ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

CDSS for Heart Disease Prediction Using Risk Factors

Rupali Wagh[1], S. S. Paygude[2]

PG Scholar, Dept. of Computer Engineering, MIT, Pune University, India[1]

Associate Professor, Department of Computer Engineering, MIT, Pune University, India[2]

ABSTRACT: Clinical Decision Support System (CDSS) is a tool which helps doctors to make better and uniform decisions. This work proposes new approach for prediction of heart diseases based on risk factors such as age, ECG, heart rate diabetes, slope, hypertension, high cholesterol, or physical inactivity, etc. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. Application of artificial intelligence techniques will be very much helpful not only to the medical professionals but also to an individual in the prediction of heart diseases because of its accuracy and reduction in cost. This work uses Neuro-Fuzzy model (NFS) to the available data for prediction of heart disease. The hybrid system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights. The neuro-fuzzy model combines the neural network adaptive capabilities and the fuzzy logic reasoning approach for prediction of heart disease. The proposed system is going to produce an intelligent system with good accuracy.

KEYWORDS: Heart disease risk factors, neuro-fuzzy approach, genetic algorithm

I. INTRODUCTION

Clinical decision support system (CDSS) helps in making diagnosis of various medical diseases. Modern day medical diagnosis is a very composite process, entailing precise patient data, a philosophical understanding of the medical literature and many years of clinical experience. Decision support systems that are implemented with the support of data mining and artificial intelligence have the capability to espouse in a new environment and to learn with instance.

Heart disease diagnosis is a complex task which requires much experience and knowledge. Traditional way of predicting Heart disease is doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Almost all the doctors are predicting heart disease by learning and experience. The diagnosis of disease is a difficult and tedious task in medical field. Predicting Heart disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. Nowadays, Healthcare industry today generates large amounts of complex data about patients, Hospitals resources, disease diagnosis, electronic patient records, medical devices etc , which contains hidden information. This hidden information is useful for making effective decisions.

There are many studies and researches on the prevention of heart disease risk. Data from studies of population has helped in prediction of heart diseases based on various risk factors such as blood pressure, smoking habit, cholesterol and blood pressure levels, diabetes etc. Researchers have used these risk factors in adapted form of simplified score sheets that allow patients to calculate the risk of heart diseases.

Quite a number of methods have been proposed for the Diagnosis of cardiovascular disease. Among these methods, Neuro-Fuzzy based approach seems promising because of its high level of diagnosis accuracy. In short, the problem with neural networks (NN) is that a number of parameter such as optimal number of hidden nodes, selection of the relevant input variables and selection of optimal set of connection weights has to be set before training can begin [01]. However, there are no clear rules on how to set these parameters. Yet these parameters determine the success of the training of the neural network [02]. Despite the promising results produced by neuro-fuzzy based system when used for

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

medical diagnoses, a lot of time is required to determine the connection weights needed to effectively train the neural network module. This often leads to high computational cost and constitutes a major setback to the real life implementation of neuro-fuzzy based systems. This research proposes an enhanced diagnostic decision support system which adopts genetic algorithm technique to optimize the performance of a neuro-fuzzy diagnostic method. The proposed system consists of a Fuzzy logic (FL) component which takes as input the key attributes of heart disease diagnosis, the NN component was designed to compute the membership function parameters needed by the FL module for prediction, while the Genetic Algorithm component was designed to automatically generate a set of optimal connection weights needed to efficiently train the NN. The proposed system is built with the aim of providing a decision support platform with the capability to aid medical practitioners in administering efficient diagnosis outcome for cardiovascular disease patients.

This system will not only prove helpful to medical professionals but it will also give patients a warning about the probable presence of cardiovascular disease even before he\she visits a hospital or goes for costly medical check-ups. The system implementation will be done using MATLAB.

II. Related work

For devising clinical decision support systems, literature presents a number of researches that have made use of artificial intelligence and data mining techniques. Among them, to support decision makers in the risk prediction of heart disease, a handful of researches have been presented. A few of the significant researches obtainable in the literature are explained below.

