International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-4 Number-3 Issue-16 September-2014

Genetic Neural Approach for Heart Disease Prediction

Nilakshi P. Waghulde[1], Nilima P. Patil[2]

Abstract

Data mining techniques are used to explore, analyze and extract data using complex algorithms in order to discover unknown patterns in the process of knowledge discovery. Heart disease is a major life threatening disease that cause to death and it has a serious long term disability. The time taken to recover from heart disease depends on patient's severity. Heart disease diagnosis is complex task which requires much experience and knowledge. Nowadays, health care industry contain huge amount of health care data, which contain hidden information. Advanced data mining techniques along with computer generated information are used for appropriate results. Neural Network is widely used tool for predicting heart diseases diagnosis. A Heart Disease Prediction System is developed using Neural Network and Genetic Algorithm. This system calculates the number of hidden nodes for neural network which train the network with proper selection of neural network architecture and uses the global optimization of genetic algorithm for initialization of neural network. For prediction, the system uses 12 parameters such as sex, age, blood cholesterol etc. From the result, it is found that genetic neural approach predicts the heart disease upto 98% accuracy.

Keywords

Neural network, genetic algorithm, data mining

1. Introduction

Knowledge discovery in databases process consist of data mining as one of the most important steps and a significant subfield in knowledge management [1]. Data mining aids in healthcare to support for effective treatment, healthcare management, customer relation management, fraud and abuse detection and decision making.

Manuscript received August 28, 2014.
Nilakshi P. Waghulde, CSE Department, North Maharashtra University SSBT's COET, Jalgaon, Maharashtra, India.
Nilima Patil, CSE Department, North Maharashtra University SSBT's COET, Jalgaon, Maharashtra, India.

Research in medical area is not limited to healthcare development such as developing new healing techniques and drugs, but also there are healthcare informatics fields such as structured data entry, constructing longitudinal patient data, image processing etc. Creative results are achieved by using the data mining techniques. In healthcare the most commonly used data mining techniques are Artificial Neural Network, Decision trees, Genetic Algorithms, Nearest neighbor method, Fuzzy logic, Fuzzy based Neural Networks, Bayesian Networks and Support Vector Machines [2].Data mining is not an easy task to achieve because of the complexity and toughness of information in medical domain.

Nowadays, the rise in the healthcare cost leads to the world's most important problems. Due to increase in world's population, the health care industries are facing many challenges and issues based on patient's severity is to be reduced and detect it earlier in a more effective way. In order to reduce the health care cost, the hospitals are trying to improve and increase the efficiency of medical resource utilization [1].

The mortality rate caused by heart disease has been changing so there is need for development of methods for heart disease prediction is of immediate scientific and practical interest. There are several algorithms which have been already developed for risk stratification and diagnostic models for heart disease prediction. The heart disease prediction is based on different sets of risk factors, such as arterial hypertension, hypercholesterolemia, diabetes mellitus, and smoking [3] [4]. It seems that objective difficulties in heart disease detection are caused by the multiplicity of risk factors to be taken into consideration. Thus the survey of the structure of the variable risk factors is needed and an efficient classification system is to be created for the prediction of heart disease. Therefore, the development of algorithms for correct classification of heart disease risk factors is an important problem. Nowadays, expert medical systems have been created for this purpose based on the computer methods of intelligent data processing.

Predicting heart disease from various factor or symptoms is a multi-layered approach which leads to

false presumptions and unpredictable effects. The large amount of data is a key resource to be processed and analyzed for knowledge extraction that enables support for decision making. Neural Network which is able to train the data through multiple layers proves to be well performing technique for heart disease prediction [5]. It is possible for doctors to analyze, model, and make sense of complex clinical data across a broad range of medical applications with the help of artificial neural networks as it is a powerful tool for medical data mining [6] [7].

One of these promising method is artificial neural networks (ANNs) which emerge as a well performing technique for heart disease prediction [8], is a highly effective tool used in classification tasks, as well as to solve many important problems, such as signal enhancement, identification, and prediction of signals and factors. ANNs has an important feature as its adaptivity in complex information processing in data mining process. This makes it possible that the ANNs are applied in cases where there is impossible to create a strict mathematical model but has a sufficiently representative set of samples. The other important characteristic of neural networks is their capacity to generalize input information and to give correct answers for unfamiliar data, which makes them effective in solving complicated classification problems [6].

