

CARDIOVASCULAR DISEASE PREDICTION USING GENETIC ALGORITHM AND NEURO-FUZZY SYSTEM

Sneha Nikam¹, Priyanshi Shukla² and Megh Shah³

Abstract--Cardiovascular disease (CVD) is related to heart and a major cause of morbidity and transience in the modern society. Diagnosis of cardiovascular disease using various medical tests is an important but complicated task which should be performed accurately. Hence a powerful tool in the prediction of heart disease with lower cost has become the need of time. A very scarce number of the systems predict 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. In this paper, we will apply NFS to the dataset which is nothing but the risk factors, for prediction and training of network will be done using back propagation algorithm and weight optimization will be done by genetic algorithm. The NFS combines neuro adaptive capability and fuzzy logic reasoning for prediction.

Keywords—Heart diseases risk factors, Prediction and Diagnosis systems, Genetic algorithm, NFS.

I. INTRODUCTION

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 be 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 diagnose the disease early and correctly[1].

Fuzzy Systems is used for solving a wide range of problems in different application domains. The neuro-fuzzy system (NFS)[7] model combines the adaptability of fuzzy inputs with neural network for accurate prediction. 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 use of Genetic algorithm is to improve the learning of the neuro-fuzzy system[7] which makes the training fast and also enhances its performance. Genetic algorithms[6] are commonly used for better solution due to its operators like selection, crossover and mutation. Accurate and reliable decision making in cardio logical prognosis can help in the planning of suitable surgery and therapy, and generally, improve patient management through the different stages of the disease.

The proposed problem thus obviously is related to unawareness among people and their resulting disregard for proper medical care especially related to cardio-logical problems. Thus this system aims to spread awareness among people by accurately predicting if they are at a potential risk of contracting a heart disease and thereby make them pro-active in making healthier life choices and follow regular checkups.

The remaining sections of the paper are organized as follows: In Section 2, a brief review of some of the works on heart disease prediction is presented. System design is described in Section 3. System implementation is explained in Section 4. The conclusions are summed up in Section 5.

II. PROPOSED SYSTEM

Soft Computing and AI 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. 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 etc. Heart

¹ Department of Computer Engineering St.Francis Institute of Technology (SFIT) Mumbai, India

² Department of Computer Engineering St.Francis Institute of Technology (SFIT) Mumbai, India

³ Department of Computer Engineering St.Francis Institute of Technology (SFIT) Mumbai, India

disease patients have lot of these visible risk factors in common which can be used very effectively for diagnosis.

Hence in this paper we are giving a technique for prediction of heart disease using major risk factors. This technique involves neural networks and genetic algorithms. A reliable method for both feature selection and classification is required. The feature selection is based on genetic algorithm and system learning is based on neuro-fuzzy system (NFS)[2]. 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. The hybrid system implemented uses the global optimization advantage of genetic algorithm. The learning is fast, more stable and accurate as compared to simple back propagation. The proposed system will be implemented in Matlab.

Neural networks mimic our ability to adapt to circumstances and learn from past experiences, fuzzy logic addresses the imprecision or vagueness in input and output. Genetic algorithms are inspired by biological evolution, can systemize random search and reach to optimum characteristics. Each of these technologies has provided efficient solution to wide range of problems belonging to different domains. However, each of these technologies has advantages and disadvantages. It is therefore appropriate that Hybridization of these three technologies is done so as to overcome the weakness of one with the strength of other.

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 [1].

The following is an overview of how our system will work is shown in fig. 1.

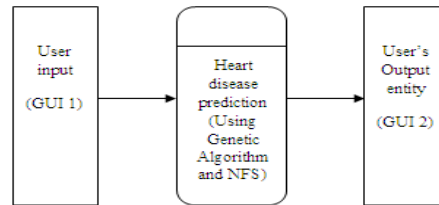


Figure. 1 System Flow Diagram

The user input will be fed to the system. The system will accurately predict the user’s possibility of contracting a heart disease.

