk factors based on data mining approach using Neural Network Classifier. For this we first determine the number of inputs, layers and hidden neurons of the neural network and then apply the back-propagation algorithm to train the networks.

II. Classification of Heart Disease

There are many types of heart disease, but in this study the researcher chooses to discuss five types that are common to happen. Five common types of heart disease are discussed below.

Congenital Heart Disease: The term congenital or hereditary heart disease refers to heart disease which is passed down through the family, and this is considered as being a congenital type as it is principally inevitable and unpreventable. Congenital heart disease refers to a problem with the heart's structure and function due to abnormal heart development before birth. Congenital heart disease is the most common type of birth defect. it is responsible for more deaths in the first year of life than any other birth defects.

Congestive Heart Failure: Congestive heart failure is when the heart does not pump adequate blood to the other organs in the body. Congestive heart failure can often result from heart problem and constricted arteries. Congestive heart failure results in a heart which works a lot less efficiently than it should and can make further problems. Symptoms regularly consist of swelling and edema, shortness of breath, and kidney problems which in turn can lead to mysterious weight gain. Even elevated blood pressure and alcohol abuse can lead to congestive heart failure. A patient may be examined for congestive heart failure if they have suffered from heart problem in the past, are alcoholic, have a family history of heart problems or show one or all of the symptoms that are caused by congestive heart failure. There are choices of examinations that aid a doctor in diagnosing this heart crisis. Treatment should begin without delay, starting with changes to diet and exercise, as patients should abolish salt from the diet altogether and sternly limit their fluid

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

intake. Further treatment should be done by a professional.

Coronary Heart Disease (Ischemic Heart Disease): Coronary heart disease or in its medical term Ischemic heart disease is the most frequent type of heart problem of all, and is also the leading reason of heart attacks. Coronary heart disease is a term that refers to damage to the heart that happens because its blood supply is decreased, and what happens here is that fatty deposits build up on the linings of the blood vessels that provide the heart muscles with blood, resulting in them narrowing. These narrowing decreases the blood supply to the heart muscles and causes pain that is identified as angina. There are a few factors which are considered as being responsible causes of coronary heart disease. One in particular is high cholesterol that can increase fat concentration in our blood and create the building up of fatty deposits. Another one of the major factors of coronary heart disease is cigarette and tobacco smoke, as a smoker's risk of getting heart problem is two times that of a non-smoker, and studies have actually revealed that after five years of quitting smoking, the risk of developing heart problem is the same as that of someone who had never smoked in their life.

Pulmonary Heart Disease: Pulmonary heart disease is a disease that comes from a lung, or pulmonary, disorder, or a complication of lung problems where the blood flow into the lungs is slowed or even totally blocked, resulting in increased pressure on the lungs. There are a number of different symptoms that typically come with pulmonary heart disease, such as shortness of breath, syncope, dyspnea, and chest pain. It is a state which is often misdiagnosed, and has frequently progressed to late stages by the time that it is actually correctly diagnosed. It has been previously chronic and untreatable with a poor survival rate. However, there are now numerous new treatments which are accessible which have extensively improved the overall prognosis of this disease.

Rheumatic Heart Disease: Rheumatic heart disease frequently derives from strep throat infections. This can be a reason for alarm for many because strep throat, while often preventable, is a quite common condition that affects many people who do not treat a minor sore throat infection in time. However, there is no reason to be because rheumatic heart disease that comes from strep throat is fairly rare. Actually, the sheer volume of cases of rheumatic heart disease has decreased considerably since the 1960's.

III. PROPOSED METHODOLOGY

The problem with risk factors related to heart disease is that there are many risk factors involved like age, usage of cigarette, blood cholesterol, person's fitness, blood pressure, stress and etc. and understanding and categorizing each one according to its importance is a difficult task. Also a heart disease is often detected when a patient reaches advanced stage of the disease. Hence the risk factors are analyzed from various sources [2]-[3]. The dataset was composed of 12 important risk factors which were sex, age, family history blood pressure, Smoking Habit, alcohol consumption, physical inactivity, diabetes, blood cholesterol, poor diet, obesity .The system indicated whether the patient had risk of heart disease or not. The data for 50 people was collected from surveys done by the American Heart Association [3]. Most of the heart disease patients had many similarities in the risk factors [4]. Table 4.1 shows the identified important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

Table 1: Risk factors values and their encodings [5]

