International Journal of Bio-Science and Bio-Technology
Vol.8, No.4 (2016), pp. 139-148
http://dx.doi.org/10.14257/ijbsbt.2016.8.4.16

Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review

J.Vijayashree and N.Ch.SrimanNarayanaIyengar*

School of Computer Engineering, VIT University, Vellore-632014, TN, India
vijayashree.j@vit.ac.in;nchsniyr@vit.ac.in

Abstract

Heart disease is one of the main sources of demise around the world and it is imperative to predict the disease at a premature phase. The computer aided systems help the doctor as a tool for predicting and diagnosing heart disease. The objective of this review is to widespread about Heart related cardiovascular disease and to brief about existing decision support systems for the prediction and diagnosis of heart disease supported by data mining and hybrid intelligent techniques .

Keywords*: Cardiovascular Disease, Decision Support System, Data Mining, Hybrid Intelligent System*

1. Introduction

The heart is a most significant muscularorgan in humans, which pumps blood through the blood vessels of the circulatory system[l]. Human life is dependent on the proper functioning of heart. Improper functioning of heart will influence other parts of human body like brain, kidney *etc*. If the blood circulation in body is inefficient, it affects both heart and brain. Generally blood arrest in heart is called as attack and blood arrest in brain is called as stroke. Human life is absolutely reliant on the efficient working of the heart and brain. The rest of this paper is presented as follows: Section 2 describes the cardiovascular disease and its pervasiveness. Section 3 describes the advantage of decision support system for the prediction of heart disease. Section 4 and 5 describes various data mining and hybrid intelligent techniques used for the prediction of heart disease.

2. Cardiovascular Disease

Cardiovascular heart disease is one of the principal reasons of death for both men and women. The term heart disease relates to a number of medical conditions related to heart which define the irregular health conditions that directly stimulate the heart and all its parts.Different types of heart related cardiovascular diseases along with description are given in Table1.

Table 1. Types of Cardiovascular Diseases

Heart-related cardiovascular diseases	Description
Acute coronary syndromes	Blood-supply to the heart muscle is swiftly obstructed
Angina	Chest pain due to a lack of blood to the heart muscle
Arrhythmia	Atypical heart rhythm
Cardiomyopathy	Heart muscle disease
Congenital heart disease	Heart disfigurements that are present at birth
Coronary heart disease	Arteries supplying blood to heart muscle becomes

ISSN: 2233-7849 IJBSBT

	obstructed
Heart failure	Heart is not propelling ample blood
Inflammatory heart disease	Tenderness of the heart muscle and/or the tissue
Ischaemic heart disease	Plaque builds up inside the coronary arteries
Rheumatic heart disease	Rheumatic fever
Valvular disease	Disease of the valves

Various risk factors along with its symptoms that contribute to heart attack are presented in Table2.

Table 2. Risk Factors and Symptoms of Heart Attack

Risk factors	Symptoms of Heart Attack
• Age • Angina • Blood cholesterol levels • Diabetes • Diet • Genes • Hypertension • Obesity • Physical Inactivity • Smoking • Work stress	• Chest Discomfort • Coughing • Nausea • Vomiting • Crushing chest pain • Dizziness • Dyspnoea (shortness of breath) • Restlessness

2.1. Prevalence of Heart Disease

According to World life expectancy, India ranked 39th position of all the countries in the world suffering from coronary heart disease. In India the death rate per 100,000 is 138.36. Population suffering from Coronary heart disease in India by age, gender and region is presented in table 3.

Table 3. Coronary Heart Disease in India by Age, Gender and Region

Year/Age	20-29	30-39	40-49	50-59	Male	Female	Urban	Rural
2000	4.51	5.48	6.11	5.81	12.9	14	12.3	14.7
2005	6.15	7.25	8.33	7.71	17.1	18.7	17.8	18
2010	8.31	9.09	10.9	10	22.4	24.5	24.6	22.2
2015	10.4	12.4	14.3	13	28.7	32.7	36	25.4

From the above table it is obvious that in country like India female suffer from coronary heart disease more than men. Till the year 2010 the population suffering from coronary heart disease in the rural areas of India is more compared to urban population, whereas from the year 2010 onwards it is vice versa. In 2015 there is a drastic variance in the population suffering from CHD in rural and urban regions.

