



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: VI Month of publication: June 2019

DOI: <http://doi.org/10.22214/ijraset.2019.6366>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Heart Disease Prediction System using Genetic Algorithm

Shaikh Sadaf Bano¹, Sarala Mary²

¹Student, ²Professor Dept. of Institute of Computer Science, MET College, Maharashtra, India

Abstract: Humans in this modern era follow a lifestyle which is sedentary following an eight to nine hours work load, due to which cardio vascular diseases is becoming one of the biggest reason for mortality worldwide. To reduce this, hospitals of healthcare industries collect information about their patients and their treatment and ailments in huge amount. Diagnosing heart diseases or cardio vascular diseases in specific is not an easy task and also is not predictable by many doctors, there are very few doctors who can predict any type of heart disease accurately. With the help of Optimization Techniques, this research paper will focus on developing a Prediction Algorithms. We will be using the Genetic Algorithm technique which is an inherently distributed algorithm where the solution can be obtained from many simple individual agents interacting between themselves. The data source we have used for experimental testing are commonly used and considered as a de facto standard for heart disease prediction reliability ranking. The results obtained show that Genetic Algorithm Technique are not only competitive with other evolutionary techniques, but also with industry standard algorithms, and can be successfully applied to heart disease prediction.

Keywords: Heart Diseases, Genetic Algorithm, Coronary Heart Disease.

I. INTRODUCTION

Heuristic search is a method which gives possible solution which can either be a point in space or a path from the initial state. It may not guarantee the best solution but will give good solution. Artificial Intelligence is the processing of computer system intelligently using various algorithms and get the best output. Genetic Algorithm is one of the artificial intelligent technique which uses operations of crossover and mutation to get the optimum result. In the paper we have explored genetic algorithm that fits good to predict any heart disease given any dataset.

The health care data which are not extracted to discover information for effective decision-making are collected in a huge amount by health care industry. Clinical decisions are often taken on the basis of doctors' perception and experience rather than the on the knowledge rich masked in the dataset. However, not every doctor is expert in diagnosis of every heart disease. The details given by the patients may entail interrelated symptoms, redundant data and signs in medical diagnosis especially when the patients suffer from more than one type of disease of the same category. Unfortunately, due to complex interdependence on a variety of factors, accurate diagnosis of disease at a premature stage is quite a challenging task. To extract data which is useful for surgical procedures and medical services from a large set of data, we use the Extraction Transformation and Loading process which will help us to remove null values, missing values, inappropriate data and ambiguity from the dataset. This processed dataset will help the medical diagnosis of any heart patient. We need to transform the data as the way the algorithm would expect it to take it as input to give the result.

A. Heart Diseases

Coronary Artery disease is when the small blood vessels that supply blood and oxygen to the heart becomes smaller or narrows to an extent that proper supply is not reachable to the heart, this disease is also called as Coronary Heart Disease.

This condition occurs when fatty material and plaque starts residing on the walls of the arteries which causes the passage of the blood vessels to narrow down.

Major risk factors involved in CHD are: a) High blood pressure b) Diabetes c) Obesity d) Smoking.

II. LITERATURE SURVEY

Chaitrali S. Dangare and Sulabha S. Apte [1] proposed a new data mining approach for prediction of heart disease using Neural Networks techniques like Backpropagation neural network (Multilayer perceptron neural network). In this approach they have used two more attributes to get accurate results for the heart disease prediction.

M.Akhil jabbar a*, Dr.Priti Chandrab , Dr.B.L Deekshatulu, [2] proposed Heart Disease prediction system using associative classification and genetic algorithm like Associative classification along with genetic algorithm to improve accuracy of the rule generation. But the proposed algorithm was Time consuming.

Nitika, Madan Lal Yadav,[3], proposed A Fuzzification Approach for Prediction of Heart Disease which used Fuzzy rule based system. This technique requires creation of fuzzy rules based on the fuzzy inputs.

Curt G. DeGroff, MD; Sanjay Bhatikar et al,[4], proposed Artificial Neural Network–Based Method of Screening Heart Murmurs in Children in which they used ANN for processing the sound of heartbeats collected from 69 patients.

P.K.Anooj,[5], proposed Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. The paper presented a weighted fuzzy rule-based system with automatic generation of fuzzy rules, in which the attributes were selected based on frequency of attributes which were then used to generate fuzzy rules.

R. Chitra and Dr.V. Seenivasagam,[6], proposed, Heart Disease Prediction System Using Supervised Learning Classifier, using Cascaded Neural Network (CNN) classifier, in which the ANN structure is the fixed size of the neurons in the hidden layer.

Shivapratap Gopakumar ; Truyen Tran et al,[7], proposed, Stabilizing High-Dimensional Prediction Models Using Feature Graphs, using Laplacian-based regularization into a regression. Model to reduce variance in the selected features.

III. DATASET DESCRIPTION

The data set is taken from Data Mining Repository of the University of California, Irvine (UCI) (Newman et al., 1998). Finally the system is validated using data sets from Cleveland, Hungary and Switzerland. In those datasets there are various attributes which is collected, but only, 14 attributes such as Age, sex, chest pain type, resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina, ST depression, and slope of the peak exercise ST segment, number of major vessels, thal and diagnosis of heart disease are presented.