One of the systems [04] uses neural based learning classifier for classifying data mining tasks showed that neural based learning classifier system performs equivalently to supervise learning classifier. IEHPS [05] intelligent and effective heart attack prediction system was built using data mining and neural networks and it proposed extracting significant patterns for heart disease prediction using K-means clustering and used MAFIA algorithm to mine the frequent patterns. Polatet al. [06], developed system using hybrid fuzzy and k-nearest neighbour approach for the prediction of heart disease, which had 87% accuracy in diagnosis

Latha Parthiban, et. al. [07] performed a work, *"Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm*". In this paper projected a move towards on the basis of coactive neuro-fuzzy inference system (CANFIS) for the prediction of heart disease. The CANFIS model diagnosed the occurrence of disease by integrating the neural network adaptive capabilities and the fuzzy logic qualitative approach and further integrating with genetic algorithm. On the basis of the training performances and classification accuracies, the performances of the CANFIS model were evaluated. The CANFIS model is shown the potential in the prediction of the heart disease as illustrated by the result.

The heart disease accounts to be the leading cause of death worldwide. 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.Earlier medical domain applications predict heart disease using computer aided diagnosis methods, where the data are obtained from some other sources and are evaluated based on computer-based applications. A clinician uses several sources of data and tests to make a diagnostic impression but all tests are not necessary and useful for the diagnosis of a heart disease. Also this process is time consuming, costly, requires lots of tests and really depends on medical expert's opinions which may be subjective. 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. To handle this problem, machine learning techniques have been developed to gain knowledge automatically from examples or raw data which saves time, money and accurate prediction of heart disease earlier than usual.

III. Genetic algorithm and their integration with NFS

The Genetic Algorithm (GA) is a global, parallel, stochastic search method, founded on Darwinian evolutionary principles. Genetic algorithms are suitable for optimizing discrete variables and non-continuous cost functions. They use the principles of gene crossover, reproduction and natural selection in evolutionary biology. When applying genetic algorithms to a problem, the first step involves representing a chromosome to describe the individuals in the population and encoded them as commonly binary or floating point. The GA then evaluates the 'fitness' of the offspring's within

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

the existing generation. While, the fitness function is dependent the problem being solved. The GA applies the mutation, crossover and selection operations to the individuals in the population to explore all promising regions in the solution space till it achieves the needed solution.

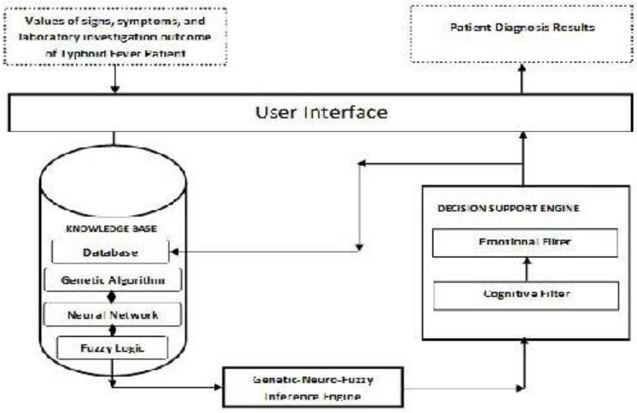

Fig 1: Architecture of the proposed genetic adaptive neuro-fuzzy inference

This advantage of the GA can be applied to neural networks to optimize the topology and/or weight parameters. For a GA-based algorithm employed in fuzzy neural networks, the initial network structure is usually generated or given firstly. The GA is then used to optimize the network's topology in terms of membership functions and/or fuzzy rules. A GA-based algorithm thus has the potential to evolve the number of fuzzy rules, and may be developed as a GA-based pruning method for fuzzy neural networks. The GA is used to determine all the parameters of the membership functions of the fuzzy controller.

IV. 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 neuro –fuzzy system (NFS), 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 neuro –fuzzy system (NFS) can be successfully applied in diagnosing heart diseases [7].

The neuro –fuzzy system (NFS) model which combines the adaptability of fuzzy inputs with neural network is used to accurate predictions. And this kind of network is more efficient than the simple neural networks. Neuro-fuzzy systems are fuzzy systems which use ANN's theory in order to determine their properties like fuzzy sets and fuzzy rules by processing data samples. The main objective of this research is to develop an Intelligent Heart Disease Prediction system.

