Effective Heart Disease Detection Using Artificial Intelligence and Fuzzy Logic

Sandeep Dubey[1], Mayank Deshmukh[2], Pratik Singh[3], Mayuresh Ghanekar[4], Prof. Ankit Anand[5]
[1, 2, 3, 4, 5]Dept. of Information Technology
[1, 2, 3, 4, 5]M.G.M.College Of Engineering and Technology, Kamothe, NaviMumbai

Abstract- *In a huge country like India there is an exceptional growth in number of heart related diseases. Heart disease is the leading cause of death for both men and women. An estimated 17.7 million people died from CVDs in 2015, representing 31% of all global deaths. In a 2005 survey, most respondents—92%—recognized chest pain as a symptom of a heart attack .According to WHO report its been found that out of total patients only 27% patients were aware about heart disease. The objective of this research is to create a "Effective and Intelligent Heart disease detection technique ". The main aim of this research is to develop an Intelligent user friendly and cost effective system to detect the heart disease at an early stage. Almost all systems that predict heart diseases use clinical dataset having parameters and inputs from complex tests conducted in labs .This system focuses risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc which are serious causes for heart related diseases. Effective feature selection is possible using advanced NEURO FUZZY Techniques and Precise prediction of disease is possible using Genetic Neural Networks and Data MiningTechniques. A novel way to overcome the limitation of existing system using artificial Intelligence and Machine Learning is proposed.*

Keywords- Effective, intelligent, cost effective, user friendly, neuro fuzzy, genetic neural network, data mining

I. INTRODUCTION

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. In today's world data mining plays a vital role for prediction of diseases in medical industry. In medical diagnosis, the information provided by the patients may include redundant and interrelated symptoms and signs especially when the patients suffer from more than one type of disease of same category. The physicians may not able to diagnose it correctly. So it is necessary to identify the important diagnostic features of a disease and this may facilitate the physicians to diagnosis the disease early and correctly.

Fuzzy Systems is used for solving a wide range of problems in different application domains. The use of Genetic Algorithms for designing Fuzzy Systems allows us to introduce the learning and adaptation capabilities.The paper briefly reviews the classical models and the most recent trends for Genetic Fuzzy Systems. Accurate and reliable decision making in oncological prognosis can help in the planning of suitable surgery and therapy, and generally, improve patient management through the different stages of the disease. To indicate that the reliable prognostic marker model than the statistical and artificial neural-network-based methods. In today's stressful & hectic lifestyle it is difficult to take care of heart health. Heart disease (HD) is a major cause of morbidity and mortality in the modern society. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource.

A system for automated medical diagnosis would enhance medical care and reduce costs. System not only helps medical professionals but it would give patients a warning about the probable presence of heart disease even.

II. LITERATURE REVIEW

Awang, R. &Palaniappan, S., "Intelligent Heart Disease Prediction System Using Data Mining Techniques" This research has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals.

Patil, S. &Kumaraswamy, Y., "Intelligent and effective heart attack prediction system using data mining and artificial neural network" IEHPS presents methodology for the extraction of significant patterns from the heart disease warehouses for heart attack prediction. Initially, the data warehouse is pre-processed to make the mining process more efficient. The pre-processed data warehouse is then clustered using the K-means clustering algorithm. Frequent Item set Mining (FIM) is performed using MAFIA (Maximal Frequent Item set Algorithm) for the extraction of association rules from the clustered dataset. Weightage is then calculated and the patterns vital to heart attack prediction are selected according to the weightage

P.K. Anooj, "Clinical Decision Support System: Risk Level Prediction of Heart Disease Using Decision Tree Fuzzy Rules "It is a weighted fuzzy rule-based clinical decision support system (CDSS) for the diagnosis of heart disease. Fuzzy rules are automatically generated to and are weighted in accordance with their importance using the attribute weightage. These weighted fuzzy rules are applied on the rule base of the fuzzy system before carrying out prediction on the designed fuzzy-based CDSS.

III. EXISTING WORK

Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. These techniques have been very effective in designing clinical support systems because of their ability to discover hidden patterns and relationships in medical data. One of the most important applications of such systems is in diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Almost all systems that predict heart diseases use clinical dataset having parameters and inputs from complex tests conducted in labs. None of the system predicts heart diseases based on risk factors such as age, family history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. Heart disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis. System based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical checkups. Hence this paper presents a technique for prediction of heart disease using major risk factors. This technique involves two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implemented uses the global optimization advantage of genetic algorithm. The learning is fast, more stable and accurate as compared to back propagation. The system was implemented in Matlab.