A. Cleaveland dataset [9]

Robert Detrano, M.D., Ph.D., collected these data at V.A. Medical Centre. All published experiments related to using a subset of 14 of the 76 attributes present in the processed Cleveland heart disease database.

Specifically, researchers use only the Cleveland database till today. The existence of heart disease in the patient is indicated in the goal field by means of an integer that can take any value from 0 (no presence) to 4.

Distinguishing disease existence (values 1-4) from non-existence (value 0) has been the focus of the experiments conducted in the Cleveland database (Ephzibah, 2010). Six of the examples have been discarded because they had missing values. Class distributions are 54% heart disease absent, 46% heart disease present.

Only 14 attributes used:

1. **(age)** age in years
2. **(sex)** (1 = male; 0 = female)
3. **(cp)** chest pain type –
 - a) Value 1: typical angina
 - b) Value 2: atypical angina
 - c) Value 3: non-anginal pain
 - d) Value 4: asymptomatic
4. **(trestbps)** resting blood pressure (in mm Hg on admission to the hospital)
5. **(chol)** serum cholesterol in mg/dl
6. **(fbs)** (fasting blood sugar >120 mg/dl) (1 = true; 0 = false)
7. **(restecg)** resting electrocardiographic results –
 - a) Value 0: normal
 - b) Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of >0.05 mV)
 - c) Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria.
8. **(thalach)** maximum heart rate achieved
9. **(exang)** exercise induced angina (1 = yes; 0 = no)
10. **(oldpeak)** ST depression induced by exercise relative to rest
11. **(slope)** the slope of the peak exercise ST segment
 - a) Value 1: upsloping
 - b) Value 2: flat

- c) Value 3: downsloping
12. **(ca)** number of major vessels (0-3) colored by fluoroscopy
13. **(thal)** 3= normal; 6 = fixed defect; 7 = reversible defect
14. **(num)** (the predicted attribute) diagnosis of heart disease (angiographic disease status)
- a) Value 0: 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

IV. PROPOSED SYSTEM

A. Selection

Selection is a process in which the genes are chosen from the population which fits best for breeding. All the genes in the population are evaluated for their fitness. The ones which are fit undergoes for selection process.

B. Crossover

It is also called as recombination .It is one way to stochastically generate new solutions from an existing population. Solutions can also be generated by cloning an existing solution, which is analogous to asexual reproduction. Newly generated solutions are typically mutated before being added to the population.

$$\text{Recombination Frequency} = \frac{\text{Recombinants}}{\text{Total Offsprings}} \times 100$$

C. Mutation

From the Initial state of the chromosomes, mutation alters one or more genes values. From the previous solution, the solution which is generated again can be changed entirely due to mutation .Hence mutation can be used for better solution in Genetic Algorithm. Mutation occurs during evolution according to a user-definable mutation probability. This probability should be set low. If it is set too high, the search will turn into a primitive random search.

D. Genetic Algorithm

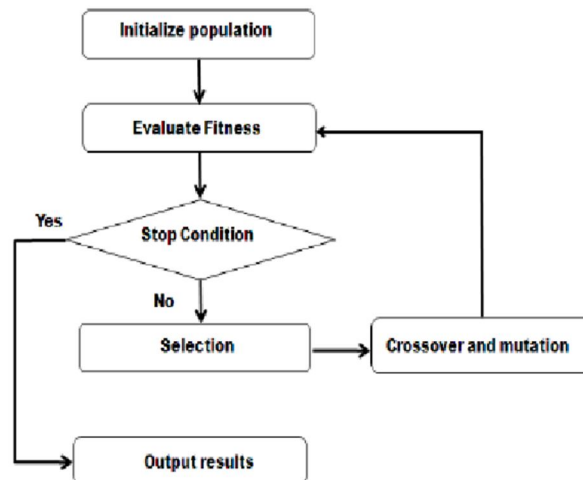


Fig 1: Flow Diagram of genetic algorithm [8]

In computer science and operations research, a genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA).

Genetic algorithms are based on the ideas of natural selection and genetics. Genetic algorithms simulate the process of natural selection which means those species who can adapt to changes in their environment are able to survive and reproduce and go to next generation.

Genetic Algorithm works as a searching algorithm, it works on a population of attributes which will help in predicting heart diseases in Human heart. Population is the collection of candidate solutions which will undergo the process of genetic algorithm. The fitness of the individual gene depends upon how good the solution represented by the individual is.

The fitness is good if the solution provided by an individual is good.