S. No.	Risk Factors	Values
1	Sex	Male (1), Female (0)
2	Age (years)	20-34 (-2), 35-50 (-1), 51-60 (0), 61-79 (1) , >79 (2)
3	Blood Cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
5	Hereditary	Family Member diagnosed with HD -Yes (1) Otherwise –No (0)
6	Smoking	Yes (1) or No (0)
7	Alcohol Intake	Yes (1) or No (0)
8	Physical Activity	Low (-1) , Normal (0) or High (-1)
9	Diabetes	Yes (1) or No (0)
10	Diet	Poor (-1), Normal (0) or Good (1)
11	Obesity	Yes (1) or No (0)
12	Stress	Yes (1) or No (0)
Output	Heart Disease	Yes (1) or No (0)

Data analysis has been carried out in order to transform data into useful form, for this the values were encoded mostly between a range [-1, 1]. Data

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

analysis also removed the inconsistency and anomalies in the data. This was needed. Data analysis was needed for correct data pre-processing. The removal of missing and incorrect inputs will help the neural network to generalize well.
In this paper, Neural Network based approach is used to determine optimum number of clusters in analyzed data.

Neural Network

Back Propagation Neural Network (BPNN) generates complex decision boundaries in feature space. BPNN in specific circumstances resembles Bayesian Posterior Probabilities at its output. These conditions are essential to achieve low error performance for given set of features along with selection of parameters such as training samples, hidden layer nodes and learning rate. In else case, the performance of BPNN could not be evaluated. For W number of weights and N number of nodes, numbers of samples (m) are depicted to correctly classify future samples in following manner:

$$m \geq O\left(\frac{W}{\epsilon}\log\frac{N}{\epsilon}\right) \quad (1)$$

The theoretical computation of number of hidden nodes is not a specific process for hidden layers. Testing method is commonly entertained for selection of these followed in the constrained environment of performance [6].
The data for risk factors related to heart diseases collected from 50 people is provided in Table 2.

Table 2: Patient's case study data in encoded form

No	Sex	Age	Blood Cholesterol	Blood Pressure	Hereditary	Smoking	Alcohol Intake	Physical Activity	Diabetes	Diet	Obesity	Stress	Heart Disease
1	F	35	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
2	M	70	Low	Low	No	No	Yes	High	Yes	Normal	No	No	No
3	F	60	High	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
4	F	36	Low	Normal	No	No	No	Normal	No	Good	No	No	No
5	M	30	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
6	F	39	Low	Normal	Yes	No	Yes	High	Yes	Normal	No	Yes	No
7	F	41	High	Normal	No	No	No	Low	No	Poor	Yes	No	No
8	M	70	High	Normal	No	No	Yes	Low	No	Poor	Yes	No	Yes
9	M	65	Normal	High	Yes	Yes	Yes	Normal	Yes	Poor	Yes	No	Yes
10	M	30	Normal	High	No	Yes	No	Normal	No	Good	No	Yes	No
11	F	31	Low	Normal	No	No	No	High	No	Normal	No	No	No
12	F	29	Low	Normal	No	No	Yes	High	No	Good	No	No	No
13	M	30	Low	Normal	No	No	Yes	Normal	No	Normal	No	No	No
14	F	45	Normal	High	Yes	Yes	No	Normal	Yes	Normal	Yes	Yes	No
15	M	25	High	Normal	Yes	Yes	Yes	Low	Yes	Normal	No	No	Yes

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

16	F	37	Normal	Normal	No	No	No	Normal	Yes	Poor	No	Yes	No
17	F	37	Normal	High	No	Yes	Yes	High	No	Poor	No	Yes	No
18	M	53	High	Low	No	Yes	No	Normal	Yes	Normal	No	Yes	No
19	M	57	High	Normal	No	Yes	No	Low	No	Poor	Yes	Yes	Yes
20	M	52	High	Low	No	No	No	Normal	Yes	Poor	Yes	No	No
21	M	48	Normal	Normal	Yes	Yes	Yes	Normal	No	Normal	No	No	Yes
22	M	62	High	High	No	Yes	Yes	Normal	Yes	Normal	No	No	Yes
23	M	56	Normal	High	No	Yes	Yes	Low	No	Poor	Yes	Yes	Yes
24	F	27	Low	Normal	No	No	No	High	No	Good	No	No	No
25	M	33	Normal	Normal	No	No	No	Normal	Yes	Good	No	No	No
26	F	33	Normal	Normal	No	No	Yes	Low	Yes	Poor	No	Yes	No
27	M	37	High	Normal	No	No	Yes	Normal	No	Normal	No	Yes	No
28	M	43	Normal	High	No	No	No	Normal	Yes	Poor	Yes	Yes	Yes
29	M	46	Low	Normal	No	No	No	Normal	Yes	Poor	Yes	Yes	No
30	F	36	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
31	F	29	Low	Normal	No	No	No	Normal	No	Good	No	No	No
32	F	47	Normal	Normal	No	No	Yes	High	Yes	Normal	No	Yes	No
33	M	58	High	High	No	Yes	Yes	Normal	Yes	Normal	No	Yes	Yes
34	M	44	High	Normal	Yes	Yes	Yes	Normal	No	Normal	Yes	Yes	Yes
35	F	36	Normal	High	No	No	No	Normal	No	Good	Yes	No	Yes
36	M	42	Low	Normal	Yes	No	Yes	Low	No	Poor	No	Yes	No
37	F	25	Low	Normal	No	No	No	High	No	Poor	No	No	No
38	F	28	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
39	F	26	Low	Normal	Yes	No	No	Normal	No	Normal	Yes	No	Yes
40	M	28	Low	Normal	No	No	No	Normal	No	Poor	No	No	No
41	F	45	High	Normal	No	No	Yes	Low	Yes	Poor	Yes	Yes	Yes
42	M	63	Low	Low	No	No	Yes	High	Yes	Good	No	No	No