IV. PROPOSED WORK

The objective of this research is to create an intelligent & cost effective system which will overcome the limitations of existing system and improve its performance. A novel way to enhance the performance of a model that combines genetic algorithms and neuro fuzzy logic for feature selection and classification is proposed.

Various soft computing methods have been used for the detection of a potential medical problem. Thus, a reliable method for both feature selection and classification is required. The feature selection is based on a new genetic algorithm and classification is based on neuro fuzzy system (NFS). Feature selection is another factor that impacts classification accuracy. By extracting as much information as possible from a given data set while using the smallest number of features, we can save significant computation time and build models that generalize better for unseen data points.

Neuro fuzzy systems are fuzzy systems which use ANN's theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples. The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction.

System with NFS and genetic algorithm using historical heart disease databases to make intelligent clinical decisions which traditional decision support systems cannot. The NFS model integrates adaptable fuzzy inputs with a modular neural network to rapidly and accurately approximate complex functions. Fuzzy inference systems are also valuable, as they combine the explanatory nature of rules (MFs) with the power of neural networks.

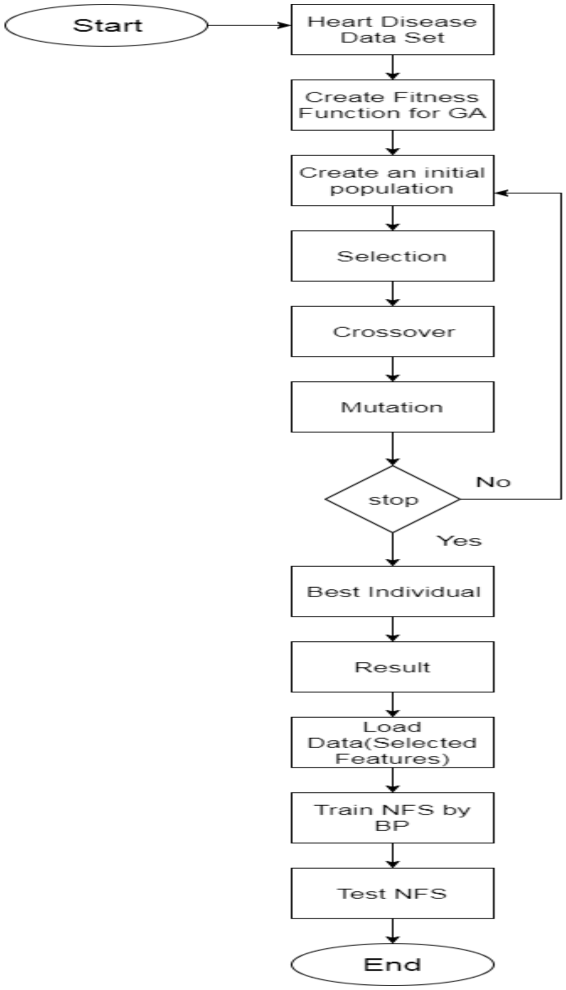

Figure1: Flow chart of proposed work

Data Set

The dataset in this research is heart disease data. It consists of heart disease patient's information. The publicly available Cleveland heart-disease database.

Software

For feature selection the coding is done using genetic algorithm commands in command window and for classification NFS tool of MATLAB is used.

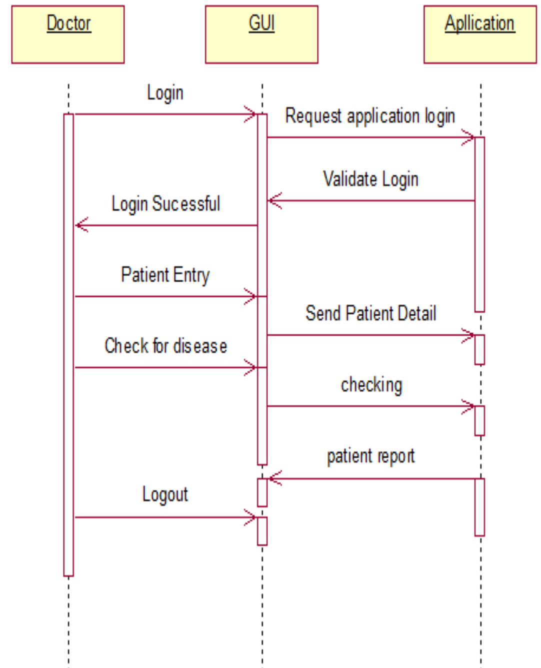

Figure 2: Sequential Diagram

Feature Selection using Genetic Algorithm:

