University Journal of Surgery and Surgical Specialities

CLASSIFICATION OF RISK LEVEL FOR ISCHEMIC HEART DISEASE IN INDIA USING ARTIFICIAL INTELLIGENCE - A CASE STUDY

*Dr.P. Amirtharaj, K.Rajeswari, Dr.V.Vaithiyanathan
*Associate Professor, Department of Cardiothoracic Surgery, Madras Medical College, Chennai.

Abstract

Cardiovascular Diseases (CVD) comprises of a group of diseases of the heart and vascular system. The major conditions include Coronary Heart Disease (CHD) or Ischemic Heart Disease(IHD) which cause 25-30 percent of deaths in most industrialized countries. India is in a risk of developing more death due to CHD. Hence a Decision Support System (DSS) is proposed to identify the level of risk for classifying the Ischemic Heart Disease of a Patient. This will help the patients in taking precautionary steps like following a balanced diet and medication which in turn may increase the life time of a patient. The features for prediction are selected after considering Indian conditions from literature and based on the Expert knowledge from Doctors. Framingham Risk score which has five attributes is used for comparison. Our proposed system will have fourteen features to be analyzed according to Indian Conditions. The system is a theoretical study which proposes implementation of Artificial Intelligence to mine the knowledge from Medical data collected.

Keywords

Decision Support System (DSS); Ischemic Heart disease (IHD); Artificial Intelligence; Coronary Heart Disease (CHD); Risk

I. Introduction

When it comes to matters of the heart Indians seem to fare badly. By the end of next year, India will bear 60% of the world's heart disease burden. What's worse, compared to people in other developed countries, the average age of patients with heart disease is at least 5-8 years lower among Indians. Sixty is the average age of heart patients in India against 63-68 in developed countries. It is slipping further to the mid-50s. Indians are also more likely to have types of heart disease that lead to worse outcomes like ischemic heart disease — a condition characterized by reduced blood supply to the heart[l]. The WHO has drawn attention to the fact that CHD is our modern "epidemic". Studies say that nowadays, sudden death during sleep has become prevalent. This is due to the lack of oxygen supply to heart.

CHD may manifest itself as any one of the following forms.
Angina pectoris effort
Myocardial Infarction
Irregularities of the heart
Cardiac failure
Sudden death

Myocardial Infarction is specific to CHD. Others are not and may cause confusion during diagnosis [2]. Rose [3] calculates the "incubation period" of CHD may be 10 years or more.

A. Coronary Heart Disease In INDIA

A large body of data exists on the occurrence of CHD in hospital patients. However, there are only two studies one in Chandigarh, screening patients over the age 30, by a 12-lead ECG and the other in Haryana were the prevalence is 65.4 and 47.8 per 1000 males and females respectively in urban population [4].This proposed study will be carried out in Chennai, Madras Medical College for the age group 30 and above.

The pattern of CHD in India has been reported to be as follows:
CHD appears a decade earlier compared with the age incidence in developed countries. The peak period is attained between 51 -60 years.
Males are affected more than females.
Hypertension [9] and Diabetes account for about 40 percent of all cases.
Heavy smoking is responsible aetologically in a good number of cases [5][6][7][8].
Serum Cholesterol [10]
The risk of CHD is 2-3 times higher in Diabetes than in Non-Diabetes [11].
A family history of CHD.
Sedentary life style [12].
Type A individuals [13].
Higher alcohol intake, defined as 75g [9]

II. RISK ASSESSMENT BASED ON CLINICAL CONDITIONS AND RISK FACTOR EVALUATION

Absolute Risk is divided into three categories namely, high, intermediate and low. Patients at High Risk deserve Intensive risk reduction therapy.

Patients at intermediate risk also require clinical intervention to the extent that therapy is safe and effective. Patients with low risk may be encouraged by their physicians to follow Health recommendations for Prevention of CHD. Each class of absolute risk may be expressed in quantitative terms.