In India population of age group between 40 and 49 suffer profoundly from the heart disease. The population suffering from heart disease in all age groups has doubled in last fifteen years. The following figure illustrates Coronary heart disease in India by age, gender and region.

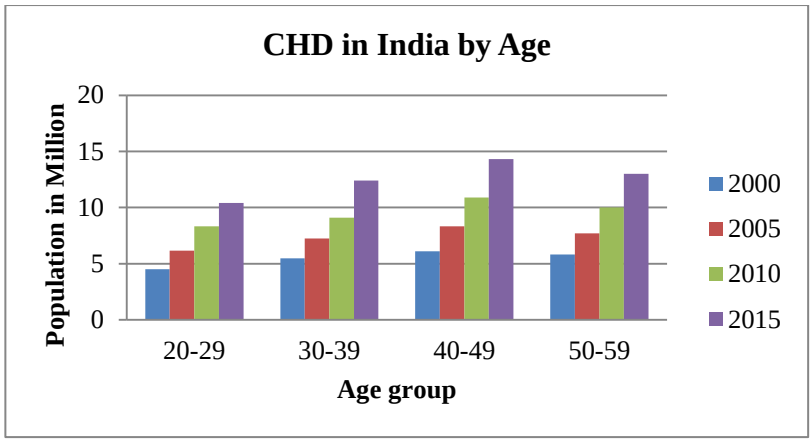

Figure 1. Age Wise Coronary heart disease in India

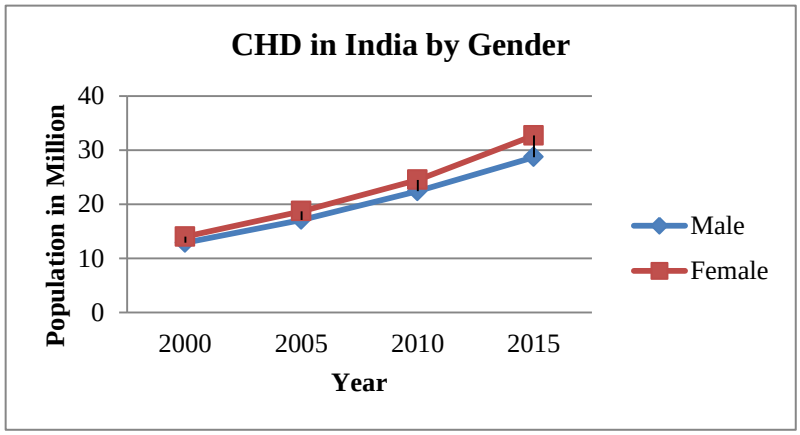

Figure 2. CHD in India based on Gender

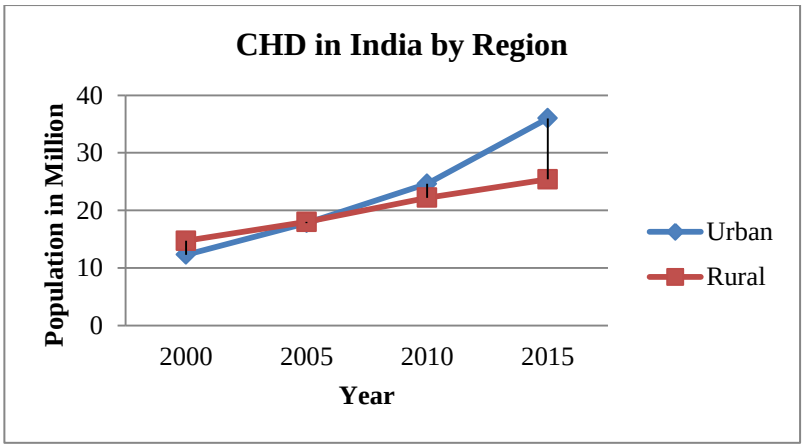

Figure 3. CHD in Urban and Rural India

3. Decision Support System for Heart Disease Prediction and Diagnosis

Medical diagnosis can be improved by the use of computer-based systems and algorithms taking decisions at the appropriate stages. Such systems are called decision support systems (DSSs). Intelligence also plays a role here. These systems help to predict and diagnose the disease based on the patient information and domain knowledge. DSS

helps in improving the quality of healthcare by providing an effective and reliable diagnosis. DSS can decrease the cost of treatment by providing a more specific and faster diagnosis efficiently and also the time is reduced compare to traditional procedures. Once placed in cloud any health organization can utilize these services.