The process of Neuro-Fuzzy with GA presented below:

1. Initialize the process of predicting Cardiovascular Disease.
2. Extract the patient's details from dataset.
3. Assign the input to NFS.
4. Selection process starts by assigning weights to each attributes in random manner.
5. Training the network using Back-propagation algo.
6. Compute output values.
7. Compute fitness using below equation

$$\text{Mean Square Error (MSE)} = \frac{\sum(\text{Output} - \text{Targets})^2}{\text{Number of Samples}}$$

8. If MSE is less than error then go to step 10, otherwise, go to step 9.
9. Select the parents and apply crossover and mutation.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

10. Train NFS with selected connection weights.
11. Study the performance of test data.

As discussed in above steps, the process initializes the prediction of Cardiovascular Diseases. It extracts the details of various medical parameters of the patient. The system then gives training to the Neuro-Fuzzy system. Training process generally consists of four sub processes, which are:1. giving the inputs and outputs of the patient's parameters to the system.2. the selection process of sample 3.training using the back- propagation algorithm and finally 4. The output values are calculated and Mean Square Error (MSE) is computed. If the Mean Square Error (MSE) is less than error then training to the Neuro-Fuzzy System is completed and if the Mean Square Error (MSE) is higher than the error then the two chromosomes or samples are selected for further process. In this further process genetic algorithm is applied, in which crossover and mutation is applied to the samples. New weights are taken for the samples and again training process is carried out until the MSE is less than error.

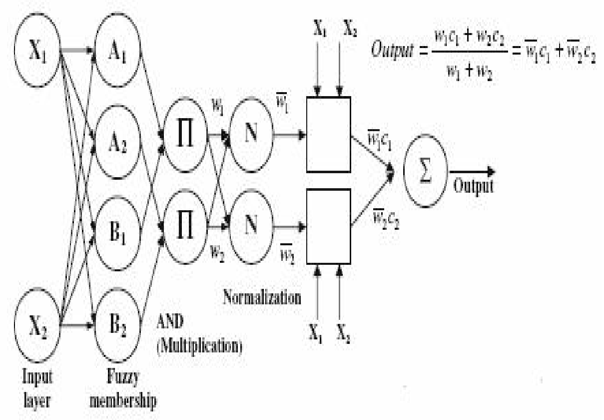

Fig. 2 A prototype NFS network and output calculation

In order to improve the learning of the *NFS*, quicker training and enhance its performance, we use genetic algorithms to search for the best number of MF for each input, and optimization of control parameters such as learning rate, and momentum coefficient. This approach is also useful to select the most relevant features of the training data which can produce a smaller and less complicated network, with the ability to generalize on freshly presented data, due to the removal of redundant variables.

The GA combines selection, crossover, and mutation operators with the goal of finding the best solution to a problem by searching until the specified criterion is met. The solution to a problem is called a chromosome, which is composed of a collection of genes. In hybrid neuro-fuzzy-genetic applications, genes are the *NFS* parameters to be optimized. The GA creates an initial population and then evaluates this population by training a network for each chromosome. It then evolves the population through multiple generations in the search for the best network parameters.

A. Heart Disease Dataset

Recent research in the field of medicine has been able to identify risk factors that may contribute toward the development of heart disease but more research is needed to use this knowledge in reducing the occurrence of heart diseases. Diabetes, hypertension, and high blood cholesterol have been established as the major risk factors of heart diseases. Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity are also associated with the major heart disease risk factors and heart disease.

Here the dataset related to the cardiovascular disease is provided to the neuro-fuzzy system. The dataset consist of the patients symptoms of cardiovascular diseases. It consists of cardiovascular disease patients' information. The system uses Cleveland databases which is publicly available.This dataset represents the information of the patients like Age in years, Sex, Blood Pressure, Blood Cholesterol, Diabetes, Electrocardiographic results, Heart rate, Physical activity, Slop of the peak, Number of major vessels colored by fluoroscopy, Thalassemia(Defect type) which is

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

considered as a most commonly risk factors of the cardiovascular disease. There are samples or instances for 12 attributes like sex, age, blood cholesterol, blood pressure, chest pain, electrocardiographic results, heart rate, physical activity, diabetes, diet number of vessels, Thalassemia to predict whether one can have a cardiovascular disease or not. The system was developed using MATLAB.

	Risk factors	**Description with encoded values**
1	Age	20-34 (-2), 35-50 (-1), 51-60 (0),61-79 (1) , >79 (2)
2	Blood pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
3	Blood cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Diabetes	Yes (1) or No (0)
5	Physical Activity	Yes (1) or No (0)
6	Slope	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3:down sloping
7	Chest Pain	Yes (1) or No (0)
8	ECG	The slope of the peak exercise ST segment Normal(0) Abnormal(1) Hyper(2)
9	Heart Rate	Below 100- Low (-1) 100 to 150- Normal (0) otherwise- High (1)
10	No.of major vessals	Number of major vessels Colored by fluoroscopy (0-3)
11	Thal	Normal-3(0) fixed defect-6(1) reversible defect-7 (2)
12	Sex	Male(1) or Female (0)
o/p	Heart Disease	Yes (1) or No (0)

Table 1: Description Of Cleaveland Heart Disease Database

The Table shows some of the important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

V. SIMULATION RESULTS

The medical records of 100 Heart Disease patients were collected from the UCI Machine learning repository.This acquired data was analyzed and pre-processed to the required format. Matrix Laboratory (MATLAB) was used to implement the proposed system.

