# Hybrid Architecture of Heart Disease Prediction System using Genetic Neural Network

**Kumkum Chaudhary**
Department of Computer Science
SNDT University, Mumbai, India
kumkum171997@gmail.com

**Radhika Naidu**
Department of Computer Science
SNDT University, Mumbai, India
radhufx@gmail.com

**Rhea Rai**
Department of Computer Science
SNDT University, Mumbai, India
rheapremrai@gmail.com

**Narendra Gawai**
Assistant Professor
Department of Computer Science and Technology
SNDT University, Mumbai, India
ngawai@rediffmail.com

***Abstract*— Data mining techniques have been used in disease diagnosis, disease risk evaluation, patient monitoring, robotic handling of surgeries and predicting effect of new medicines but lack behind in certain factors such as accuracy, speed, performance etc. This paper proposes and evaluates Neural Network and Genetic Algorithm for diagnosing risk of heart disease. Risk factors viz. Blood Pressure, Blood Chest Pain Type, Heart Rate, Cholesterol, ECG, Diabetes, Sex, Physical Activity, etc. have been taken as inputs to the system. The system classifies the input samples and predicts whether heart disease is present or absent. The results of proposed system have been compared with system which was developed by using traditional algorithms; in terms of Accuracy, Mean Square Error and Regression and found better. This hybrid system will predict the presence of heart disease in more efficient manner.**

***Keywords — Genetic Algorithm, Neural network, Naive Bayes, Decision Tree, Mean Square Error, Risk factors.***

## I. INTRODUCTION

The current scenario for the diagnosis of heart disease uses clinical dataset having parameters and inputs from complex tests conducted in labs. None of the present 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. Patient suffering from heart disease have the above mentioned risk factors in common which can be used to diagnose diseases effectively.

This paper proposes a system which is computer-based clinical decision support and can reduce medical errors, improve patient safety and reduce unnecessary changes in practice, and improve the prognosis of the patient's medical history to integrate patients. The main objective of this study is to develop a prototype of heart disease forecasting system using data mining and neural network concepts. a huge knowledge and accurate data in the field not only helps users by providing effective treatment, but also help to reduce the cost of treatment and improve the visualization and ease of explanation.

There are some methods in the literature individually to diagnose heart disease such as Decision Tree, Naive Bayes, K-means, SVM, etc. These algorithms possess many drawbacks such as they cannot handle large, noisy and missing data which leads to difficulty in interpreting the result. There is no automated diagnosis method to diagnose the disease. The system based on above mentioned risk factors would not only help medical professionals but also 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. Medical organizations invest heavily in this type of activity in order to focus on the risks involved and possible events.

## II. REVIEW OF LITERATURE

Several data mining algorithms are used to find pattern that can be used for predicting and in decision support areas. For developing effective and intelligent heart disease prediction system author Priyanga, P., and N. C. Naveen(2017) used Naıve Bayes technique for classifying task of data mining. In this system, patients must provide values to the attributes for getting precise result. By using UCI dataset; data was trained. Trained data was compared with user input value. Traditional data mining techniques does not yield accurate result but Na¨ıve Baye's managed to yield result close to accuracy. For classification purpose Na¨ıve Bayes was used and result was in the form of low, average, high and very high. Basically here classification and prediction both were performed. Accuracy of the system is dependent on algorithm and database used, and here Naıve Bayes algorithm got 86% accurate result which is more than all other traditional data mining techniques[7]. Purushottam, Kanak Saxena, Richa Sharma(2015) designed a system using decision tree algorithm that creates rules for predicting heart disease. Decision tree algorithm is used for solving regression and classification problems too. Decision Tree is to create a training model which can use to predict class. Decision tree solves problem by using tree representation. In tree representation internal nodes corresponds to attribute leaf nodes corresponds to class label. Accuracy of the system was 90% which was better, depending on its performance. Reasonable accuracy in result of predicting system can be provided by using a single data mining technique. To boost accuracy level hybrid data mining techniques can be used. Malav, Amita and kalyani kadam(2017) carried out predictive analysis which was done on UCI dataset by using K-means and ANN algorithms. A model was developed which was used for classification using ANN and K-means. The aim was to classify the data

according to heart disease better which will lead to more reliable and efficient diagnosis[4]

## III. System Architecture

The architecture of the proposed system is as displayed in the figure below. The major components of the architecture are as follows: patient database, preprocessing, tokenization, training the model, test the model, design fitness function, application of genetic algorithm, results collection and prediction of heart disease.