3.1. Knowledge Discovery in Database

Decision support systems was developed using a knowledge base. Knowledge discovery in database uses data mining process which extracts useful information from data set and transforms it into a reasonable structure for further use. Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases [22]. [19] Defines data mining as "a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database". Data mining uses two strategies: supervised and unsupervised learning. A training set is used to learn the model parameters in supervised learning whereas no training set is used in unsupervised learning.

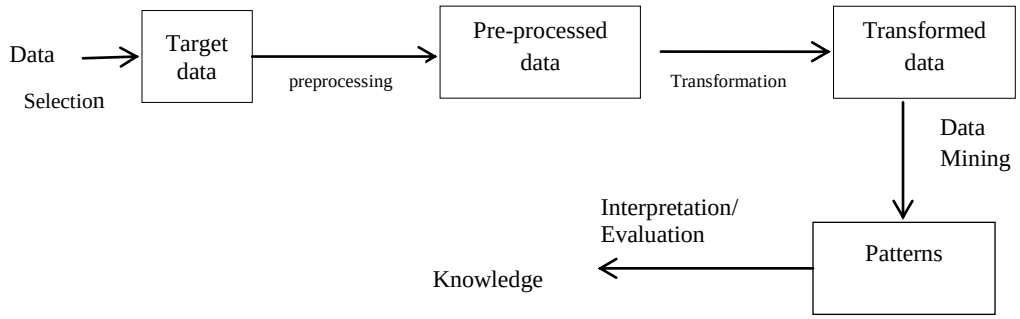

Figure 4. Basic Methodology for Knowledge Discovery in Database

3.2 Data Mining Algorithms

3.2.1. Neural Networks (NN)

Neural network is a parallel, distributed information processing structure consisting of numerous quantities 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 and each conveys the equivalent signal. The NN can be classified in two main groups according to the way they learn. They are supervised learning and unsupervised learning.

In supervised learning the network compute a response to each input and then compares it with the target value. If the computed response differs from the target value, the weights of the network are adapted according to a learning rule. Examples of supervised learning are Single-layer perceptron and Multi-layer perceptron. In unsupervised learning the networks learn by identifying special features in the problems they are exposed to. Example for unsupervised learning is self-organizing feature maps.

3.2.2. Naïve Bayesian Classifier

Naïve Bayes [1] is a classification algorithm based on Bayes theorem, which calculates a probability by counting the frequency of values and combination of values n historical data. Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred.

$$Pro(B \text{ given } A) = Pro(A \text{ and } B)/Pro(A)$$

Advantage of this algorithm is it requires only a small amount of training data for estimating the parameters essential for classification.

3.2.3. Decision Tree

Berry and Linoff defined decision tree as "a structure that can be used to divide up a large collection of records into successive smaller sets of records by applying a sequence of simple decision rules. With each successive division, the members of the resulting sets become more and more similar to one another". ID3(Iterative Dichotomiser 3) is one of the decision tree models which builds a decision tree from fixed set of training instances. C4.5 is the latest version of ID3 induction algorithm, C5.0 is an extension of C4.5 decision tree algorithm and J48 decision tree is the implementation of ID3 algorithm.

3.3. Genetic Algorithm

In Artificial Intelligence, Genetic Algorithm is a search technique which uses the process of natural selection. Genetic algorithms provide solutions to optimization and search problems by using techniques such as inheritance, mutation, selection, and crossover. A typical genetic algorithm requires genetic representation of the solution domain and a fitness function to evaluate the solution domain.

4. Surviving Techniques for Heart Disease Prediction using Data Mining Techniques

A Web based clinical decision support system which uses medical profiles like age, blood pressure, *etc*. is proposed to predict the prospect of patients attaining heart disease [1]. Naïve Bayes Data Mining algorithm answers complex what-if queries. The system is implemented on PHP platform which is ascendable, steadfast and expandable.

A survey [3] on different data mining techniques used for the prediction of heart disease and found hybrid approaches as best prediction model compared to single model by comparing previous researcher's findings .

Using accuracy and sensitivity [4] as measures, author evaluated three data mining techniques such as Decision Tress (C4.5), neural networks (MLP) and Naïve Bayes. Results illustrate that with more number of attributes neural networks and Naïve Bayes perform very responding than Decision tree. Results also illustrate that neural networks perform the best with a classification accuracy of 0.897 and with sensitivity of 0.9017.

