

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)**SciVerse ScienceDirect**

Procedia Engineering 41 (2012) 1818 – 1823

**Engineering  
Procedia**[www.elsevier.com/locate/procedia](http://www.elsevier.com/locate/procedia)

International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012)

## Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks

K.Rajeswari <sup>1\*</sup>, Dr.V.Vaithyanathan<sup>2</sup>, Dr. T.R. Neelakantan<sup>3</sup><sup>1</sup>Research Scholar, School of Computing, SASTRA University, Thanjavur 613 401, India<sup>2</sup>Associate Dean and CTS Chair Professor, School of Computing, SASTRA University, Thanjavur 613 401, India<sup>3</sup>Associate Dean and L&T ECC Chair Professor, School of Civil Engineering, SASTRA University, Thanjavur 613 401, India

---

### Abstract

Feature Selection in Data Mining refers to an art of minimizing the number of inputs under evaluation. An artificial neural network is the simulation of a human brain which learns with experience. Efficiency of a model or a system in terms of cost, time and accuracy will greatly improve if proper features of a system are selected. This proposed method uses Artificial Neural Network for selecting the interesting or important features from the input layer of the network. A Multi Layer Perceptron Neural Network is used for selection of interesting features from a Ischemic Heart Disease (IHD) data base with 712 patients. Initially the number of attributes was 17 and after feature selection the number of attributes was reduced to 12. All combination of features are attempted as inputs of a Neural Network. When the input features is 12 the predicted accuracy during training is high as 89.4% and during testing is high as 82.2 %.. Further removal of features lowers the accuracy and hence the interesting features selected for prediction is concluded to be as 12 for this IHD data set.

© 2012 The Authors. Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Centre of Humanoid Robots and Bio-Sensor (HuRoBs), Faculty of Mechanical Engineering, Universiti Teknologi MARA”

Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

*Keywords:* Data Mining, Feature Selection, Multi Layer Perceptron, Neural Network, Ischemic Heart Disease (IHD)

---

### 1. INTRODUCTION

IHD is a cardiovascular disease (CVD) which is alarmingly on rise throughout the world. Especially Indian population with IHD and death due to this is dangerously increasing. According to [1], CVDs will be the largest cause of death and disability by 2020 in India. In 2020, 2.6 million Indians are predicted to die due to coronary heart disease which constitutes 54.1 % of all CVD deaths. Nearly half of these deaths are likely to occur in young and middle aged individuals (30-69 years). Hence research on IHD has become significant. This paper focuses on reducing the number of features examined for IHD identification thereby reducing the doctor's time, patient's time and cost. The research objectives are (1) to observe neural networks, especially back propagation neural networks (2) to study about IHD and (3) to try various combinations of parameters or features for IHD identification. The organization of the paper is as follows. Chapter 2 discusses briefly about back propagation algorithm. Chapter 3 discusses about IHD and the parameters used to identify IHD. Chapter 4 discusses about methods and materials used for the research. Chapter 5 discusses about the findings and Chapter 6 gives the conclusion and scope for further research.

\* K.Rajeswari. Tel. : +919766909526 ;  
E-mail address: [raji.pccoe@gmail.com](mailto:raji.pccoe@gmail.com)

## 2. BACKPROPAGATION ALGORITHM

The feed forward neural network trained with back propagation algorithm is attempted to identify if a person is affected by IHD or not and the presence of IHD is further classified as high, medium and low risk levels. Backpropagation is a neural network learning algorithm. A neural network has a set of connected input and output units through hidden layer units in which each connection has a weight associated with it as shown in Fig. 1[13].