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

43	F	55	High	High	No	No	No	Normal	Yes	Normal	Yes	Yes	Yes
44	F	44	Low	Normal	No	No	No	Normal	No	Normal	No	No	No
45	M	35	Low	Normal	No	No	Yes	High	No	Normal	No	No	No
46	F	42	Normal	Normal	No	No	Yes	High	Yes	Good	No	No	No
47	F	43	Normal	Normal	No	No	No	Low	No	Poor	Yes	No	No
48	M	65	Normal	Normal	No	No	Yes	Low	No	Normal	Yes	Yes	Yes
49	M	74	Normal	High	No	Yes	Yes	Normal	Yes	Normal	Yes	Yes	Yes
50	M	36	Normal	High	No	Yes	No	Normal	No	Poor	No	No	No

IV. Simulation And Results

The performance of proposed technique has been studied by means of MATLAB simulation.

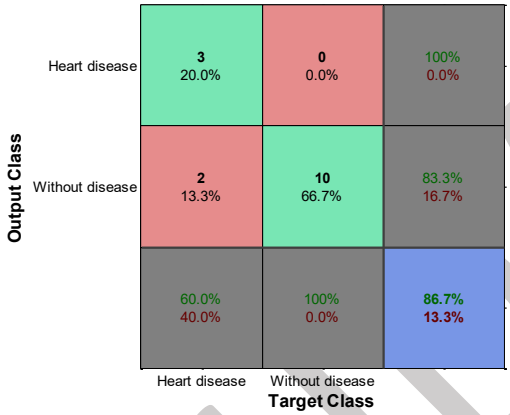

Figure 1: Confusion matrix for NN method

The row and column are the labels for detection and no detection of heart disease from database. There are 2 sets of classes and each class having different set of detection. Total 15 samples of patients are taken out of which 3 patients are classified correctly and none of the samples were misclassified but in normal category 10 samples are classified correctly out of 15 samples and 2 samples are misclassified. The confusion plot indicates the accuracy i.e. 86.7% for this approach.

V. Conclusion

Data mining and machine learning algorithms in the medical field extracts distinctive concealed patterns from the medical data. They can be utilized for the examination of vital clinical parameters, expectation of different diseases, estimating assignments in pharmaceutical, extraction of medical knowledge, treatment planning support and patient administration. Various algorithms are exist in literature for the prediction of heart disease using major risk factors. Back propagation provides the input to our network to give better results and it was found that it is effective to predict the risk of heart disease when the person provide the required attributes value.

Reference

[1] National Heart Lung and Blood Institute, "What is Echocardiography?", Available at http://www.nhlbi.nih.gov/health/dci/Diseases/echo/echo_whatis.html

[2] Centre for Disease Control and Prevention, Online available at: http://www.cdc.gov/heartdisease/risk_factors.htm

[3] American Heart Association, Online available at: http://www.heart.org/HEARTORG/Conditions

[4] D. Isern, D. Sanchez, and A. Moreno, "Agents Applied in Health Care: A Review", International Journal of Medical Informatics, Vol. 79, Issue 3, pp. 146-166, 2010.

[5] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", Proceedings of IEEE Conference on Information and Communication Technologies (ICT), pp. 1227–1231, April 2013.

[6] Christos Stergiou and Dimitrios Siganos, "Neural Networks", Report available at: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html