The efficacy of different decision tree algorithms used for classification and prediction of heart disease such as ID3, C4.5, C5.0 and J48 has been scrutinized [5]. ID3 is one of the decision tree models which build a decision tree from fixed set of training instances. C4.5 is the latest version of ID3 induction algorithm, C5.0 is an extension of C4.5 decision tree algorithm and J48 decision tree is the implementation of ID3 algorithm. Author also conferred attribute selection measures or the split criteria: Information gain, Gini Index & Gain ratio and performance evaluation measures: Sensitivity, specificity & accuracy.

The heart dataset contains large volumes of which consumes more time for classification so by using attribute selection methods the dimensionality of data is reduced. In both cases Naïve Bayes classification technique produced enhanced results. This is observed when the performance of four classification algorithms: Naïve Bayes, Decision Tree, K-NN and Neural Network are investigated on complete heart disease dataset and reduced dataset [7].

Table 4.Comparison Heart Disease Prediction System using Data Mining Classification Techniques

Reference	Data Mining Techniques Compared	Accuracy Obtained		Number of Attributes used	Result: Best technique
Purusothaman G *et al* (2015)	Single data mining models: Decision Tree	76%			Hybrid model
	Associative Rules	55%			
	K-NN	58%			
	Artificial Neural Networks	85%			
	Support Vector Machine	86%			
	Naïve Bayes	69%			
	Hybrid models	96%			
Srinivas K *et al* (2010)	Decision Tress (C4.5)	82.5%		15	Neural networks (MLP)
	Neural networks (MLP)	89.75			
	Naïve Bayes	82%			
	SVM	82.5%			
Chaitrali S *et al* (2012)	Decision Trees	96.66%	99.62%	13 &15	Neural networks
	Naive Bayes	94.44%	90.74%		
	Neural Networks	99.25%	100%		
John Peter T *et al* (2012)	Naïve Bayes	83.70%		13	Naïve Bayes
	Decision Tree	76.66%			
	K-NN	75.18%			
	Neural Network	78.485			
Hlaudi DM *et al* (2014)	J48	99.0741%		11	J48, REPTREE and SIMPLE CART algorithm
	Bayes Net	98.148%			
	Naive Bayes	97.222%			
	Simple Cart	99.0741%			
	REPTREE	99.0741%			
Gnanasoundhari SJ *et al* (2014)	Naive Bayes	52.33%			Weighted Associative Classifier
	Neural network	78.43%			
	Weighted Associative Classifier	81.51%			
	Support Vector Machine	60.78%			
Anbarasi M *et al* (2010)	Naive Bayes	96.55%		6	Decision Tree
	Classification by clustering	88.3%			
	Decision Tree	99.2%			

Five data mining algorithms: J48, Bayes Net, and Naive Bayes, Simple Cart, and REPTREE are considered [8] for diagnosing heart disease, 11 attributes and Waikato Environment for Knowledge Analysis (WEKA) tool is used for prediction. J48, REPTREE and SIMPLE CART algorithm are evidenced to be best.

A survey four data mining classifiers: Naive Bayes, Neural network, Weighted Associative Classifier (WAC) algorithm and Support Vector Machine (SVM) was made [9]and used the same for predicting the heart disease with condensed number of attributes. Results demonstrate that WAC affords more accurate results in predicting the heart disease.

A three dimension survey conducted depending upon the type of dataset [10] for dataset consisting of labelled features classification model is appropriate and for unlabelled features clustering model is appropriate and to increase the performance of dataset with more optimization, bio-inspirational based techniques are appropriate. Amongst the four classification models decision tree, ID3, SMV and neural networks, support vector machine is exceedingly used by researchers. Amongst the three clustering models K-means, Fuzzy C-means and hierarchical clustering, K-means is exceedingly used by researchers. Amongst the three bio-inspirational based techniques Ant Colony Optimization, Artificial Immune system and Particle swarm optimization, Particle swarm optimization is exceedingly used by researchers.

Classification techniques such as logistic regression (LR), decision trees, and Artificial neural networks (ANNs) performance is compared to predict the patient's attainment heart disease[11] Using lift chart and error chart performance is compared. Results demonstrate that artificial neural networks have the least of error rate and have the highest accuracy. Thus artificial neural networks are the best method to classify and predict the heart disease.