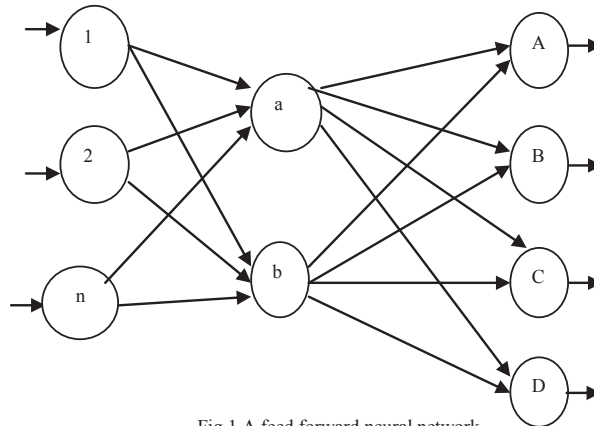


Fig 1.A feed forward neural network

There are two phases

1. Learning phase where the network learns by modification of weights.
2. Testing phase where an unknown input is tested for proper learning of neural network.

In Fig.1, the nodes 1, 2, 3,.., n represents the input nodes of input layer. These n nodes represent the features or parameters determining IHD. Nodes a,b represents nodes of hidden layer. Nodes A, B, C, D represents nodes of output layer.

The Back-propagation algorithm [13] is as follows.

### Input

- The training samples
- The learning rate  $\eta$
- A multilayer feed – forward network
- Network

### Method

Initialize all weights and biases in network

**While** terminating condition is not satisfied

{

**for** each training sample X in **samples**

{

// **Propagate the inputs forward**

- for each hidden or output layer unit j

{

$$I_j = \sum_i w_{ij} O_i + \theta$$

$$O_j = 1 / 1 + e^{-I_j}$$

// compute the output of each unit j.

- // Back propagate the errors
- for each unit j in the output layer
- $Err_j = O_j (1 - O_j) (T_j - O_j)$  ;
- // Compute the error

```

▪ for each unit j in the hidden layers , from the last to the first hidden layers
▪ Errj = Oj ( 1- Oj ) Σk Errk wjk;
  //Compute the error wrt the next higher layer , k
▪ for each weight wij in network {
▪ Δ wij = (1)Errj Oi; // weight increment
▪ wij = wij + Δ wij ; } // weight update
▪ for each bias θj in network {
▪ Δ θj = (1)Errj ; // bias increment
▪ θj = θj + Δ θj ; } // bias update
}}

```

Fig 2. Backpropagation Algorithm

### 3. Ischemic Heart Disease

A large body of data pertaining to the IHD patients is available. 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 where the prevalence is 65.4 and 47.8 per 1000 males and females respectively in urban population [2]. In epidemiology, the prevalence of a disease in a statistical population is defined as the total number of cases of the disease in the population at a given time, or the total number of cases in the population, divided by the number of individuals in the population [15]. Our study was carried out on a database collected from Madras Medical College, Chennai in India for the age group 30 and above.

Indians are more prone to heart disease that lead to worse outcomes like IHD a condition characterized by reduced blood supply to the heart [14]. The pattern of IHD in India has been reported to be as follows:

- In India, IHD appears in earlier ages compared with developed countries. The tough period is attained between 51-60 years of age
- 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 [3][4][5][6].
- A family history of IHD. Children of parents with heart disease are more likely to develop it themselves.
- Sedentary life style, a casual life with no physical activity [10].
- Type A individuals who are time conscious, tightly bound to job and are restless [11].
- Higher alcohol intake, defined as 75g [7] Research has revealed an association between moderate alcohol consumption and lower risk for CHD [16][17].

### 4. Materials and methods

#### 4.1. Data Preparation

The following data were collected and analyzed for Indian heart risk score prediction based on extensive study and expert opinions. from doctors with respect to Indian body conditions, life style and eating habits. After discussion with cardiologists a three-stage questionnaire was prepared. Diagnosis was done through data collection for each individual patient as given in tables 1,2 and 3.

Stage I includes physical examination parameters. Stage II includes Co Morbid features collection and Stage III includes attributes on personal habits and hereditary.

Table 1 Stage one diagnosis

Stage 1						
Age	Gender [ F – Female M- Male]	Menopause [0 for pre menopause]	Ht in cm	Wt in kg	Body Mass Index in kg/m2	Waist Measure in cm
30	F	0	165	65	19.69	75

If sex is female, pre or post menopause details are recorded. Waist circumference is noted as it is proved to be an important risk determinant factor for CHD[12].