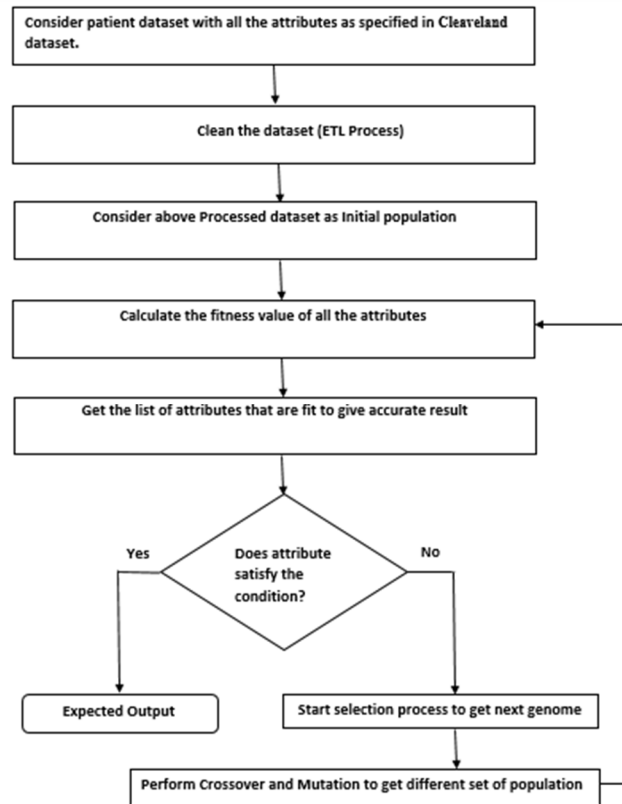


Fig 2: Flow Diagram of Proposed System

Step 1: We consider all the details taken from the patients in the hospital as our dataset as discussed earlier in the paper, the dataset acts as the Initial population in our algorithm

Step 2: We do the cleaning process to remove the missing data from dataset, to make sure the data is complete and no ambiguity is present using the Extract, Transform and Load Process (ETL).

Step 3: From the population we evaluate the fitness of all the attributes which will help us to get better and accurate results.

Step 4: The Output of the step 3 will be the attributes which will be considered fit. We will apply the required conditions to match the values in the given dataset for predicting a specific Heart Disease on these attributes.

Step 5: If these attributes satisfy the conditions, then we will get the Expected Output.

Step 6: If the attributes do not satisfy the conditions then the attributes will go through the selection process to select the next genome for better solution.

Step 7: The output of the Selection process will go through Crossover and mutation where the attributes which are fit will be mutated to get a different set of population.

Step 8: The mutated attributes again go through the process of evaluating the fitness of the population.

Step 9: The Algorithm stops when we get the Expected results which gives us the information of the Heart disease for which the algorithm was used.

The 14 attributes which are considered are best fit attributes which can be used to predict ant type of heart disease in the Human Heart.

For example:

If person A has the following data:

Age: 30

Gender: 0



Chest pain: 2
Resting Blood pressure: 120
Cholesterol: 50
Fasting blood sugar: 100
Resting electrocardiographic: 2
Thalach: 80
Exang: 1
Old Peak: 100
Slope: 2
Ca: 3
Thal: 3
Num: 0

The result from the above data shows that Person A has symptoms of Coronary Artery Disease.

V. CONCLUSION

Heart diseases have become very common in India and other parts of country. We need to diagnose the type of disease early so that proper medication can be done to save the patients. From the Genetic algorithm, we can apply various Crossover and Mutation functions through which we can predict any Heart Disease. The proposed flowchart would give the optimum result and help to predict any type of heart disease. In the future, I would use particle swarm optimization technique and compare both the algorithm to see which algorithm gives the better result.

REFERENCES

- [1] Chaitrali S. Dangare and Sulabha S. Apte, data mining approach for prediction of heart disease using Neural Networks techniques, International Journal of Computer Engineering and Technology (IJCET),(IJCET) , ISSN 09766-076. 6367(Print), ISSN 0976 – 6375(Online) Volume 3, Issue 3, October-December (2012), © IAEME30
- [2] M.Akhil jabbar a*, Dr.Priti Chandrab , Dr.B.L Deekshatuluc : Heart Disease Prediction System using Associative Classification and Genetic Algorithm ,International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT, 2012 .
- [3] Nitika, Madan Lal Yadav, A Fuzzification Approach for Prediction of Heart Disease, International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue5- May 2013
- [4] Curt G. DeGroff, MD; Sanjay Bhatikar, PhD; Jean Hertzberg, PhD; Robin Shandas, PhD; Lilliam Valdes-Cruz, MD; Roop L. Mahajan, PhD, Artificial Neural Network–Based Method of Screening Heart Murmurs in Children, Circulation. 2001;103:2711-2716
- [5] P.K.Anooj, Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules, Journal of King Saud University - Computer and Information Sciences, Volume 24, Issue 1
- [6] R. Chitra and Dr.V. Seenivasagam, Heart Disease Prediction System Using Supervised Learning Classifier, Bonfring International Journal of Software Engineering and Soft Computing, Vol. 3, No. 1, March 2013
- [7] Shivapratap Gopakumar ; Truyen Tran et al, Stabilizing High-Dimensional Prediction Models Using Feature Graphs, IEEE Journal of Biomedical and Health Informatics (Volume: 19 , Issue: 3 , May 2015)
- [8] https://www.researchgate.net/figure/Genetic-algorithm-flowchart_fig2_263224226
- [9] <https://archive.ics.uci.edu/ml/datasets/heart+Disease>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)