The presence of heart disease identified using three classifiers like Naive Bayes, Classification by clustering and Decision Tree [12]. Author reduced the number of attributes from 13 to 6 using genetic algorithm. Interpretations demonstrate that decision tree outperforms Naive Bayes, Classification by clustering technique after integrating feature subset selection with moderately high model construction time. Naïve Bayes achieves consistently before and after reduction of attributes with the same model construction time. Classification via clustering achieves poor compared to other two methods. Author intends to extend the work for predicting the intensity of the disease using fuzzy methods.

Data mining techniques such as classification, clustering, fuzzy system and association rules are studied and investigated [13] for the prediction of heart disease.

5. Surviving Techniques for Heart Disease Prediction using Hybrid Intelligent Techniques

Genetic algorithm and Fuzzy logic techniques are used for predicting heart disease where feature selection is done by Genetic algorithm and Classification and prediction is done by fuzzy logic [2]. Authors compares the performance of proposed method (GAFL system) with fuzzy entropy based method (NNTS) using the metrics like accuracy, specificity and sensitivity. Number of features is reduced from 13 to 7 and the accuracy of the proposed method is 86%.

Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm is presented in [14]. This model combines neural network, fuzzy logic and genetic algorithm. The proposed model improves training performance and classification accuracy.

Authors introduced a new classification approach for the classification of heart disease, which uses Artificial Neural Network and feature subset selection in [15]. Feature subset selection reduces the number of attributes. Pre-processing is through using Principal Component Analysis (PCA). Results demonstrate that the proposed approach indicates enhanced accuracy over traditional classification techniques.

The objective of [16] is to used genetic algorithm for determining the weights of neural networks and to evaluate two types of learning algorithms namely Feed forward neural

network algorithm and Fitting algorithm. Author validated the accuracy to be 97.75% and improvement of Feed-forward and Fitting neural network to be 1.29% and 1.37% by applying datasets on Feed Forward ANN model, Fitting ANN model and GA trained Feed Forward ANN model, GA trained Fitting ANN model.

[17] Implemented a system for predicting the heart disease using Data mining techniques: K Means and Weighted Association rule. Results demonstrate that K-means with decision tree technique make the system more accurate and efficient compared to the weighted association rule with Apriori algorithm.

Heart disease prediction system using three data mining classification techniques (Decision Trees, Naive Bayes, Neural Networks), two more attributes: Obesity and smoking along with consistent 13 attributes are used in [6]. J48 decision tree algorithm which uses pruning method for building a tree and data mining tool Weka 3.6.6 are used. Results show that neural networks produce accurate results by comparing the accuracy of classification techniques with 13 and 15 input attributes. A Multi-layer Perceptron Neural Networks (MLPNN) is used for improving the accuracy.

Artificial neural networks with back propagation error method are used to classify the cerebrovascular disease [18]. The neural network was trained with 16 input attributes using back propagation algorithm with sigmoid function on one hidden layer which improves the accuracy.

6. Conclusion

Many DSS exists to predict the heart disease with various techniques. The World life expectancy statistics implies that heart disease is prevailing more in number. So it is necessary to build an efficient intelligent trusted automated system which predicts the heart disease accurately based on the symptoms according to gender/age and domain knowledge of experts in the field at the lowest cost.

References

[1] D. Ratnam, P. HimaBindu, V. MallikSai, S. P. Rama Devi and P. RaghavendraRao, "Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve BayesAlgorithm", International Journal of Computer Science and Information Technologies, vol. 5, no. 2 **(2014)** pp.2384-2388.

[2] T. Santhanam and E. P. Ephzibah, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model", Indian Journal of Science and Technology, vol.8, no. 9, **(2015)**, pp.797–803.

[3] G. Purusothaman and P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease",Indian Journal of Science and Technology, vol. 8, no. 12, **(2015).**

[4] K. Srinivas, G. RaghavendraRao and A. Govardhan, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", Proceedings of 5th International Conference on Computer Science & Education, China, **(2010)**, pp. 24–27.

[5] K. Thenmozhi and P. Deepika, "Heart Disease Prediction Using Classification with Different Decision TreeTechniques", International Journal of Engineering Research and General Science, vol. 2, no. 6, **(2014)**, pp.6-11.