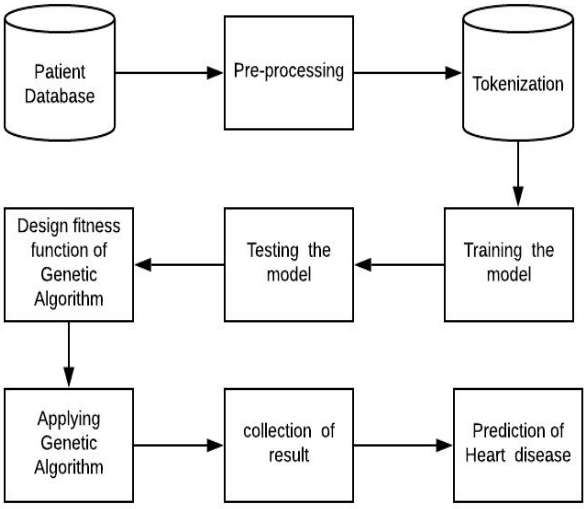


Figure 1. System architecture

*1. Patient database*

The dataset as provided by University of California, Irvine Machine Learning Repository is initially imported for the analysis of this system work. The dataset consists of the following attributes: age, chest pain, blood pressure, cholesterol, diabetes, ECG, heart rate, physical activity, slope, thalassemia, ca and sex. The final predicted attribute will be specified in 'num'. The attributes are further elaborated in Table 1.

2. Preprocessing

Preprocessing is a significant stage in the knowledge discovery process. Real world data tends to be noisy and inconsistent. Data processing techniques like data cleaning etc help in overcoming these drawbacks. Normalization of the dataset helps in classify the data which further makes the data to smoothly allow algorithms to execute with efficient results. To carry out normalization, normalize function is used. this helps in bifurcating the data into classes. Then a variable will be created that is 'num' which will hold the predicted attribute.

3. Tokenization

In tokenization, the data will be clubbed into set of meaning sentences or chunks for further processing. This will further enhance the efficiency of the data that has undergone preprocessing.

4. Training the model

In the training part, the backpropagation algorithm as mentioned above will be implemented. backpropagation helps in finding a better set of weights in short amount of time. The training is done on basis of the dataset input to the system. Herein 'min max' function is implemented so as to gain a matrix of minimum and maximum values as specified in its argument. This function is applied for training of the network. The efficiency of the system can be improved every instance as many times the model is trained, the number of iterations etc. The whole dataset provided which consists of 13 attributes and 872 rows will help the model undergo training. Training can also be implemented by splitting the data in equalized required amount of data partitions. In the user interactive GUI, as the user will select train network option after entering his data at the backend the .csv file of UCI dataset will be read and normalization will be carried out so as to classify the data into classes which becomes easier to be fed onto the neural network. the neural network that is created here will be consisting of three layers namely: input layer, hidden layer and output layer. Hidden layers can be customized to 2 or 3 as per users requirements. To generate a network, train() function is implemented so as to pass the inputs. this network will be stored in .mat file. After the network is generated, we check for mean square error.

5. Testing the model

Testing will be conducted so as to determine whether the model that is trained is providing the desired output. As the data is entered for testing, the .csv file will be retrieved to crosscheck and then compare and the results of the newly entered data will be generated. On basis of how the model is trained with the help of the dataset, the user will input values of his choice to the attributes specified and the results will be generated as the whether there is a risk of heart disease or not.

6. Design fitness function of genetic algorithm

The genetic algorithm is applied so as to initialize neural network weight. The genetic algorithm is used to evaluate and calculate the number of layers in the neural network along with the total number of weights used and bias. The initial population is generated at random. Bias is used such that the output value generated will not be 0 or negative. On basis of the mean square error calculated during testing, the fitness function of each chromosome will be calculated. Ater selection and mutation is carried out in genetic algorithm, the chromosome consisting of lower adaptation are replaced with optimizedd one that is better and fitter chromosomes. If at all, the best fit is not selected (worst fit is selected) then the process continues until the best fit is selected. This genetic algorithm concept along with Multilayer Feed Forward Network is used to predict the presence or absence of cardiovascular disease in the patient.

7. Prediction of heart disease

This component will help in predicting the severity of the cardiovascular disease. When user will input data, the weights will will be cross checked with the given inputs. The prediction neural network will consist of 13 nodes as a part of input layer considering that 13 attribute values will be input to the system. Then the hidden layer and one node in the output layer which will provide the result.

The predicted will be generated in the form of a 'yes' or 'no' format considering all the risk factors whether they lie in the criteria as per the model is trained.