The major problem associated with the neural network is the selection of hidden neurons in structure of neural network [3].This is very important while the neural network is trained to get very small errors which may not respond properly in prediction. There exists an overtraining issue in the design of Neural Network training process. Over training is similar issue to the over fitting of data in the neural network. This issue is to be solved because the neural network trains the data by matching the data so closely as to lose its generalization ability over the test data. One of the searching algorithms which find an optimal solution to a problem is Genetic Algorithm [9].Genetic algorithm is based on Darwin theory about evolution "Survival of the fittest". Genetic algorithm are a way of solving problems by mimicking processes the nature uses-Selection, Crossover, Mutation and Accepting to evolve solution to the problem [10]. So when the two data mining techniques such as neural network and genetic algorithm combines helps to increase the accuracy of prediction.

The rest of the paper is structured as follows. Section 2, follows Related Work completed on the application of data mining techniques to heart disease prediction. In Section 3, the Genetic Neural Network with proper selection of neural network architecture for data mining is proposed for heart disease prediction. The experimental results of the heart attack disease system for prediction using Genetic-Neural Approach are explained in Section 4. Further, the paper presents the conclusion based on the results analysis of Genetic Neural Approach for Heart Disease Prediction and defines the future work to be carried for further research in healthcare service.

2. Literature Survey

Huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods hence calls for technological interventions so as to simplify management of those data. The decision making can be improved by using data mining in discovering patterns and trends in large amounts of complex data. Several ways are carried out in finding efficient technique of medical diagnosis for various diseases. Popular data mining tasks are association rules, classification, clustering, prediction and sequential patterns.

Classification techniques are capable of processing a large amount of data. Classification is one of the most widely used methods of Data Mining in Healthcare organization. The common classification techniques used in healthcare are Bayesian Networks, Support Vector Machines, Nearest neighbor method, Decision trees, Fuzzy logic, Fuzzy based Neural Networks, Artificial Neural Network, Genetic Algorithms [11].

In 2012, R. Bhuvaneswari et al., [12], use Naive Bayes classifier in medical applications. Two of the well-known algorithms are used in data mining classification are Backpropagation Neural Network(BNN) and Nave Bayesian (NB) calculate the priors, the probability of the object among all objects based on the previous experience. Bayesian technique is constructed on the probability concept. The posterior from the prior is calculated by bayes rules. Depending on the precise nature of the probability model, Naive Bayes classifiers is used to trained very efficiently in a supervised learning setting.

In 2011, Milan Kumari et al., [11], resolves cardiovascular disease dataset using different data mining algorithms, such as Support Vector Machine, Artificial neural networks (ANNs), Decision Tree, and RIPPER classifier. The authors analyze the performance of these algorithms through several statistical analysis factors such as accuracy and error rate. Accuracy of RIPPER, Decision Tree, ANN and SVM are 81.08%, 79.05%, 80.06% and 84.12% respectively. While the results of error rates for RIPPER, Decision Tree, ANN and SVM are 2.756, 0.2755, 0.2248 and 0.1588 respectively. Out of these four classification models SVM predicts cardiovascular disease with least error rate and highest accuracy.

In 2012, Mai Shouman et al., [13], combine different classifiers through voting to outperform other single classifiers. Decisions of multiple classifiers are associated by using aggregation technique called as Voting. The idea of applying multiple classifier voting is to divide the training data into smaller equal subsets of data and building a classifier for each subset of data. The results show that applying voting could not enhance the K-Nearest Neighbor accuracy in the diagnosis of heart disease.

In 2013, Senthil Kumar et al., [14], proposed a method that uses components of fuzzy logic like Fuzzification, Advanced Fuzzy Resolution Mechanism and defuzzification. Fuzzification is a process to transfer crisp values into fuzzy values. In the analysis of heart disease a fuzzy resolution mechanism uses predicted value with five layers, each layer has its own nodes. The results are tested with Cleveland heart disease dataset. Fuzzy Resolution Mechanism was developed using MATLAB. Defuzzification process converts the fuzzy set into crisp values.

In 2013, NABEEL AL-MILLI., [15], developed heart disease prediction system that uses the backpropagation algorithm technique to develop multilayer neural networks in a supervised manner. The error-correction learning rule is the basis for the back propagation algorithm. The algorithm uses a forward pass and a backward pass through the different layers of the network. The forward pass use to fix the synaptic weights of the networks. In the backward pass, the synaptic weights are all adjusted in accordance with an error-correction rule. Error signal is calculated as the difference between the desired output and the actual response of the network. Then the error signal is back propagated in the network. The actual response of the network moves nearer to the desired response by adjusting the synaptic weights in a statistical sense in the network. The generalized delta rule which minimizes the error is used for the weight adjustment in the network. Thus a medical decision support system can be developed particularly in the diagnosing of heart disease.