In this heart disease prediction system one can know whether the person having the disease or not with greater accuracy than other prediction systems. Accuracy is measured on the basis of Mean Square Error means that if the mean square error is low then accuracy is high and if mean square is high then accuracy is low.
Accuracy is determined using the following formula:

$$Accuracy= (TP+TN)/ (TP+FP+TN+FN)$$

Where, TP-True positive, TN- True Negative, FP-False positive and FN-False negative

As shown in Table 2, for cardiovascular disease prediction 100 records is used for testing the performance of the system on the basis of TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) which calculate the accuracy of the system. This represents that when neuro-fuzzy system with genetic algorithm is applied then the result or the prediction of the cardiovascular disease is more accurate than when it is applied without genetic algorithm.

Technique	Neuro-Fuzzy System(without genetic algorithm)	Neuro-Fuzzy System(With genetic algorithm)
True Positive	38	40
True Negative	44	50
False Positive	05	0
False Negative	12	10
Accuracy	82%	90%

Table 2: Comparison of the NFS with GA and Without GA Approach for CVD Prediction by Using 100 Records.

B. Simulation Results

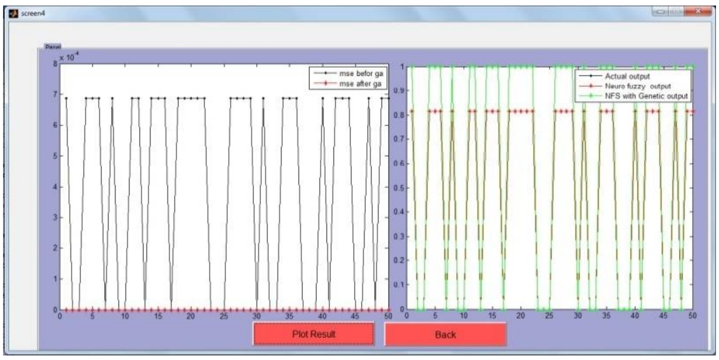

Figure 3: Comparison graph

The above figure 3 shows various comparisons. In the first part the value of Mean Square Error (MSE) before application of Genetic algorithm (GA) are compared with the values of MSE after the application of genetic algorithm. From this graph, it can be easily deduce that the value of MSE after the application of GA (red) is far less than that of before application of GA (black). The substantial reduction in MSE will lead to increase the accuracy of the system and helps in realistic prediction of the CVD.

ISSN(Online): 2320 - 9801
ISSN (Print) : 2320 - 9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

In the second part, simply using neuro-fuzzy system (without application of GA) the assigned values of outputs (which isinput to the NFS) i.e. '0', '1' are not recognized perfectly (red). On the other hand the assigned values to the outputs (which is input to the NFS) like '0', '1' etc. are clearly recognized using neuro-fuzzy system in connection with genetic algorithm (green). This comparison graph proves that neuro-fuzzy system with genetic algorithm can produce more realistic and accurate results.

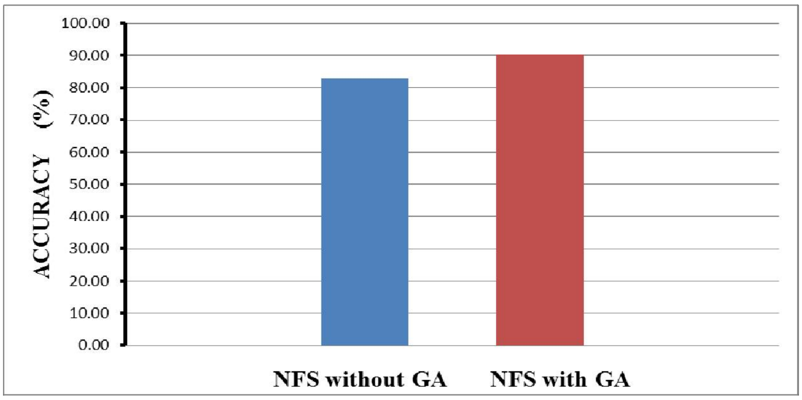

Figure 4: Accuracy of the NFS with and without GA

It can be seen from Figure 4 that the proposed technique (NFS with GA) had an overall average diagnosis accuracy of 90% as against that of the (NFS without GA) method which was 82% and this seem promising and could help increase the overall accuracy in the CDSS of heart disease and other diseases in the world.

