CARDIOVASCULAR DISEASE PREDICTION USING BINARY AND GENETIC ALGORITHM - HYBRID METHODOLOGY

P. VAISHNAVI
SCSE, M.Tech Integrated Software Engineering, (Student) VIT University Chennai, Tamil Nadu, India
vaishnavi.p2018@vitstudent.ac.in

Y. LAKSHMI SAI CHARITHA
SCSE, M.Tech Integrated Software Engineering, (Student) VIT University Chennai, Tamil Nadu, India
ylakshmi.saicharitha2018@vitstudent.ac.in

T. JAYASRI
SCSE, M.Tech Integrated Software Engineering,(Student)VIT University Chennai, Tamil Nadu, India
thiyyagura.jayasri2018@vitstudent.ac.in

Abstract *-- Cardiovascular diseases are those diseases that related to heart. Heart diseases are not short term diseases like fever or cold. They take years of time to diagnose and are hard to detect and predict based on symptoms. It is 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. If there are any errors or mistakes in those predictions, the life of patient might be in danger. Hence a Powerful tool in the prediction of heart disease with lower cost has Become the need of time. Detection of such cardiovascular i.e heart diseases might be done with the help of some common symptoms like regular illness or even be predicted using risk factors such as age, family history, diabetes,hypertension ,high cholesterol, tobacco smoking, alcohol intake ,obesity or physical in-activity ,etc. A very scarce number of the systems predict heart diseases based on these risk factors. 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, we will Apply ANN and binary classification to the dataset which is nothing but the risk factors for Prediction and training of network*

Keywords **-- Cardiovascular diseases, genetic algorithm, neuro adaptive capability, ANN, Binary classification.**

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 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. Genetic algorithms are commonly used for better solution due to its operators like selection, crossover and mutation. Accurate and reliable decision making in cardiological prognosis can help in the planning of suitable surgery and therapy, and generally, improve patient management through the different stages of the disease. Prediction of diseases isn't an easy task to perform. We might even need more than one soft computing and machine learning, data mining techniques to understand the situation and predict

.

II. OBJECTIVE

- Identifying the best suitable model with highest accuracy and lowest overall error rates.
- Understanding the various neural networks for all the models.

III. ACKNOWLEDGEMENT

The methodology of this research includes identifying the most popular Throughout the writing of this research paper, we have received a great deal of support and assistance. We would first like to thank our first guide and tutor, Prof. Jagadeesh Kannan Raju whose expertise and guidance was invaluable in the completion of the research. In addition, we would like to thank our parents for their wise counsel and sympathetic ear. They were always there for us. Finally, there are our friends, who were of great support in deliberating over our problems and findings

IV. LITERATURE REVIEW

[1] A Novel Approach for Prediction of Cardio Vascular Disease: An Improved Genetic Algorithm Approach Using Classifiers: The proposed methodology is to find cardiovascular heart diseases and detect disease by using data

mining methods and genetic algorithm. Proposed ensemble structure contains three modules such as pre-processing and data acquisition. Classifier training for disease discovering as well as prediction with 3 layered strategy. [2] Adaptive genetic algorithm with fuzzy logic (AGAFL): The objective of adaptive genetic algorithm with fuzzy logic (AGAFL) model is to predict heart disease which will help medical practitioners in diagnosing heart disease at early stages. The 9 model consists of the rough sets-based heart disease feature selection module and the fuzzy rule-based classification module. [3] Heart Disease Prediction System using Genetic Algorithm: In this Heart Disease Prediction System using Genetic Algorithm paper simple genetic algorithm is used without any other combination of other soft computing techniques. [4] Hybrid Architecture of Heart Disease Prediction System using Genetic Neural Network: The main objective of Hybrid Architecture of Heart Disease Prediction System using Genetic Neural Network is to develop a prototype of heart disease forecasting system using data mining and neural network concepts. [5] Coronary Heart Disease prediction using genetic algorithm-based decision tree: The aim of Coronary Heart Disease prediction using genetic algorithm-based decision tree is to support the doctors in taking decision to classify healthy and coronary heart disease (CHD) patients using popular modified decision tree by using genetic algorithm. [6] Prediction of heart disease using Artificial Neural Network: In this Prediction of heart disease using Artificial Neural Network paper, they have compared the results of applying all the neural network techniques. [7] Study and Analysis of Prediction Model for Heart Disease: An Optimization Approach using Genetic Algorithm: The main idea behind writing this paper is to study diverse prediction models for the heart disease and selecting important heart disease feature using genetic algorithm. [8] Identification of heart disease using fuzzy neural genetic algorithm with data mining techniques: The Fuzzy validation of parameters of anomalous information smoothly decide the anomalies in the extensive size database. The presentation of fuzzy set idea effectively manages the vulnerability issues. [9] Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm: 13 One of the factors affecting the performance of artificial neural network is the initial weights utilized in the network structure. In this regard, the proposed model sought to ameliorate the performance of neural network through enhancing the primary weights used in it. [10] Design of Heart Disease Diagnosis System using Fuzzy Logic: Proposed diagnosis system shows the fuzzy membership functions, which have been used in the implementation part, fuzzy expert system designing (ranges for all input attributes) and fuzzy rule base. [11] Analysis of heart disease prediction system using artificial neural networks: Until now, all the research methodology has been focused on clinical parameters, but they can use additional attributes in consideration too in this research paper .The proposed method is focused on 4 stages. Stage 1 will be the data profiling and attribute selection in which clinical parameters-as shown in UCI parameters as well as risk factors –medical history, smoke , hypertension , hypotensive etc will be taken into consideration. Stage 2 will be the optimization phase to reduce the noise elimination in dataset. Stage 3 will be the neural network implementation with a combined approach where recurrent fuzzy neural network or radial basis function RBF neural network can be used for prediction and backpropagation for error reduction. Stage 4 will show the final predicted output showing presence or absence of the disease. [12] Detection of heart disease applying fuzzy logics and its comparison with neural networks: The process starts with data processing. The second step includes the reduction of number of attributes for the processing in the fuzzy method. After the successful development of fuzzy rule model the early stage prediction will be made lastly, the result will be analyzed and will be ended for development phase. [13] Enhanced decision tree algorithm using genetic algorithm for heart disease prediction: 15 In this paper the methodologies used are basic decision tree algorithm, genetic algorithm, enhanced decision tree algorithm based on genetic algorithm, fitness function for the rule, crossover and mutation operations for the rule. [14] Cardiovascular disease prediction using genetic algorithm and neuro-fuzzy system: In this paper, the heart disease prediction is done by using genetic algorithm and fuzzy system. The methodologies performed here are genetic algorithm, NFS(neuro fuzzy system), back propagation algorithm and feature selection. In this paper they applied 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. [15] Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks: This Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks proposes a genetic algorithm (GA) based trained recurrent fuzzy neural networks (RFNN) to diagnosis of heart diseases. The results showed that 97.78% accuracy was obtained from testing set. In addition to the accuracy, root mean square error, the probability of the misclassification error, specificity, sensitivity, precision and Fscore are calculated. The results were found to be satisfying based on comparision[16] A Computational Intelligence Method for Effective Diagnosis of Heart Disease using Genetic Algorithm: In this paper a novel method for the diagnosis of heart disease has been proposed using Genetic Algorithms. In this approach an effective association rules are inferred using Genetic Algorithm approach which uses tournament selection, crossover, mutation and new proposed fitness function.