### *A. Flowchart*

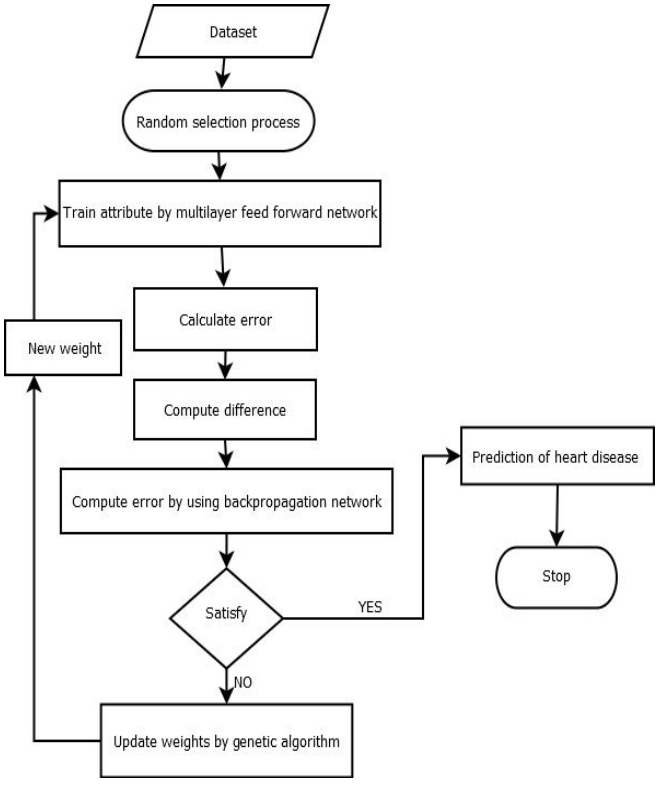


Figure 2. Genetic neural network algorithm flowchart

The above diagram depicts basic flow of the system. Initially, dataset is imported into the matlab software and selection process is performed. The dataset is then trained to recognise the pattern using multilayer feed forward network. Back propagation technique is carried out to recognize the pattern and genetic algorithm to optimize the weights . After all the processes are successfully completed, result is obtained in the form of classifier i.e. Yes or No.

## IV. PROPOSED DESIGN ARCHITECTURE

### *A. Dataset*

There are many disease prediction systems which do not use some of the risk factors such as age, sex, blood pressure, cholesterol, diabetes, etc. Without using these vital risk factors; result will not be much accurate. In this paper; 12 important risk factors are used to predict heart disease in accurate manner. Dataset is imported from UCI Machine Learning Repository.[1] This system is developed using MATLAB 2015.

| Attribute | Description | Domain value |
|---|---|---|
| Age | Age in years | 20-34(-2),<br>35-50(-1),<br>51-60(0),<br>61-79(1),<br>>79(2) |
| Chest pain | Chest pain type | Typical angina(1)<br>Atypical angina(2)<br>Non-anginal(3)<br>Asymptotic(4) |
| BP | Blood pressure | Below 120 mm Hg-Low(-1),<br>120-139 mm Hg-Normal(0),<br>Above 139 mm Hg- High(1) |
| Cholesterol | Cholesterol | Below 200 mg/DL-Low(-1),<br>200-239 mg/DL-Normal(0),<br>240 mg/DL and above -High(1) |
| Diabetes | Blood sugar | Yes(1)<br>No(0) |
| ECG | Resting ECG result | Normal(0)<br>ST-T wave abnormality(1)<br>LV hypertrophy(2) |
| Heart Rate | Maximum heart rate achieved | 71 to 202 |
| Physical Activity | Exercise induced angina | Yes(1)<br>No(0) |
| Oldpeak | ST depression induced by exercise relative to rest | 0-6.2 |
| Slope | Slope of peak exercise ST segment | Upsloping(1)<br>Flat(2)<br>Downsloping(3) |
| Ca | Number of major vessels coloured by fluoroscopy | 0-3 |
| Thal | Defect type | Normal(3)<br>Fixed defect(6)<br>Reversible defect(7) |
| Sex | Sex | Male(1)<br>Female(0) |
| Num | Heart disease | 0-1 |

Table 1. Heart disease patient dataset

https://archive.ics.uci.edu/ml/datasets/heart+Disease

### B. *Genetic Algorithm*

The technique mentioned in this paper will optimize the weights of neural network. It deals with the population i.e individual input string. First it will select the input string and assign a fitness value. Based on those fitness value a new offspring will be generated. Then followed by the crossover process it will generate possibly a fit string so as to obtain optimized weight. The new string generated at each stage is possibly a better than the previous one. This is how the weights are optimized at each stage of genetic process.

Following steps are used to optimize the weights :-
Step 1: First initial population is randomly selected.

Step 2: Each chromosome is evaluated using fitness function.

Step 3: A selection process is done using fitness function to generate new population.

Step 4: New generated population goes through crossover and mutation process.