VI. CONCLUSION AND FUTURE WORK

A Multi-technic decision support system powered by genetic algorithm, neural network, and fuzzy logic concepts for the diagnosis of heart disease has been investigated in this study. An improved genetic algorithm concept was used to automatically supply the optimal set of weights needed to effectively train the neural network module. Usually, the membership function parameters of FIS are manually set thereby making it difficult for the FIS to provide accurate diagnosis results when confronted with new cases. To address this problem, the trained, validated, and tested neural network module was configured to automate the provision of membership function parameters for the fuzzy inference system, that is, building some form of learning and tuning capability into the fuzzy inference system. With this development, the fuzzy inference system was able to provide timely and reliable diagnosis outcome for new cases. The outcome of the evaluation process conducted in this research shows that the proposed system (NFS with GA) had it attained a diagnosis accuracy of 90% as compared to 82% of the (NFS without GA) method. Also, in terms of time taken to diagnose a patient, the proposed system also performed better than the conventional (NFS without GA).Therefore, the proposed technique (NFS with GA) has the capability to alleviate the key problems associated with Neuro-Fuzzy Based diagnostic methods if fully embraced and as well it could be adopted to solve challenging problems in several other domains.

REFERENCES

1. Koehn P. Combining Genetic Algorithms and Neural Networks: The Encoding Problem, A thesis presented for the Master of Science Degree, the University of Tennessee, Knoxville,1994
2. Zadeh LA The calculus of fuzzy if/then rules. AI Expert 7: 27-27,1992
3. Ebene S. M. Metev and V. Ebenezer O. O, Oyebade K. O.," Heart Diseases Diagnosis Using Neural Networks Arbitration", I.J. Intelligent Systems and Applications, vol. 12,pp. 75-82,2015
4. Hai H.Dam, Hussain A.Abbass and Xin Yao, "Neural – Based Learning Classifier Systems", IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.1, pp.26-39, 2008
5. Shantakumar B.Patil and Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research, Vol.31, No.4, pp.642-656, 2009
6. Polat , K., S. Sahan, and S. Gunes, Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour based weighting preprocessing. Expert Systems with Applications 32 p. 625–631,2007

7. Latha Parthiban and R. Subramanian, “Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm”, International Journal of Biological and Life Science, Vol. 15, pp. 157 - 160, 2007
8. G.Camps-Valls, L.Gomez-Chova, J.Calpe-Maravilla, J.D.Martin-Guerrero, E.Soria-Olivas, L.Alonso-Chorda, J.Moreno, “Robust support vector method for hyperspectral data classification and knowledge discovery.” vol.42, no.7, pp.1530–1542, July.2004
9. N. Al-Milli, “Backpropagation Neural Network for Prediction of Heart Disease”, Journal of Theoretical and Applied Information Technology, vol. 56, 2013.
10. Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Technique, 978-1-4244-1968-5/08/$25.00 ©2008 IEEE.
11. P.K. Anooj,Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules”, Journal of King Saud University – Computer and Information Sciences 2012
12. "Coronary heart disease statistics fact sheets 2008/2009," British Heart Foundation, London 2008
13. *L. Zaret, M. Moser, and L. S. Cohen, Yale university school of medicine heart book. New York: Hearst Books, 1992.*
14. B. Phibbs, The human heart: a basic guide to heart disease. Philadelphia: Lippincott Williams & Wilkins, 2007.
15. A. Selzer, Understanding heart disease. Berkeley: University of California Press, 1992
16. O. S. Randall and D. S. Romaine, The encyclopaedia of the heart and heart disease. New York, NY: Facts on File, 2005.
17. Wood D, De Backer, Prevention of coronary heart disease in clinical practice: recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. Atherosclerosis 140: pp.199–270, 1998
18. 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
19. 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

BIOGRAPHY

Rupali Wagh is a PG student in the Computer Engineering dept, MIT. She received her bachelor of Information Technology (B.E.I.T.) degree in 2008 from, Pune University, India. Her research interests are Computer Networks (wireless Networks), Artificial Intelligence, Data mining, Data Science etc.