In order to train the proposed system the input needed would have to be a comprehensive dataset. This will be obtained from the Cleveland University database which is freely available for use. The dataset will be further compacted and processed by taking 14 precise parameters that contribute as the risk factors for heart diseases. The dataset to be used is represented in Fig. 2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
2	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
3	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
4	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
5	37	1	3	130	250	0	0	167	0	3.5	3	0	3	0
6	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
7	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
8	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
9	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
10	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
11	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
12	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
13	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
14	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
15	44	1	2	120	263	0	0	173	0	0	1	0	7	0
16	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
17	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
18	48	1	2	110	229	0	0	168	0	1	3	0	7	1
19	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
20	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0

Figure. 2 Dataset representation

The output that would be shown to the user is proposed to be of the following format as shown in fig. 3. The user would insert the values for the required parameters and submit to the system for prediction purpose.

Figure. 3 System accepting user input

The system predicts the possibility of a person being prone to heart disease. The system also depicts a graph stating the risk percent that the user faces based on the predictive evaluations [Fig. 4].

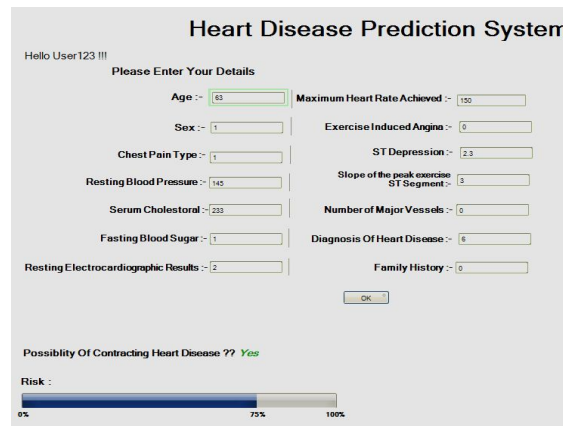


Figure. 4 System displaying predicted results

The system would be bound by accuracy constraints at preliminary levels which would be overcome as the system intercepts more and more distinguished inputs. Thereby enabling further system learning.

III. IMPLEMENTATION

The implementation would be done with the help of feature selection and a sequential execution of algorithms namely; Genetic Algorithm followed by system learning implemented via Neuro-Fuzzy system. The final system would be implemented as shown in Fig. 5.

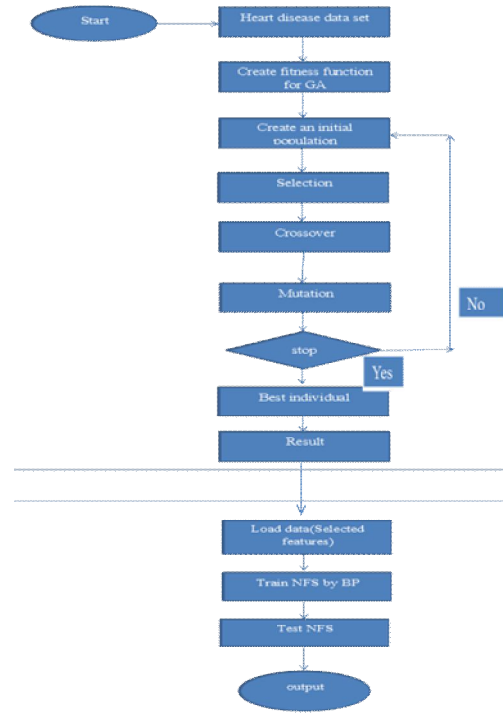


Figure. 5 Flow Chart of proposed work

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. GA 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. GA causes 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 given in (1).

$$fitness = \frac{w_f}{\sum_{i=1}^{n_f} f_i} \quad (1)$$

W_f for the number of features; f_i is the mask value of the i^{th} 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.

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) as $parent1$ 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
8. $d1$, and the distance between $child$ and $parent2$ as $d2$.
9. Calculate the fitness of $child$, $parent1$, and $parent2$ as f , $f1$, and $f2$ respectively.
10. if ($d1 < d2$) and ($f > f1$) then
11. replace $parent1$ with $child$
12. else
13. if ($d2 \leq d1$) and ($f > f2$) then
14. Replace $parent2$ with $child$.
15. end if
16. end for
17. Extract the best (highly-fitted) individuals as your final solution.