Feature selection reduces the dimensionality of data by selecting only a subset of measured features (predictor variables) to create a model. Reducing the number of features (dimensionality) is important in statistical learning. For many data sets with a large number of features and a limited number of observations, such as bioinformatics data, usually many features are not useful for producing a desired learning result and the limited observations may lead the learning algorithm to over fit to the noise. Reducing features can also save storage and computation time and increase comprehensibility.

GAs is basically a domain independent search technique, they are ideal for applications where domain knowledge and theory is difficult or impossible to provide. 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. 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. GAs cause the initial population to evolve towards a population that is expected to contain the best solution. A fitness function is needed in the Genetic Algorithm to evaluate whether an individual is "fit" to survive.

The principle is that individuals with high classification accuracy and small number of features has a high fitness value, and thus high probability to be pass its genes to the next generation. A single objective fitness function that combines the two goals into one was designed to solve the multiple criteria problem. The formula is as below

$$fitness = w_f \Big/ \sum_{i=1}^{n_f} f_i$$

wf for the number of features; fi is the mask value of the ith feature, '1' represents that feature i is selected; '0' represents that feature i is not selected.

Selected chromosomes are subjected to mutation and to crossover. The crossover operator exchanges genetic information between two strings. The strings of two randomly selected solutions are broken up at randomly chosen position, and parts of the strings are exchanged. One hopes that two solutions with good properties create an even better one. New genetic material is introduced by the mutation operator. The values of individual genes are changed and hence, new solutions are chosen. Mutation becomes important when after some generations the number of different strings decreases because strong individuals start dominating. In a situation of strong dominance of few strings, the crossover operator alone would not bring any changes and the search for an optimal solution would be ended.

Following are some steps in genetic algorithm

1. Initialize population by selecting random individuals from the space S.
2. for the specified number of generations do
3. for the size of the population do
4. Select two individuals (with uniform probability) asparent1 and parent2.
5. Apply crossover to produce a new individual (child).
6. Apply mutation to child.
7. Calculate the distance between child and parent1 as d1, and the distance between child and parent2 as d2.
8. Calculate the fitness of child, parent1, and parent2 as f, f1, and f2 respectively.
9. if (d1 < d2) and (f> f1) then
10. replace parent1 with child
11. else
12. if (d2 <= d1) and (f> f2) then
13. Replace parent2 with child.
14. end if
15. end for
16. Extract the best (highly-fitted) individuals as your final solution.

Neuro-Fuzzy system:

Fuzzy logic and neural networks are natural complementary tools in building intelligent systems. While neural networks are low-level computational structures that perform well when dealing with raw data, fuzzy logic deals with reasoning on a higher level, using linguistic information acquired from domain experts. However, fuzzy systems lack the ability to learn and cannot adjust themselves to a new environment. On the other hand, although neural networks can learn, they are opaque to the user. Integrated neuro-fuzzy systems can combine the parallel computation and learning abilities of neural networks with the human-like knowledge representation and explanation abilities of fuzzy systems. As a result, neural networks become more transparent, while fuzzy systems become capable of learning. A neuro-fuzzy system is a neural network which is functionally equivalent to a fuzzy inference model. It can be trained to develop IF-THEN fuzzy rules and determine membership functions for input and output variables of the system. Expert knowledge can be incorporated into the structure of the neuro-fuzzy system. At the same time, the connectionist structure avoids fuzzy inference, which entails a substantial computational burden.The structure of a neuro-fuzzy system is similar to a multi-layer neural network. In general, a neuro-fuzzy system has input and output layers, and three hidden layers that represent membership functions and fuzzy rules.A neuro-fuzzy system has total five layer.