In 2013, Syed Umar Amin et al., [4], developed genetic neural network hybrid system. This system uses the global optimization advantage of genetic algorithm for initialization of neural network weights. A backpropagation algorithm is used to train the networks with optimize initialization of synaptic weights by Genetic Algorithm.

The determination of neural network structure is a challenging task. Even though significant progress has been carried out in classification related areas of neural networks, there are number of issues that are not been solved successfully or completely applying to the neural network. A small network could not be able to provide good performance owing to its limited information processing power. A large network, on the other hand, may have some connections redundant Neural network trains the input data pattern through different layers with local convergences which does not provides optimal solution to the problem [4]. Thus the selection of proper network architecture plays key role in training the input data.

There is being needed to have decision support system for predicting the existence or absence of heart disease. Wrong diagnosis or poor clinical decisions leads to mortality, All clinicians are not equally good in predicting the heart disease in which diagnosis plays a very important role. In the case of heart disease time is a precious, proper diagnosis at the right time saves life of many patients. The system can be considered assisting the doctor to come to decision making. Further, proposed solution to overcome these limitations in medical data mining for heart disease prediction is presented.

3. Proposed System

Diagnosing heart disease is considered as a non-linear problem that shows the complex causal relationship between the variables. However, there is a new computational paradigm called an artificial

neural network, which is suitable for problems of extreme complexity not addressable with our conventional technologies, either by the conventional computer programming or statistical method. Several studies have shown that an artificial neural network can be successfully applied in diagnosing heart diseases [3][4]. Therefore, the purpose of this study is to evaluate the application of neural network in predicting the presence of heart disease.

In the neural network, the hidden neuron can influence the error on the nodes to which their output is connected. It can greatly degrade the generalization capability of the neural network which leads to the significant deviation in prediction result to the problem. To overcome this, an approach is proposed which is able to find minimum number of hidden nodes. The Neural Network Training Problem consists in determining the synaptic weights of a neural network to get the desired output for a set of input vectors. As the Genetic Algorithm is able to find global optimize solution to the problem it can be used for the initialization of neural network weights. Thus, the proposed method with Genetic-Neural approach can be used to design system for the heart disease prediction.

This hybrid system uses backpropagation algorithm for learning and training the neural network. The Multi-Layer Neural Network is optimize by calculating the number of nodes in hidden layer to minimize the over fitting; which causes the overestimation of complexity of the target problem that leads to significant deviation in prediction. As the initialization of the Neural Network weights is a blind process which makes it difficult to find out globally optimized initial weights and the network output would run towards local optima hence the overall tendency of the network to find out a global solution is greatly affected. So the problem of local optimum solution is solved by optimizing the initial weights of neural network. For this a genetic algorithm is used which is specialized for global searching. Thus the system uses the backpropagation algorithm to train the network by using the weights optimized by Genetic Algorithm. Error is calculated using equation (1) to measure the differences between desired output and actual output that has been produced in feed forward phase. Error then propagated backward through the network from output layer to input layer as represented below. The weights are modified to reduce the error as the error is propagated.

$$Error=1/2(Output_{desired}-Output_{actual})^2 \quad (1)$$

This process will be repeated iteratively until convergence is achieved (targeted learning error or maximum number of iteration). The Genetic-Neural Approach for Heart Disease Prediction is used to test data to the optimum value and predict whether the patient have a heart disease or not. The Neural Network with Genetic Algorithm Approach for Heart Disease Prediction is shown in Figure 3.1.

