

# Heart Disease Prediction using Data Mining

Sarvesh Chowkekar<sup>1</sup>, Rohan Ringe<sup>2</sup>, Viral Gohil<sup>3</sup>, Naina Kaushik<sup>4</sup>

<sup>1</sup> Student VIII SEM, B.E., Computer Engg. , RGIT, Mumbai, India

<sup>2</sup> Student VIII SEM, B.E., Computer Engg. , RGIT, Mumbai, India

<sup>3</sup> Student VIII SEM, B.E., Computer Engg. , RGIT, Mumbai, India

<sup>4</sup> Professor, Department of Computers, RGIT, Mumbai, India

\*\*\*

**Abstract**—Heart disease 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. 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. 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. The system implemented uses the global optimization advantage of genetic algorithm for initialization of neural network weights.

**Keywords** - data mining, heart disease, risk factors, prediction, Genetic Algorithms (GA).

## 1. INTRODUCTION

Heart diseases are the number one cause of death globally: more people die annually from Heart diseases than from any other cause. 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. 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 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 diagnosis the disease early and correctly. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This practice may lead to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patient.

Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity are major threat in occurrence of heart disease.

## 2. LITERATURE REVIEW

[1] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg  
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 for initialization of neural network weights. The learning is fast, more stable and accurate as compared to back propagation. The system was implemented in Matlab and predicts the risk of heart disease with an accuracy of 89%.

[2] Latha Parthiban and R.Subramanian  
In this paper, a new approach based on coactive neuro-fuzzy inference system (CANFIS) was presented for prediction of heart disease. The proposed CANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the presence of the disease. The performances of the CANFIS model were evaluated in terms of training performances and classification accuracies and the results showed that the proposed CANFIS model has great potential in predicting the heart disease.

[3] Kavita Rawat, Kavita Burse  
The proposed paper shows a method performs feature selection and parameters setting in an evolutionary way. The wrapper approach to feature subset selection is used in this paper because of the accuracy. The performance of the ANFIS classifier was evaluated in terms of training performance and classification accuracy. The objective of this research is to simultaneously optimize the parameters and feature subset without degrading the ANFIS classification accuracy. To verify the effectiveness of the proposed approach, it is tested on ovarian cancer dataset.

[4] Dr. Anooj P.K  
The proposed clinical decision support system for risk prediction of heart patients consists of two phases, (1) automated approach for generation of weighted fuzzy rules, and (2) developing a fuzzy rule-based decision support system. In the first phase, we have used the mining technique, attribute selection and attribute weightage method to obtain the weighted fuzzy rules. Then, the fuzzy system is constructed in accordance with the weighted fuzzy rules and chosen attributes. Finally, the experimentation is carried out on the proposed system using the datasets

obtained from the UCI repository and the performance of the system is compared with the neural network-based system utilizing accuracy, sensitivity and specificity

### 3. SYSTEM ANALYSIS

#### A. Problem Definition

One of every three deaths in the U.S. in 2013 were from heart disease and other cardiovascular diseases, while heart disease and stroke were the No. 1 and No. 2 killers worldwide according to American Heart Association's 2016 Heart Disease and Stroke statistics Update. Early detection of medical problems such as ovarian cancer, prostate cancer and diabetes is important to increase the chance of successful treatment. 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. 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.

#### B. Proposed System

A system for automated medical diagnosis would enhance medical care and reduce costs. System 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 check-ups.

With using hybrid data mining techniques we could design more accurate clinical decision support systems for diagnosis of diseases. We can build an intelligent system which could predict the disease using risk factors hence saving cost and time to undergo medical tests and checkups and ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases.

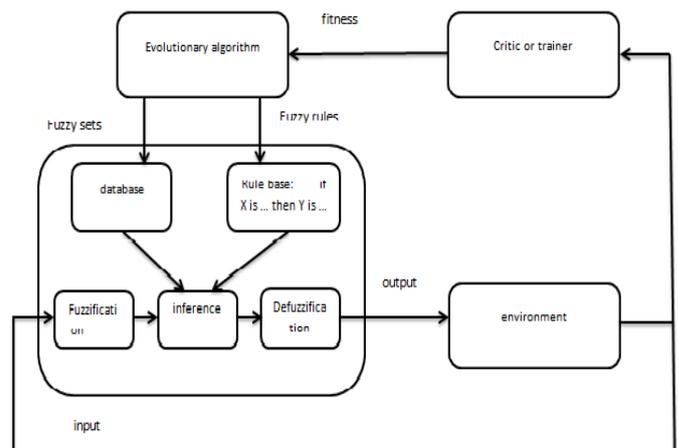


Figure 1 : block diagram of neuro-fuzzy with genetic algorithm

### DATA MINING TECHNIQUES

Data mining techniques are used to explore, analyze and extract medical data using complex algorithms in order to discover unknown patterns. The problem with risk factors related to heart disease is that there are many risk factors involved like age, usage of cigarette, blood cholesterol, person's fitness, blood pressure, stress and etc. and understanding and categorizing each one according to its importance is a difficult task.

### NEURO FUZZY

First step in the Genetic Neuro Fuzzy System is data fuzzification. Fuzzy Variables can represent the knowledge of the Subject Matter Experts more properly as compared to actual values. Each input variable has been represented as fuzzy variables using triangular fuzzy membership functions named 'Normal' and 'High'. The range of values for each of the membership variables has been taken as suggested by the medical experts. The first step is to initialize the weights of neural network using the 'configure' function available in MATLAB. Then these configured weights are passed to the genetic algorithm for optimization according to the fitness function.

### Genetic Algorithm

Genetic Algorithm A Neuro Fuzzy System can be obtained by feeding the fuzzified data into the Back Propagation Network. To get more reliability with larger number of input parameters, the back propagation network has been initialized with optimized weights and biases. Thus a Genetic Neuro Fuzzy System has been developed in which the Genetic

Algorithm has been used for optimizing the initial weights and biases of Neuro Fuzzy System.

**Risk Factors**

	Risk factors	values
1	Sex	Male(1) or Female(0)
2	Age(years)	20-30(-2),31-50(-1),51-60(0),61-80(1),>81(2)
3	Blood Cholesterol	Below 200 mg/dL – low(-1) 200-239 mg/dL –normal(0) 240 mg/dL and above – high(1)
4	Blood Pressure	Below 120 mm Hg -Low(-1) 120-129 mm Hg -normal(0) Above 139 mm Hg – High(1)
5	Smoking	Yes(1) or No(1)
6	Alcohol intake	Yes(1) or No(0)
7	Physical inactivity	Low (-1) , Normal(0) or High(1)
8	Diabetes	Yes(1) or No(0)
9	Diet	Poor(-1),Normal(0) or Good(1)
10	Stress	Yes(1) or No(0)
output	Heart Disease	Yes(1) or No(0)

Table 1 : Risk factors

**4. CONCLUSIONS**

In this project, we apply 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]Genetic Neural Network Based Data Mining in Prediction Heart Disease Using Risk Factors

Syed Umar Amin<sup>1</sup>, Kavita Agarwal, Dr. Rizwan Beg, Department of Computer Science & Engineering Integral University, Lucknow, India syed.umar.amin@gmail.com, kavitalucknow@gmail.com, rizwanbeg@gmail.com

[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] <http://www.heartfailurematters.org/>