

New Improved Genetic Algorithm for Coronary Heart Disease Prediction

Waheeda
Dhokley
M.H.S.S.C.O.E
Mumbai-08

Tahreem Ansari
M.H.S.S.C.O.E
Mumbai-08

Naeema Fazlani
M.H.S.S.C.O.E
Mumbai-08

Heena
Mohd.Hafeez
M.H.S.S.C.O.E
Mumbai-08

ABSTRACT

Heart disease prediction is treated as most complicated task in the field of medical sciences. Thus there arises a need to build a decision support system for detecting heart disease of a patient. Almost all system predicting heart disease use inputs from complex tests conducted in labs. In this project we are developing a system which will predict heart based on the risk factors such as tobacco, smoking, alcohol intake, age, family history, diabetes, hypertension, high cholesterol, physical inactivity, obesity. These common risk factors can be used effectively for diagnosis of heart disease[1]. System based on the such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of the heart disease even before he/she visits a hospital or goes for costly medical checkups.

General Terms

Artificial Intelligence, heart disease, risk factor

Keywords

Genetic Algorithm, Coronary Heart Disease, Prediction

1. INTRODUCTION

Heart disease is one of the most major reasons of death in India or other Asian countries. In 2003 approx. 17.3 million people died all over the globe and out of this, 10 million were only due to the coronary heart disease [2]. A long without changing lifestyle there are many such factors such as smoking, alcohol, obesity, high blood pressure, diabetes etc. which are responsible for the risk of having heart problem. However, with the resent studies, with the introduction of artificial intelligence and medical sciences, we can actually help in preventing any such kind of diseases. In the health care industry, data mining plays an important role for predicting diseases. For detecting a disease number of tests should be required from the patient [3]. But with the help of data mining technique the number of test should be reduced. This reduced test plays an important role in time and performance [4].Genetic Algorithm which uses the phenomena of mutation and crossover over various generations. The weights which are used for BP can be optimized first and then given as input to our network to give much better results. Fig 1.1 shows the Architecture of most important risk factors.

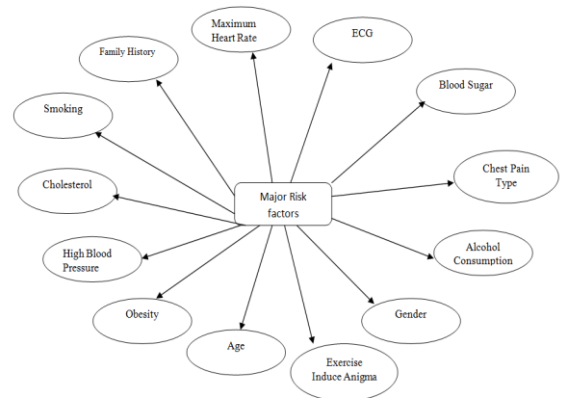


Fig 1: Architecture of most important risk factors.

As in this project we are going to make use of genetic algorithm for faster results .We take the input to user in terms of risk factors and then based on these risk factors we predict whether that person having heart disease or not. These risks are explained in detail below:

Table 1. The risk factor for heart disease

Sr. No.	Risk Factors	Ranges
1.	Family History	0-No 1-Yes
2.	Smoking	0-No 1-Yes
3.	Cholesterol	150 mg/dL above
4.	High Blood Pressure	120/80 mmHg above
5.	Obesity(Physical Inactivity)	0-No 1-Yes
6.	Age	1-If Female AND age>=55 1-If Male AND age>=45 0-Otherwise
7.	Gender	0-Female 1-Male
8.	Alcohol Consumption	0-No 1-Yes

9.	Chest pain Type	1-Typical Angina 2-Atypical Angina 3-Non Anginal 4-Asyngtomatic
10.	Blood Sugar	120 mg/dL above
11.	ECG	0-Normal 1-having ST wave abnormal 2-left_ventricular hypertrophy
12.	Exercise Induce Anigma	0-No 1-yes
13.	Maximum Heart Rate	100 bpm above

2. GENETIC ALGORITHM

In the field of artificial intelligence, a genetic algorithm (GA) is a search heuristic method that imitates the process of natural evolution. This heuristic method is routinely used to generate useful solutions to optimization and search problems [5]. The input to the algorithm is called as chromosome that contains the parameters that have unique characteristics [6]. Each chromosome consists of a collection of genes.

A gene expresses the characteristic of the input. A collection of such chromosomes form a population. The solution is obtained with the inclusion of fitness function, the genetic operators like selection, crossover and mutation in every generation. Its main advantage is that GA only uses the fitness function but not gradient or other attached information in the optimizing process [7].