Table 1 : RISK CLASSIFICATION

Risk Classification	10 Year Absolute Risk for CHD
Low	<10%
Medium	10-20%
High	>20%

A. From ingham Risk Analysis

Framingham heart study has five features to be entered, namely Age, Total Cholesterol, Smoker/Non smoker, Systolic Blood pressure, HDL. Points are added together to predict the 10-year risk score of Myocardial Infarction. Sample score for Age is as follows [14]. Similar scores are given to other features. The scores are summed to obtain the Risk factor

TABLE II. AGE SCORE

Age	Points
20-34	-9
35-39	-4
40-44	0
45-49	3
50-54	6
55-59	8
60-64	10
65-69	11
70-74	12
75-79	13

B. Indian Heart Risk Predicting Features

The following features are proposed to be collected and analyzed for Indian Heart risk score prediction based on extensive study and expert opinion from doctors with respect to Indian body conditions, life style and eating habits. After discussion with cardiologists a three stage questionnaire was prepared. Diagnosis is done through data collection for each individual patient as given in tables III, IV and V.

Stage I includes Physical Examination parameters. Stage II includes Co Morbid feature collection and Stage III includes attributes about personal habits and hereditary.

TABLE III. STAGE ONE DIAGNOSIS

Stage 1				
1-Age	2.Gender	3.Menopause	4.Height	5. Weight
6.BMI		7.WaistMeasure		

If Sex is Female, Pre or Post Menopause details are recorded

TABLE IV. STAGE TWO DIAGNOSIS

Stage 2				
8.SBP	9.DBP	10.Diabetes	11.Cholestrol	12.Thyroid Y[1] N[0]

SBP is Systolic Blood Pressure and DBP is Diastolic Blood Pressure.

TABLE V. STAGE THREE DIAGNOSIS

Stage 3				
13. Personal habits Smoker[1] No [0]	14. Family History Y[1] N[0]	15. Genetic factors Y[1] N[0]	16. Type A Stress [1] No Stress [0]	17. Sleep Disturbance Y[1] N[0]

Based on the data collected, the immediate risk analysis is classified as No Risk, Low Risk, Medium Risk or High Risk. Table V data is collected from experienced doctors in cardiology based on the importance of every feature collected in Tables III, IV and V. Three expert opinion is collected. Two identical opinion is taken into consideration for deriving conclusion. Varied opinion data of a patient is removed from the dataset.

TABLE VI. OUTPUT

Absolute Risk for CAD			
No Risk	Low Risk	Medium Risk	High Risk

III. Artificial intelligence in medical diagnosis

Nowadays many systems in health care domain are of multi-purpose type which support various complex tasks along-with diagnosis and combine more than one AI technique [17]. Medical care could be enhanced and costs could be reduced by means of an automated medical diagnosis system[15]. The importance of the role of data mining techniques in providing better patient care and effective diagnostic capabilities by finding patterns and extracting knowledge increases with the increase in the volume of stored data [16].

A.Classification of CAD R isk using A rtificial Neural Networks

An artificial neural network is the simulation of human brain. It is a supervised learning technique mainly used for Non Linear Classification. Popular Neural network algorithms include Hopfield, Multilayer perceptron, Self-Organising Map, Radial Basis Function, Adaptive Resonance Theory networks, Counter Propagation networks, Back Propagation networks etc. Our paper deals with Back propagation networks. Back Propagation Network or Feed Forward Networks is a network where there is no feedback. The information flows only in forward direction. This is a systematic method for training multilayer artificial neural network [19].The network is trained by supervised learning method. As it is a gradient descent method, it reduces the total squared error of the output computed by the net. As the work deals with realistic decision making, the system should reduce the error.