Table 2 Stage two diagnoses

Stage 2				
Co Morbid Factors				
SBP	DBP	Diabetes	Cholesterol	Thyroid
120	80	1	1	0
140	90	0	1	0

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

Table 3 Stage three diagnoses

Stage 3				
Personal Habits	Family History	Genetic Factors	Type A Personality	Sleeping Disturbance
1	0	0	1	1
0	1	1	1	1

Based on the data collected, the immediate risk analysis was classified as ‘No Risk’, ‘Low Risk’, ‘Medium Risk’ or ‘High Risk’. Table 4 data is collected from experienced doctors in cardiology based on the importance of every feature collected in tables 1,2 and 3. Three expert opinions were collected. Two identical opinions were taken into consideration for deriving conclusion. Varied opinion data of a patient was removed from the dataset.

Table 4. Output

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

#### 4.2. Feature Selection

In the IHD 17 features from 712 patients were collected. After collection of these features Expert doctor’s opinion on Risk of Ischemic heart disease was obtained after ECG was taken. Multilayer perception network was used. In the dataset 70% was used for training and 30 % for testing. When all the features were used for training, the accuracy was 70 % during training and 69 % during testing. The next step started with removing the first feature F1 which is Age and the accuracy during training fell to 63 % and during testing was 58 %. Similarly as in Table 1, every single feature was removed subsequently and the accuracy during training and testing period was noted. During the first iteration it was noted that the accuracy was worse when the following attributes were removed

1. Age
2. Gender
3. Menopause
4. Diabetes
5. Cholesterol
6. Smoker/Alcohol.

Hence according to our data set, we can conclude in iteration 1- the above attributes contribute for identifying IHD. Table 5 below shows the accuracy obtained when the features F1 to F17 are removed one by one. The description of Features are given in the first column

Table 5. Sample pruning

	~F1	~F2	~F3	~F4	~F5	~F6	~F7	~F8	.	.	~F17
Age	X										
Gender		X									
Menop			X								
Ht				X							
Wt					X						
BMI						X					
WC							X				
SBP								X			
DBP									X		
Diab											
Chol											
Thyroid											
Personal habits											
Family History											
Genetic											
Type A											
Sleep											X

During the second iteration, every time two features were removed. Here at a time two features in combination are taken out and the accuracy was checked. For example {F1,F2} are removed, followed by {F1,F3}, {F1,F4}, {F1,F5},.....{F2,F3}, {F2,F4}, {F2,F5} ..... {F3,F4}, {F3,F5}, {F3,F6},..... {F4,F5}, {F4,F6}, {F4,F7},..... {F5,F6}, {F5,F7}, {F5,F8},..... {F16,F17} is checked for accuracy. We found that accuracy was worse when the following pairs were removed.

$$\{F1, F2\}$$

Thus second iteration identified Age and Gender as important features.

During the third iteration, three features at a time were removed. For example: F1, F2, F3 together were removed from the list of independent variables or factors or as inputs. Accuracy is noted. The process continues with 4 feature removal, 5 feature removal, and when 6 features are removed accuracy is worse with any combination. At this stage we stop removing further combination of features.

**5. Results and Discussion.**

The prevalence of chronic disease determinants like smoking, diet, excersise lag, stress, sleeplessness and their immediate outcomes like increase in hyper tension has accelerated urban Indians experiencing heart attacks with a nearly a

fifth of hospital – based patients being less than 40 years of age [18- 22]. This work has used Back propagation Neural Network to classify the risk measure based on the 17 inputs obtained. The Feed forward Neural Network has 17 data inputs and 4 outputs. These 17 variables were taken for study after discussion with experts and after reviewing various literature works. The Software SPSS tool is used for training and testing the data set with the different combination of features. It is concluded that the following 12 features are the important inputs to determine the risk level of IHD.

{Age, Gender, Menopause, Body Mass Index, Waist Circumference, Systolic Blood Pressure, Diastolic Blood Pressure , Diabetes , Cholesterol, Hereditary, Personal habits, Type A}.