The GA's fitness function is required to estimate the individuals' optimization degree by optimizing computation. Those individuals who have much greater fitness will have more chances to be duplicated to the next generation.

2.1 Process of Genetic Algorithm:

- i. *Chromosome representation:* Represents the strings into suitable encoded format. The three genetic operators are briefly outlined below[8]:
- ii. *Selection* is the process of selecting the parents among the population so that they can be need for crossover and mutation operations. Selection method represents the mechanism that determines the number and the type of parent chromosome to be selected. It is one of the traditional selection techniques. The chromosome is selected based on the probability proportional to its fitness value [9]. Fitter chromosomes have a better chance to be selected in this approach.
- iii. *Crossover* is an operator that selects the parents from the pool of population and interchanges the location of the values based on the crossover point fixed.
- iv. *Mutation* is a process of flipping or changing the gene values based on its given probability (mutation probability) value in binary or real valued representation respectively. Mutation operator accelerates and explores the search space

[8]. It helps to escape from local minima and an appropriate value for this operation can direct to the optimal solution as it maintains the diversity in the population.

Genetic algorithms combine the high performance notions to achieve better performance for getting optimal solution'

3. EXISTING SYSTEM

Heart disease is a disease that effects on the operation of heart. Now a day, in the world Heart disease is the major reason of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases [10]. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost across 23.6 million people will die due to Heart disease as written in. In Existing system, the single data mining technique is used to diagnose the heart disease [11].

There is no previous research that identifies which data mining technique can provide more reliable accuracy in identifying suitable treatment for heart disease patients.

One of the systems has used Back-Propagation in neural network which is regarded as the best prediction algorithm where we have a non-linear relationship between the data and the target output.

So, accuracy reaches high enough to 75%. In this project we are developing a system which will predict heart based on the risk factors such as tobacco smoking, alcohol intake, age, family history, diabetes, hypertension, high cholesterol, maximum heart rate, poor ECG, physical inactivity, obesity and thus help healthcare experts and practitioners to make intelligent clinical decisions which traditional decision support systems cannot[12]. Disadvantages of existing system:

- Misuse of information
- More time consuming
- Not able to provide the optimal solution

4. PROPOSED SYSTEM

In the proposed system we are going to use genetic algorithm for faster results. The system predicts more accurately the disease with reduced number of attributes. The propose system will be take input as a risk factors and then system will perform preprocessing to give probability that the person having heart disease or not.

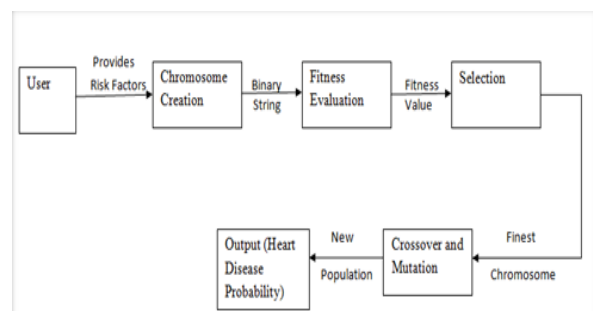


Fig 2: Block Diagram of Proposed System

4.1 Proposed Algorithm

Genetic Algorithm is used to determine the attributes which contribute more towards the diagnosis of heart diseases. It consists of Selection, Cross over, Mutation, Accepting.

- i. Select two parent's chromosomes from a population according to their fitness.
- ii. Create a pair of offspring chromosomes by applying the genetic operators - crossover and mutation.
- iii. Place the created offspring chromosomes in the new population.

4.2 Genetic Algorithm

Step 1: Takes risk factors from the user i.e. Family history, Cholesterol, Smoking etc.

Step 2: Pass this input to Chromosome that gives binary string as output.

Enter String: genetic

Enter Samples: genezzz

pqrstio

Based on match create chromosome as follows:

1111000

0000110

Step 3: Calculate Fitness value for each samples as follows:

Fitness values 4

2

Step 4: Perform Selection operation based on some condition i.e. samples having fitness value greater than or equals to 2 will be selected.

Selected String: genezzz

pqrstio

Step 5: Perform Crossover operation.

Step 6: Perform Mutation operation if needed to get better output.

Step 7: Finally the result of patients will be predicted that patient having heart disease or not.