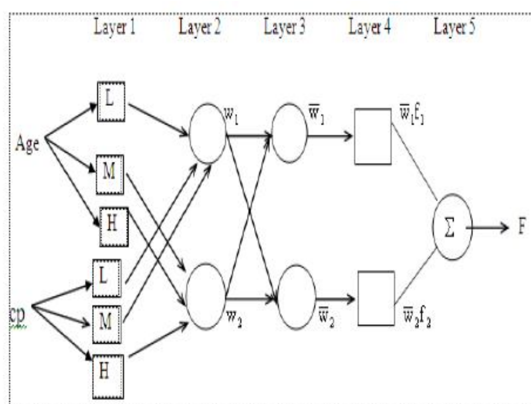

Figure 3: Architecture of the neuro-fuzzy system

In Layer 1

The node function is the membership of fuzzy set with its related input. The fuzzy if then rules has input variables, membership function and output variable. The parameter are determined by Gaussian membership function

$$O_i^1 = \mu_{cp}(x) = e^{-\frac{1}{2}\left(\frac{x-c_i^1}{\sigma_i^1}\right)^2}$$

where c and σ represent the membership function centre and width respectively in order to determine coordinates of Gaussian membership function.

Layer 2

Nodes are fixed to calculate the firing strength of rule. The output is derived by the product of all incoming values. Inputs from the nodes in the Layer 1 are multiplied with Layer 2 and the firing strength of the rules are generated. The output of the Layer 2 is given by

Wi=μ Age(x) x μcp(y)…………………

where wi is the firing strength of rule i.

Layer 3

In this layer nodes calculates the weight, they are normalized. The following steps are used to frame the fuzzy rules for the proposed system.

I: Get the highest and lowest values of the attributes.

II: Find out the degree of entropy of each attribute.

III: Select the attributes whose degree of entropy is more than the threshold value.

IV: Choose the attribute X found in step III that has the highest degree of entropy, and find the highest attribute values and the lowest attribute values of the attribute X of each species based on the results obtained in step I. Arrange these attribute values in an ascending sequence.

V: Find out the statistical distribution of the attribute values of X of the training instances falling in each interval. The rules are generated from the training instances. If most of the training instances fall in an interval belonging to the species Z1, and the corresponding linguistic term of this interval is Y1, then generate the following fuzzy rule:

IF X is Y 1 THEN the flower is Z1.

The ith node calculates the portion of the ith rules firing strength to the sum of all rules firing strengths

$$\overline{w_i} = \frac{w_i}{\sum_{i=1}^{m} w_i}$$

where the output are called normalized firing strengths is of this layer.

Layer 4

The output of this layer is a linear combination of input multiplied by the normalized firing strength. The consequent of the rules

$$\overline{W_i} f_i = \overline{Wi}(\mu Age(x)+\mu_{cp}(y)+\ldots\ldots\ldots\ldots)$$

Where $\overline{wi}$ is a normalized firing strength from layer 3 and re performed by the nodes in this layer.{Agei,cpi,….} are the parameter set of this node.

Layer 5

This layer is the simple summation of overall output... is the defuzzification layer. Each neuron in this layer represents a single output of the neuro-fuzzy system. It takes the output fuzzy sets clipped by the respective integrated firing strengths and combines them into a single fuzzy set.

$$\sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

The Back-Propagation Training Algorithm*:*

A neuro-fuzzy system is essentially a multi-layer neural network, and thus it can apply standard learning algorithms developed for neural networks, including the back-propagation algorithm.When a training input-output example is presented to the system, the back-propagation algorithm computes the system output and compares it with the desired output of the training example. The error is propagated backwards through the network from the output layer to the input layer. To determine the necessary modifications, the back-propagation For each example e in the training set do

1. O = neural-net-output (network,e) ; forward pass
2. T = teacher output
3. Calculate error (T - O) at the output units
4. Compute delta_wi for all weights

from hidden layer to output layer ;
backward pass

5. Compute delta_wi for all weights
from input layer to hidden layer ;
backward pass continued
6. Update the weights in the network end.
7. until all examples classified correctly or
stopping criterion satisfied
8. return(network)