The results of the Genetic Algorithm including the 14 parameters selected would be further fed to the NFS system to implement system learning [Fig. 6].

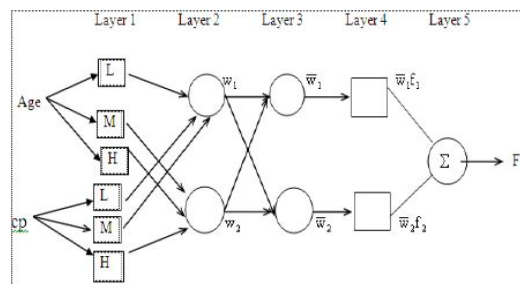


Figure 6: Architecture of the neuro-fuzzy system

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 parameters are determined by Gaussian membership function in order to ensure optimal selection. Refer (2)

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

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 is generated. The output of the Layer 2 is given by(3)

$$W_i = \mu_{Age}(x) \times \mu_{cp}(y); \quad (3)$$

w_i 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 i^{th} node calculates the portion of the i^{th} rules firing strength to the sum of all rules firing strengths. Equation (4) is

$$w_i = \frac{w_i}{\sum_{i=1}^m w_i} \tag{4}$$

where the output are called normalized firing strengths 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 is given by (5).

$$w_i f_i = w_i \mu_{Age}(x) + \mu_{cpi}(y) + \dots \tag{5}$$

Where w_i 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. It 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. Refer (6)

$$\sum_i \overline{w_i f_i} = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

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. The neuron activation functions are modified as the error is propagated. To determine the necessary modifications, the back-propagation algorithm differentiates the activation functions of the neurons.

The result would thus be a Binary output stating whether the individual is prone to contract a heart disease or not [Fig.5]. Along with this, the system also aims at displaying graphs that project the degree of risk that the user faces, in terms of percentage [Fig 4]. The results can be used by the user/common public in either of the following ways:

Either implement changes in lifestyle and become pro-active in curbing the risk factors so as to reduce the chances of contracting a heart disease. Or, in case the person has tested positive for the possibility of a heart disease, the person could schedule a visit to the specialist doctor and go ahead with all the required tests and treatments that the doctor recommends.

The main objective of this system is to help the common public become more aware of their lifestyle in terms of heart related complications and help save cost and effort spent over tedious health check-ups by limiting the need to carry out expensive tests if and only if the person has tested positive for the possibility of a heart disease.

IV. CONCLUSION

In this paper application of a neuro-fuzzy genetic approach is proposed to predict the cardiovascular disease. The objective of the work is to predict chances of cardiovascular disease. The proposed work also helps to make system more accurate and efficient. With the help of genetic algorithm error rate will be reduce and implementing the combination of neuro-fuzzy system will increase the classification performance and accuracy.

But the fact is computer cannot replace humans hence comparison between computer prediction result with lab report, doctors opinions can give best way to evaluate the auto prediction results.

REFERENCES

- [1] 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.
- [2] 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.
- [3] P.K. Anooj, "Clinical Decision Support System :Risk Level Prediction of Heart Disease Using Decision Tree Fuzzy Rules", International Journal of Research and Reviews in Computer Science(IJRRCS), Vol. 3, No. 3, pp. 1659-1667, June 2012.
- [4] K.Priya, T.Manju, R.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.
- [5] Chaitrali Dangare, Dr. Mrs. Sulabha Apte "A data mining approach for prediction of heart Disease using neural networks", International Journal of Computer Engineering & Technology(IJCET).Volume 3, Issue 3,Pp. 30-40, October - December (2012).
- [6] Georges R. Harik, Fernando G. Lobo, and David E. Goldberg "The Compact Genetic Algorithm", IEEE Transaction on Evolutionary Computation, Vol.3 , No.4, November1999.
- [7] Leszek Rutkowski and Krzysztof Cpałka "Flexible Neuro-Fuzzy Systems", IEEE Transaction on Neural Network, Vol.14, No.3,May2013.