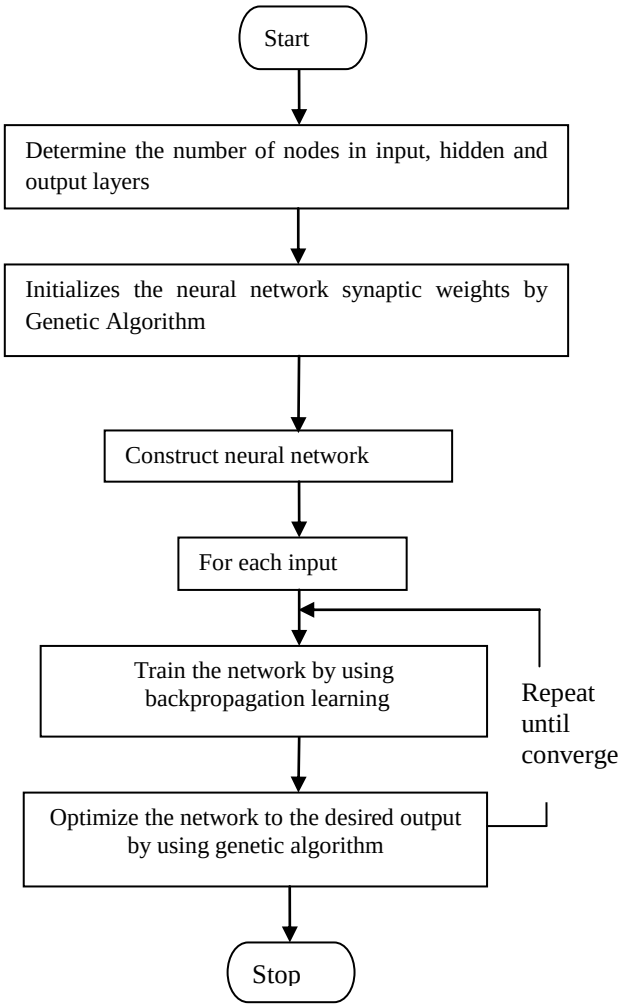

Figure 3.1: Flowchart of Genetic-Neural Approach for Heart Disease Prediction

The neural network uses the genetic algorithm fitness function to initializes the weights that makes it possible to have global optimal convergence. Neural network architecture is constructed by identifying the input and output layer neuron along with number of hidden layers and hidden nodes identification. As

Genetic Algorithm is adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics which can be used to initialize the neural network weights. Thus Genetic-Neural Network takes advantage of global optimization of genetic algorithm for initialization of neural network. Then the genetic algorithm fitness function is used to predict the heart disease.

4. Results and Discussion

The experimental results of the heart attack disease system for prediction using Genetic-Neural Approach are explained in this section. The system was developed using MATLAB R2012a. Global Optimization Toolbox and the Neural Network Toolbox were used for implementing the algorithm. The data for risk factors related to heart diseases are collected from 50 people who are provided by American Heart Association [4].

Table 4.1: Risk Factor Values and their Encoding

Name	Description
Sex	Male(1), Female(0)
Age	20-34(-2),35-50(-1),51-60(0),61-79(1),>79(2)
Blood Cholesterol	Below 200 mg/dL - Low (-1), 200-239 mg/dL - Normal (0), 240 mg/dL and above - High (1)
Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0), Above 139 mm Hg- High (-1)
Hereditary	Family Member diagnosed with HD - Yes (1) Otherwise No (0)
Smoking	Yes (1) or No (0)
Alcoholic Intake	Yes (1) or No (0)
Physical Activity	Low (-1) , Normal (0) or High (1)
Diabetes	Yes (1) or No (0)
Diet	Low (-1) , Normal (0) or High (1)
Obesity	Yes (1) or No (0)
Stress	Yes (1) or No (0)
Heart Disease	Yes (1) or No (0)

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, and obesity. The system indicates whether the patient had risk of heart disease or not. Most of the heart disease patients had many similarities in the risk factors. The 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 [4].

The approach uses Backpropagation Neural Network as the learning algorithm in Neural Network for this heart disease prediction system. However, Backpropagation learning depends on the several parameters in the Multi-Layer Feedforward Neural Network such as number of neurons in the hidden layers initialization of neural network weights. Due to this, Genetic Algorithm is used to obtain the optimal parameter value and weight for the Backpropagation learning so that the performance of Genetic Algorithm along with Multi-Layer Feedforward Neural Network is increased.