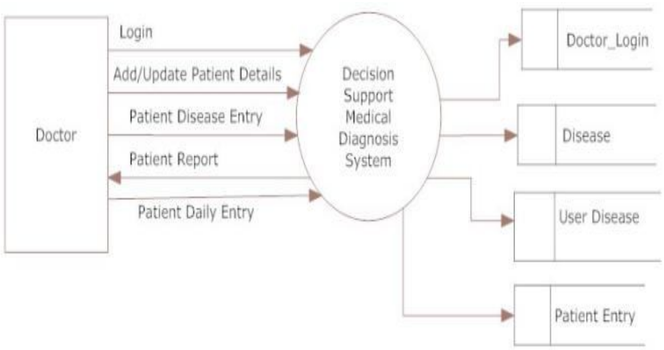

Figure 4: DFD Context Level0

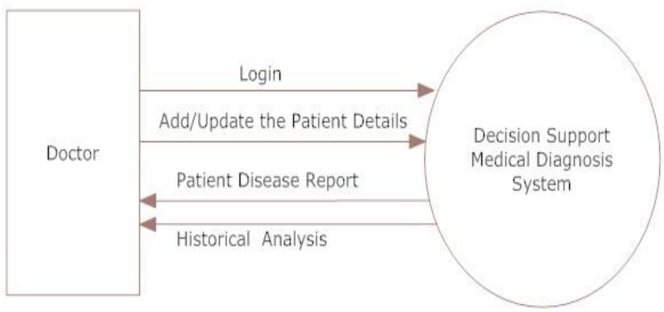

Fig 5: DFD Level 1

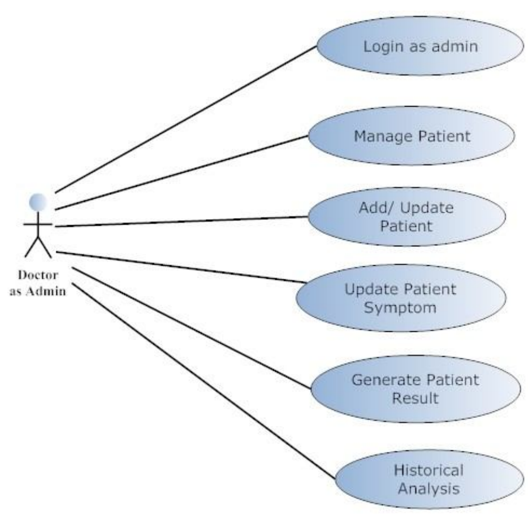

Fig 6 : Use case Diagram

V. CONCLUSION

In this paper we apply a combined genetic-neuro-fuzzy approach to the heart disease diagnosis problem. The objective of the work is to find the presence of heart disease. The proposed work also helps to minimize the cost and maximize the accuracy. Feature selection or extraction is an important part of this research. With the help of feature selection process, the computation cost decreases and also the classification performance increases.

REFERENCES

[1] 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.

[2] Awang, R. &Palaniappan, S., “Intelligent heart disease predication system using data mining technique”. IJCSNS International Journal of Computer Science and Network Security. Vol. 8, No. 8, 2008.

[3] Patil, S. &Kumaraswamy, Y., “Intelligent and effective heart attack prediction system using data mining and artificial neural network, “European Journal of Science Research. Vol 31, No. 4.2009.

[4] P.K. Anooj, “Clinical Decision Support System : Risk Level Prediction of Heart Disease Using
Decision Tree Fuzzy Rules” , IJRRCS, Vol. 3, No. 3, pp. 1659-1667, June 2012.

[5] 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.

[6] Ms. Preeti Gupta, Ms. Punam Bajaj, “Heart Disease Diagnosis System Based On Data Mining And Neural Network”, International Journal Of Engineering Sciences & research Technology ISSN: 2277-9655 Scientific Journal Impact Factor: 3.449 (ISRA), Impact Factor: 1.852

[7] MuratKarabatak, M. CevdetInce,"New feature selection method based on association rules for diagnosis of erythemato-squamous diseases", ELSEVIER Expert Systems with Applications 36 (2009) 12500–12505.

[8] Sean N. Ghazavi, Thunshun W. Liao, “Data mining by fuzzy modeling with selected features", ELSEVIER Artificial Intelligence in Medicine (2008) 43, pp.195—206.

[9] K.Priya, 2 T.Manju,3R.Chitra, “ Predictive Model Of Stroke Disease Using Hybrid Neuro-Genetic Approach”, International Journal Of Engineering And Computer Science :2319-7242 Volume 2 Issue 3 March 2013 Page No. 781-788.

[10] K.Srinivas, Dr.G.Raghavendra Rao, Dr.A.Govardhan , “Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques”, The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010