B. The Mathematical Model

When creating a functional model of the biological neuron, there are three basic components of importance. First, the synapses of the neuron are modeled as weights. The strength of the connection between an input and a neuron is noted by the value of the weight. Negative weight values reflect inhibitory connections, while positive values designate excitatory connections [18]. The next two components model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination. Finally, an activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1. The Back propagation algorithm, in particular, adaptively changes the internal network free parameters based on external stimulus. After trained, a neural network can make predictions about the membership of every test example. MLP is trained with the Back propagation algorithm suffers from the high number of parameters that need to be tuned, like learning rate, number of neurons, momentum rate, etc. However, the motivations to select this algorithm arise after observing that they have been used to solve problems in different domains, moreover, the output can be directly used for ranking purposes [21].

1. The algorithm was popularized by [20] and the same has been summarized below.Train the network with the given set of input patient data and obtain the output after propagating through the feed forward network.

2. Compare the result obtained with the desired output as given by expert doctors.

3. Calculate the error.
Adjust the weight values and repeat the process until error becomes vary negligible

The activity of hidden layer neurons is determined by the activity of input neurons and the connecting weights between the input layer and hidden layer. The activity of the output neurons depend on the activity of the hidden layer neurons and the connecting weights between the hidden and output layers. Initial weight and bias assignments influence the performance of the net. To get the best results, initial weights and biases are assigned between -0.5 to 0.5 or between -1 to +1. Faster learning of a Back propagation network can be obtained by Nguyen widrow initialization. Also Learning Rates should be increased to improve the performance.

C. Proposed Work
The database as described in Table III, IV, V is e collected for 125 patients. For every patient, Expert opinion from three experienced doctors about the Risk Level namely No Risk, Low Risk, Medium Risk, High Risk is obtained. If at least opinion of two doctors is similar, then the dataset is taken for training or testing. Otherwise the dataset is neglected.

The neural network of the proposed system is shown below in figure 1. The nodes 1,2,3..., 17 are the input nodes in the input layer. The 17 attributes collected during diagnosis of each patient is normalized in the range of 0 to 1 .The numerical attributes can be scaled in to analog form in the range between 0 to 1 using different methods. Here we have followed Linear Data scaling [19].
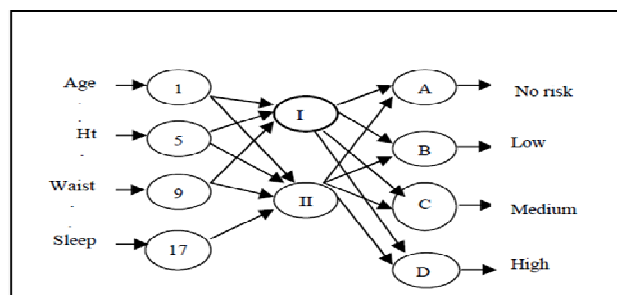

Figure 1. Neural Network Model for Ischemic Heart Disease.

D. Results and Discussion
After getting expert opinion, 110 patients data set is taken for analysis. 15 datasets are rejected due to lack of similarity in expert opinion. The different desired classification is as in Table VI.

TABLE VI. CLASSIFICATION OF RISK

Risk Cadre	No of Patients
No Risk	31
Low Risk	26
Medium Risk	26
High Risk	27

The Table VII below shows the Average Prediction Accuracy obtained with Back Propagation Neural network. Our future work will focus on reducing the number of features using optimization techniques like Genetic Algorithms; thereby will improve the Prediction Accuracy.

TABLE XII. AVERAGE PREDICTION ACCURACY

Method	Training	Validation	Testing
Back Propagation Neural Network	77%	82%	90%

The idea is to develop an intelligent system, which will support a person who may be a nurse or a volunteer in a rural or urban area to make a measure of risk of IHD for the P. This will reduce the time a Doctor has to spend in analyzing the risk a person. The Figure 2 shows the Neural network in the output screen.
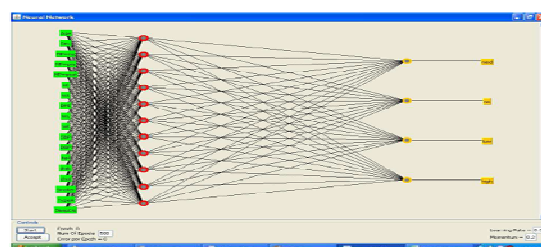
The number of epochs is set to 500. Error is 0.0326. Learning rate and momentum are set to 0.3 and 0.2 respectively. To improve the classification process, in future work the three stages can be separately fed to three different neural networks and the results may be combined to a net.