The Multi-Layer Feedforward Neural Network is constructed by calculating the number of nodes in input, hidden and output layers. The input nodes are taken as 12 equals to the number of risk factor associated in heart disease prediction system and output node is taken as 1 equals to the predicted output as 'Yes' or 'No'. the number of hidden nodes is computed from mean square error(mse) which is generated as the neural network is train by using Backpropagation Learning Algorithm considering 1 to 10 hidden nodes. The Genetic Algorithm (GA) is applied to initialize the neural network weight. The Genetic Algorithm is used to calculate the number of hub that is the number of layers in the neural network along with the total number of weights and bias used to initialize the network for each generation by genetic algorithm. Hence there are (I*Hn+10) + (10*O+2) number of total weights and biases. Fitness function is calculated for each chromosome based on mean square error. After selection, crossover and mutation in GA, the chromosomes with lower adaptation are replaced with better ones, and the better and fitter chromosomes (optimized solutions) that correspond to the interconnecting weights and thresholds of neural network are generated further, the Genetic Algorithm along with the Multi-Layer Feedforward Neural Network (MLFNN) is used to predict the heart disease. Thus, Genetic-Neural Approach for Heart Disease Prediction helps to increase the accuracy of medical resource utilization in order to reduce the health care.

4.1 Performance Metrics

The Genetic-Neural Approach for Heart Disease Prediction is evaluated to compute the accuracy.

Accuracy is determined as follows: Accuracy = (TP + TN) / (TP + FP + TN + FN) where TP denotes true positives. True positives refer to the positive tuples that were correctly labeled by the classifier. TN denotes true negatives. True Negatives are the negative tuples that were correctly labeled by the classifier. FP denotes false positives. False positives are the negative tuples that were incorrectly labeled by the classifier. FN denotes false negatives. False negatives are the positive tuples that were incorrectly labeled by the classifier. The resulted values of TP, TN, FP, and FN is computed by training the Heart Disease Dataset by using Genetic-Neural Approach for computation of Accuracy.

4.2 Experimental Results

The neural network is constructed by taking 12 nodes in input layer, 1 to 10 hidden nodes and node at which minimum mean square error occurred is taken as the number of hidden nodes and 1 node as output. Results shows that the minimum mean square error occurred at node 6 so number of hidden nodes are 6 as shown in Figure 4.1.

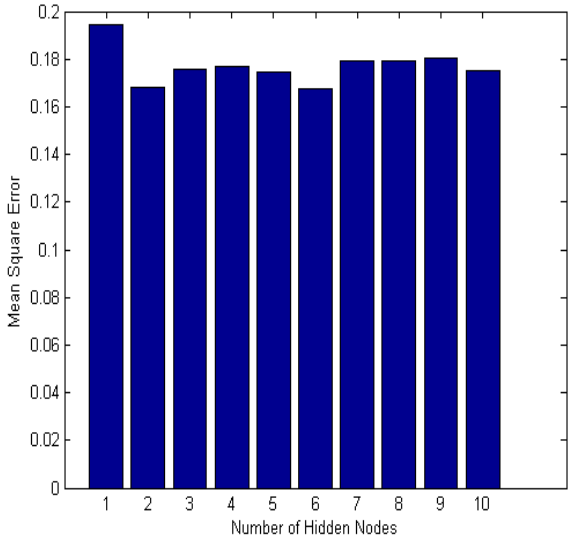

Figure 4.1: Resulted hidden number of nodes for neural network.

Thus, the results comparison of the Genetic Neural Approach using 6 and 10 hidden nodes in hidden layer along with the resulted values of TP, TN, FP and FN are shown in Table4.2.

Table 4.2: Comparison of the Genetic-Neural Approach for Heart Disease Prediction by Using 6 and 10 Hidden Nodes

Technique	Genetic Neural Approach For Heart Disease Prediction using 6 hidden nodes	Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors using 10 hidden nodes
True Positive	17	10
True Negative	32	33
False Positive	0	0
False Negative	1	7
Time Required	22.6322 seconds	25.3133 seconds
Accuracy	98%	84%

5. Conclusion and Future Scope

The Data mining techniques are nowadays mostly used in healthcare industry for predicting diseases. When these techniques applied in patient medical dataset has resulted in innovations, standards and decision support system that have significant success in improving the health of patients and the overall quality of medical services. In this study, a experiment is conducted with Heart Disease dataset by considering the Multi-Layer Neural Network along with Backpropagation Learning Algorithm used to train the network. Genetic Algorithm is used to optimize the initialization of neural network weights. This work demonstrates about Genetic Neural Network based prediction of heart disease by improving the accuracy as 98% using optimize neural network architecture and predicts whether the patient is suffering from heart disease or not.