The accuracy is a measure of sensitivity and specificity. Accuracy is check using the Feed Forward Neural Network classifier after removal of every feature. The accuracy of classifier is found to be the best with above listed 12 features.

## 6. Conclusion

This paper focuses on feature reduction using Artificial Neural Networks. As the medical field is with tens or hundreds of thousands of features or attributes, feature reduction will be very much useful. The 17 features designed as determinant features for Ischemic Heart Disease is reduced to 12 and the accuracy is checked in Madras Medical College. The accuracy is found to be excellent. This work can be extended to different localities to find its validity with different population. Also in our future work, statistical relevance between attributes will be analyzed . Variable ranking method using correlation coefficient or mutual information will be made.

## ACKNOWLEDGMENT

The Authors thank Dr. P. Amirtharaj, Cardio Thoracic Surgeon, Madras Medical College for his excellent guidance throughout the work.

## REFERENCES

- [1] World Health Organization . The World Health Report 2002. Geneva, Switzerland: WHO, 2002.
- [2] Dewan, B.D. et al (1974), Indian heart j., 26:68.
- [3] I Sinha,B.C (1970). Jr. Ind. Med. Assoc, 55: 171.
- [4] Slone, D. et al(1978). N. Eng. J. Med. 298:1273.
- [5] R Sharper A.G. et. Al (1981). Brit. Med. J. 283:179.
- [6] Bain, C. et. Al(1978). Lancet, 1:1087.
- [7] WHO (1985). Primary Prevention of CHD EURO Rep and Studio 98. Copenhagen.
- [8] Keys, A. 91980). Seven Countries : a multivariate analysis of death and CHD, Harvard University Press, Cambridge, M.A.
- [9] WHO (1985). Tech. Rep. Ser., 727.
- [10] Miller, N.E. et.al (1979) Lancet, 1:111.
- [11] Jenkins, C.D.et a(1974). N. Eng. J. Med., 290:1271.
- [12] Palla Venkataramana, Palakuru C Reddy (2002). ‘Association of overall and abdominal obesity with coronary heart disease risk factors : comparison between urban and rural Indian men’, Asia pacific Journal Clin Nutr 11(1):66-71
- [13] Jiawei Han and Micheline Kamber, “DataMining Concepts and Techniques”, Elsevier Second Edition
- [14] <http://timesofindia.indiatimes.com/city/delhi/TOI-campaign-against-heart-disease-a-success/articleshow/7193581.cms>
- [15] <http://en.wikipedia.org/wiki/Prevalence>
- [16] <http://pubs.niaaa.nih.gov/publications/aa45.htm>
- [17] Hennekens, C.H. Alcohol and risk of coronary events. In: Zakhari, S., and Wassef, M., eds. Alcohol and the Cardiovascular System. NIAAA Research Monograph No. 31. NIH Pub. No. 96-4133. Washington, DC: U.S. Govt. Print. Off., 1996. pp. 15-24.
- [18] Reddy KS. Rising burden of cardiovascular disease in Indians. In: Sethi KK, editor. Coronary artery disease in Indians. A global perspective. Mumbai: Cardio logical Society of India: 1998
- [19] Malhotra S, Gupta M, Chandra KK, Grover A, Pandhi.P Prehospital delay in patients hospitalized with acute myocardial infarction in the emergency unit of a North Indian tertiary care hospital. Indian Heart J 2003; 55:349-53
- [20] Siwach SB, Singh H, Sharma D, Katyal VK. Profile of young acute myocardial infarction in Haryana. J Assoc Physicians India 1998; 46:424-6.
- [21] Yavagal ST, Rangarajan R, Prabhavathi. Clinical profile of acute myocardial infarction in women. J Indian Med Assoc 1994; 92:283-4.
- [22] Anand SS, Pais P, Pogue J, Yusuf S. A comparison of practice patterns for acute myocardial infarction between hospitals in Canada and India. Indian Heart J 1997; 49: 35-41.