4.3 Flow Chart

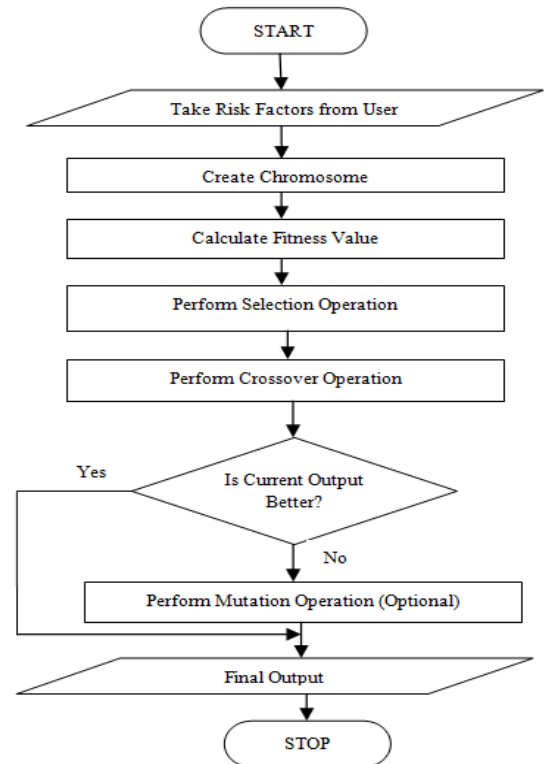


Fig 3: Flow Chart of Proposed System

5. CONCLUSION

Almost all system predicting heart disease use inputs from complex tests conducted in labs.

In this project we are developing a system which will predict heart based on the risk factors such as tobacco smoking, alcohol intake, age, family history, diabetes, hypertension, high cholesterol, physical inactivity, obesity.

These common risk factors can be used effectively for diagnosis of heart disease.

6. FUTURE ENHANCEMENT

With the help of this proposal we can also predict stroke disease problem in future easily.

From this proposed system we can get approximate 98% accurate outputs and in future we can get 100% complete and accurate optimal solution.

7. ACKNOWLEDGMENTS

The co-authors of this paper express their gratitude to our first author and internal project guide **Er. Waheeda Dhokley** for valuable suggestions, guidance and constant encouragement for successful completion of the project.

The authors of this paper also thanks to all other faculty and non-teaching staff members of our department for their support and peers for having stood by us to complete this project synopsis report.

8. REFERENCES

- [1] Dawan A, Sharma M, " Prediction of heart disease using a hybrid technique in data mining classification", Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference, - 11-13 March 2015
- [2] Tabatabaie M, Hooman N, Arjmandi-Rafsanjani K, Isafreshi R, "Ambulatory blood pressure monitoring for children with β -thalassemia major: a preliminary report", US National Library of Medicine National Institute of Health-2013 Jul;7
- [3] POONAM, PAWAR, DEEPALI VORA, "A Survey On Prediction of Heart Disease for Diabetic Patients", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015
- [4] S. Sudha, "Disease Prediction in Data Mining Technique – A Survey", International Journal of Computer Applications and Information Technology, ISSN: 2278-7720- 2013;2(1)17-21
- [5] <http://www.seminarstopics.com/seminar/disease-prediction-using-data-mining>
- [6] Optimize by Generation, RapidMiner Studio Core, http://docs.rapidminer.com/studio/operators/data_transformation/attribute_space_transformation/generation/optimization/optimize_by_generation_yagga.html
- [7] Game Development, using chromosomes to go from input to output in Genetic Algorithm, <http://gamedev.stackexchange.com/questions/96708/using-chromosomes-to-go-from-input-to-output-in-genetic-algorithms>, Accessed on 25 January 2016
- [8] Anand A. Chaudhari, " Fuzzy & Data Mining based Disease Prediction Using K-NN Algorithm", International Journal of Innovations in Engineering and Technology (IJJET)
- [9] T. Santhanam and E. P. Ephzibah " Heart Disease Prediction Using Hybrid Genetic Fuzzy Model", Indian Journal of Science and Technology, Vol 8(9), 797–803, May 2015
- [10] Kristian Guillaumier, " Generic Chromosome Representation and Evaluation for Genetic Algorithms"
- [11] Sikha Dixit, Appu Kuttan. K.K, " ARTIFICIAL NEURAL NETWORK BASED DATA MINING APPROACH FOR HUMAN HEART DISEASE PREDICTION", International Journal of Computer Engineering and Technology (IJCET), Volume 5, Issue 6, June (2014)
- [12] Ibrahim Umar Said, Jamila M. Muhammad, Manoj kumar Gupta, " Intelligent Heart Disease Prediction System by Applying Apriori Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 9, September 2014