As the healthcare domain is dynamic and this issue is a challenge to the data mining. It is also a forcing motivation to the data mining applications in healthcare. This dynamism gives way to new horizons and more data mining applications will be employed to discover new patterns and associations. In the view of the subjects examined in this study, future data mining studies seem to take place, not limited but in considerable weight, in distributed data mining applications and text mining algorithms. With the help of data mining algorithms, the classification performance increases. This can be further enhanced and expanded with more prediction algorithm for major life threatening diseases.

Acknowledgment

The authors feel a deep sense of gratitude to Prof. Dr. G. K. Patnaik HOD of Computer Science and Engineering Department for his motivation and support during this work. The authors are also thankful to the SSBT's COET, Jalgaon for being a constant source of inspiration.

References

[1] Tipawan Silwattananusarn and Kulthida Tuamsuk, "Data Mining and Its Applications For Knowledge Management:A Literature Review from 2007 to 2012", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol.2, no.5, Sept 2012.

[2] D. Tomar and S. Agarwal, "A Survey on Data Mining Approaches for Healthcare", International Journal of Bio-Science and Bio-Technology, vol.5, no.5, 2013.

[3] K. Priya, T. Manju and R. Chitra, "Predictive Model of Stroke Disease Using Hybrid Neuro-Genetic Approch", International Journal of Engineering and Computer Science, vol.2, no.3, pp 781-788, Mar 2013.

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

[5] K. Srinivas, G. R. Rao and A. Govardhan, "Survey on Prediction of Heart Morbitity Using Data Mining Techniques", International Journal of Data Mining and Knowledge Management Process,vol.1, no.3, May 2011.

[6] S. Rajasekaran and A. V. Pai, "Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis and Application", Eight Economy Edition, PHI, 2003.

[7] L. Hongiun, R. Setiono and H. Liu, "Effective Data Mining Using Neural Networks", IEEE Trans. On Knowledge and Data Mining, vol.8, no.6, Dec 1996.

[8] P. Gaur, "Neural Networks in Data Mining", International Journal of Electronics and Computer Science Engineering, vol. 1, no. 3, Aug 2013.

[9] P. Bajpai and D. M. Kumar, "Genetic Algorithm: An Approach to Solve Global Optimization Problem", vol. 1, no. 3, pp 199-206, 2010.

[10] A. Rangel-Merine, L. L. Bonilla, R. Linares and Y. Miranda, "Optimization Method based on Genetic Algorithms", International Journal of Apeiron, vol.12, no. 4, pp 393-409, Oct 2005.

[11] M. Kumari and S. Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction", International Journal of Computer Science and Technology, vol. 2, no. 2, pp. 304-308, Jun 2011.

[12] R. Bhuvaneswari and K. Kalaiselvi, "Naïve Bayesian Classification Approach in Healthcare Applications", International Journal of computer Science and Telecommunication", vol. 3, no. 1, pp. 106-112, Jan 2012.

[13] M. Shouman, T. Turner, and R. Stocker, "Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients", International Journal of Information and Education Technology, vol. 2, no. 3, pp. 220-224, Jun 2012.

[14] A. S. Kumar, "Diagnosis of Heart Disease Using Advanced Fuzzy Resolution Mechanism, International Journal of Science and Applied Information Technology, vol. 2, no. 2, pp. 22-30, Mar 2013.

[15] N. Al-Milli, "Backpropagation Neural Network for Prediction of Heart Disease", Journal of Theoretical and Applied Information Technology, vol. 56, no. 1, Oct 2013.

Nilima P. Patil received the Master of Engineering (Computer Science & Engineering) degree from the North Maharashtra University Jalgaon Maharashtra) in 2011& Bachelor of Engineering from Sant Gadge Baba University Amravati. In 2007, she joined the Department of Computer Engineering at SSBTs College of Engineering &Technology Bambhori, Jalgaon, Maharashtra, India as Assistant Professor. During the period of 2005-2006, she worked as lecturer at Jawaharlal Darda institute of technology Yavatmal(MS).
Her area of interest is image pattern & pattern recognition, data mining. During 8 years of teaching carrier, she published 4 papers in international journals, 5 papers in international &national conference. She is life member of ISTE.

Nilakshi P. Waghulde, pursuing Master of Engineering degree in Computer Science and Engineering from SSBT's COET, North Maharashtra University, India, in 2014. From 2012 to 2014, she was a Lecturer with Government College of Engineering, Jalgaon, Maharashtra, India. She has published two papers in scientific journals and conferences. Her research interests include weightless neural networks, hybrid neural systems, and applications of neural networks.