Through this work, a patient may be directed to an expert in cardiology if the patient is found to be of High Risk or medium risk for immediate treatment or follow up procedure. A patient of low risk cadre may be suggested to have controlled food habits, smoking habits and also may be directed to meet a doctor for further guidelines. According to the Planning Commission, India is short of 600,000 doctors and 1 million nurses -for every 10,000 Indians, there is barely one doctor available. [22]. As doctors especially cardiologists are less in number we should develop a system to utilize the resource efficiently.

IV. Conclusion
In this paper, we have proposed a Clinical Decision Support System (CDSS) for reliable heart disease risk classification using Artificial Intelligent techniques. The system is designed for Indian Population. Coronary Heart Disease or Ischemic Heart Disease can be handled successfully if more research is encouraged in this area. As this approach focuses on CAD Risk analysis, for a sample population, future work may be directed for further analysis.

REFERENCES

http://timesofindia.indiatimes.com/city/delhi/TOI-campaign-against -heart-disease-a- success/articleshow/7193581 .cms

Pedoe,H.T (1982). In Epidemiology of Diseases, D.I.. Miller and R.T.D. Farer(eds),Blackwell, Oxford

Rose,G. (1982) European Heart J.3, Suppl B.

Dewan, B.D. et al (1974),Indian heartj.,26:68

Sinha,B.C (1970). Jr. Ind. Med. Assoc,55 : 171.

Slone, D. et al(1978). N. Eng. J. Med. 298:1273

Sharper A.G. et. Al (1981). Brit. Med. J. 283:179

Bain, C. et. Al(1978). Lancet, 1:1087.

WHO (1985). Primary Prevention of CHD EURO Rep and Studio 98. Copenhagen

Keys, A. 91980). Seven Countries : a multivariate analysis of death and CHD, Harvard University Press, Cambridge, M.A

WHO(1985). Tech. Rep. Sen, 727.

Miller, N.E.et.al (1979) Lancet, 1:111.

Jenkins, C.D.et a(1974). N. Eng. J. Med., 290:1271.

Third Report of the National Cholestrol Education Program (NCEP) [2002].

Latha, P. and Subramanian, R. (2008) 'Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm', International Journal of Biological and Medical Sciences, Vol. 3, No. 3.

Wasan, S. K., Bhatnagar, V. and Kaur, H. (2006) "The Impact of Data Mining Techniques on Medical Diagnostics', Data Science Journal, Vol. 5, pp. 119-126.

Shahina B., Ahmed, M. U. and Funk, P. (2009) 'Case-based systems in health sciences: a case study in the field of stress' WSEAS Transactions on Systems, Vol. 8, No. 3, pp. 344-354.

Simon haykin "Neural networks", Prentice Hall 2nd edition.

S.N Sivanandam, S.Sumathi, S.NDeepa "Introduction to Neural Networks using MATLAB 6.0' TataMc-GrawHill Companies.

D.E., Rumelhart, G.E. Hinton, and R. J. Williams. Learning internal representations by back-propagating errors. Nature 323:533—536, 1986 (reprinted in J. A. Anderson, and E. Rosenfeld, editors, Neurocomputing.MIT Press, Cambridge, 1988).

Abdi, H. (1994). A neural network primer. Journal of Biological Systems, 2(3), 247-283.

hthp:/Avww.medicalbuyer.co.in/2007/telemecicine-an-optimistic -outlook- 3045-41.html