Step 5: After all the process are done based on fitness function it will decide which weight are optimized to feed into neural network.

### C. *Artificial Neural Network*

After the weights are optimized it is fed into neural network which uses back propagation technique to train the network. The process of neural network consist of activation function which is calculated at hidden layer and output layer. The weights obtained at output layer will be compared with the previous weights so as to calculate error. By calculating the error new weights will be generated and it will again fed into neural network. This process will continue until the error function is minimum.

Following steps are used in neural network :
Step 1: In first step weights are initialised.

Step 2: Forward propagate : At this stage we first pass the input through the layer and straight calculate the output.

Step 3: After output is generated loss function is calculated. Loss function is difference between the desired output and the actual output.

**Loss = Desired - Actual**

Step 4: Again the weights are optimized to reduce loss function.

Step 5: This is the step where back propagation takes place. The output obtained is again fed into neural network.

Step 6: Weight updation :

**New weight = old weight - Derivative Rate *learning rate**

Step 7: It will iterate until weights are converged.

## V. Comparison

| Features\Algorithms | Decision Tree | Naive Bayes | K-Means | ANN |
|---|---|---|---|---|
| Training of data set | Low | Low | Low | High |
| Type of data in dataset | Interval or categorical data | Numeric | Categorical and numeric data | All types of data |
| Resource consumption | Medium | High | Medium | High |
| Estimation or prediction | No estimation but prediction is possible | Yes | Yes | Yes |
| Explicitness | No | No | No | Yes |
| Size of datasets | Small | Small to intermediate | Small to intermediate | Small to intermediate |
| Ability to handle missing data | Moderate | Moderate | No | Yes (High Ability) |
| Training Speed | Fast | Fast | Fast | Moderate |
| Prediction Speed | Moderate | Moderate | Moderate | High |
| Average predictive accuracy | Low | Low | Low | High |
| Can handle large amount of data processings | Yes | No | No | Yes |
| Easy to understand | Yes | Yes | Yes | Moderate |

Fig.8 Comparison between different algorithms

## VI. Result and Discussion

In this paper, the framework for genetic algorithm and neural network for analyzing medical information. This medical data will be analyzed and processed to predict how

much is the severity of having a heart disease. The dataset also undergoes preprocessing for polishing of the data. This also helps in classifying the data for further processing. This data is then classified into classes and backpropagation algorithm along with genetic algorithm is implemented onto it. This output helps in generating the prediction of a cardiovascular disease. Herein all the attributes are taken into consideration for predicting the result.

## VII. Conclusion and Future work

Hence, we have developed a hybrid system using genetic and neural network which will predict presence or absence of heart disease. In real life, many relationships are non-linear and complex; neural network has an ability to learn and model such kind of relationships. Thus making the model generalize and predicts unseen data. System which is developed with the help of algorithm that learns hidden relationship in the data without imposing fixed relationships in the data; will definitely yield much accurate result. As future work, we will work to predict presence of heart disease with reduced number of risk factors. Fuzzy Logic whose high precision and rapid operation is a key feature can be implemented with genetic and neural network to enhance system.

## References


[1] UCI Machine Learning Repository: Flags Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/heart Disease. [Accessed: 14-Feb-2019].

[2]. Kanchan, B. Dhomse, and M. Mahale Kishore. ”Study of machine learning algorithms for special disease prediction using principal of component analysis.” Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016 International Conference on. IEEE, 2016.

[3]. Jabbar, M. Akhil, B. L. Deekshatulu, and Priti Chandra. ”Classification of heart disease using artificial neural network and feature subset selection.” Global Journal of Computer Science and Technology Neural and Artificial Intelligence 13.3 (2013).

[4]. Malav, Amita, and Kalyani Kadam. ”A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means.”

[5]. Shrivastava, Shiva, and Neeraj Mehta. ”Diagnosis of Heart Disease using Neural Network.” Blood 1 (2016): 4.

[6]. Saxena, Kanak, and Richa Sharma. ”Efficient heart disease prediction system using decision tree.” Computing, Communication and Automation (ICCCA), 2015 International Conference on. IEEE, 2015.

[7]. Priyanga, P., and N. C. Naveen. ”Web Analytics Support System for Prediction of Heart Disease Using Naive Bayes

[8]. Saxena, K. and Sharma, R. (2019). *Efficient heart disease prediction system using decision tree - IEEE Conference Publication*. [online] Ieeexplore.ieee.org. Weighted Approach (NBwa).” 2017 Asia Modelling Symposium (AMS). IEEE, 2017.

[9]. Bhargava, Neeraj, et al. ”An approach for classification using simple CART algorithm in WEKA.” Intelligent Systems and Control (ISCO),