[6] S. Chaitrali, Dangare and S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, vol. 47, no. 10, **(2012)**, pp. 0975 –888

[7] J Peter and K. Somasundaram, "An Empirical Study on Prediction of Heart Disease Using Classification DataMining Techniques", Proceedings of IEEE International Conference on Advances In Engineering, Science And Management (ICAESM), **(2012)**, pp. 514-518.

[8] H. D. Masethe and M. A. Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science(WCECS), San Francisco, USA. **(2014)**.

[9] S. J. Gnanasoundhari, G. Visalatchi and M. Balamurugan, "A Survey on Heart Disease PredictionSystem Using Data Mining Techniques", International Journal of Computer Science and Mobile Application, vol. 2, no. 2 **(2014)**, pp. 72-77

[10] K. Lokanayaki and A. Malathi, "Exploring on Various Prediction Model in Data Mining Techniques for Disease Diagnosis", International Journal of Computer Applications, vol. 77, no. 5, **(2013)**, pp. 0975 – 8887.

[11] A. Khemphila and V. Boonjing, "Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients",Proceedings of International conference on Computer Information Systems and Industrial Management Applications (CISIM), **(2010)**, pp. 193-198.
[12] M. Anbarasi, E. Anupriya, Iyengar N.Ch.S.N, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology, vol. 2, no.10, **(2010)**, pp. 5370-5376.
[13] S. Vijayarani and S. Sudha, "A Study of Heart Disease Prediction in Data Mining", International Journal of Computer Science and Information Technology & Security (IJCSITS), vol. 2, no. 5, **(2012)**, pp. 2249-9555.
[14] L. Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Sciences, vol. 3, no. 3, **(2007**), pp. 157-160.
[15] M. AkhilJabbar, B. L. Deekshatulu, P.Chandra, "Classification of Heart Disease using Artificial Neural Networkand Feature Subset Selection", Global Journal of Computer Science and Technology Neural & Artificial Intelligence, vol. 13, no. 3 **(2013),** pp. 0975-4350.
[16] P. Gupta and B. Kaur, "Accuracy Enhancement of Heart Disease Diagnosis System Using Neural Network and Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 103, no. 13, **(2014)**, pp. 11-15.
[17] A. Wilson, G. Wilson and J. Likhiya, "Heart Disease Prediction using the Data Mining Techniques", International Journal of Computer Science Trends and Technology (IJCST), vol. 2, no. 1, **(2014)**, pp. 2347-8578.
[18] O. Olabode and B. T. Olabode, "Cerebrovascular Accident Attack Classification Using Multilayer Feed Forward Artificial Neural Network with Back Propagation Error", Journal of Computer Science, vol. 8, no. 1, **(2012)**, pp.18 – 25.
[19] U. Fayyad, "Data Mining and Knowledge Discovery in Databases: Implications for scientific databases", Proceedings of 9th International Conference on Scientific and Statistical DatabaseManagement, Olympia, Washington, USA, **(1997)**, pp. 2-11.
[20] www.searo.who.int/india/cardiovascular_diseases/Commision_on_Macroeconimic and Health
[21] www.worldlifeexpectancy.com
[22] www.world-heart-federation.org
[23] B. Thuraisingham, "A Primer for Understanding and Applying Data Mining, IT Professional, **(2000)**, pp. 28-31.

Authors

J. Vijayashree, She is working as Assistant Professor in Computer Application Division at VIT University, Vellore, Tamil Nadu, India. Her area of interest includes Data Mining and Artificial Intelligence.

N. Ch. S. N. Iyengar, He is a Professor, SCS Engineering at VIT University, Vellore, TN, India. His research interests include Distributed Computing, Information Security, Intelligent Computing, and Fluid Dynamics (Porous Media). He had much teaching and research experience with a good number of publications in reputed International Journals & Conferences. He chaired many Intl. Conf. delivered Key note lectures, served as PC Member/Reviewer. He is Editorial Board member for many Int'l Journals like *Int. J. of Advances in Science and Technology*, of SERSC, *Cybernetics and Information Technologies* (CIT)- Bulgaria, Egyptian Computer Science Journal -Egypt, IJCA & IJConvC of Inderscience -China, *etc*., Also Editor in Chief for International Journal of Software Engineering and Applications(IJSEA) of AIRCC, Advances in Computer Science (ASC) of PPH and Many more.