V. PROPOSED METHODOLOGY

With the help of genetic algorithm, they enhanced or modified the decision tree, using genetic algorithm in decision tree functionality was extended. The model consists of the rough sets-based heart disease feature selection module and the fuzzy rule-based classification module. The generated rules from fuzzy classifiers are optimized by applying the adaptive genetic algorithm. First, important features which effect heart disease are selected by rough set theory. The second step predicts the heart disease using the hybrid AGAFL classifier. The adaptive genetic algorithm (AGA) is an improved version of the genetic algorithm, in which, adaptive mutations are employed for achieving desired optimizing results. **Decision Tree model**:
Model of computation in which an algorithm is considered to be basically a decision tree, i.e., a sequence of branching operations based on comparisons of some quantities, the comparisons being assigned unit computational cost. A genetic algorithm employs mutations to each parent chromosome, where random interchanging of genes occurs. In the proposed

adaptive mutation, the rate of mutation calculation is based on the chromosome's fitness. The performance of mutation is based on the rate of mutation. Fitness function was used in genetic algorithm to generate the optimized rules; more the fitness value more the optimal rule and vice versa. The optimized prediction models using genetic algorithm performance is better than traditional prediction models. It is a well-known technique to solve optimization problems. It is also employed to optimize the network parameters including learning rate, momentum of the network and the number of MFs(Membership Functions) for each input. Genetic
Algorithm could be a faster and more accurate alternative to the 24 existing time-consuming methodologies. We can also find the accurate decision even though there are less attributes in the dataset. It is an efficient and effective technique for both optimization and machine learning applications.

Genetic algorithm is good for noisy environments and it is stochastic and robust. Error rate will be reduced with the help of genetic algorithm. Genetic algorithm uses association rule mining concept to infer best and effective rules from the given input data set. In this genetic algorithm, we can use decision tree, naive bayes classifier, support vendor machine etc. for the prediction of heart disease. These classification algorithms are selected because they are very often used for research purposes and have been potential to give efficient results for heart disease prediction

VI. SYSTEM DESIGN

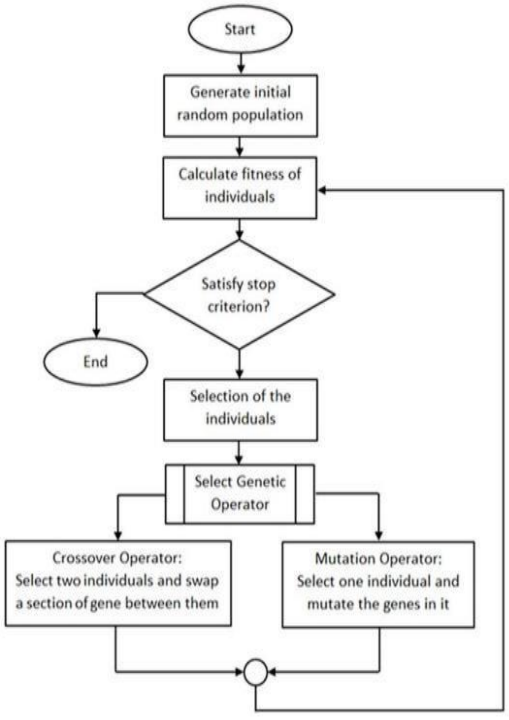

Fig_1: Architecture Diagram / Flow Diagram / Flowchart

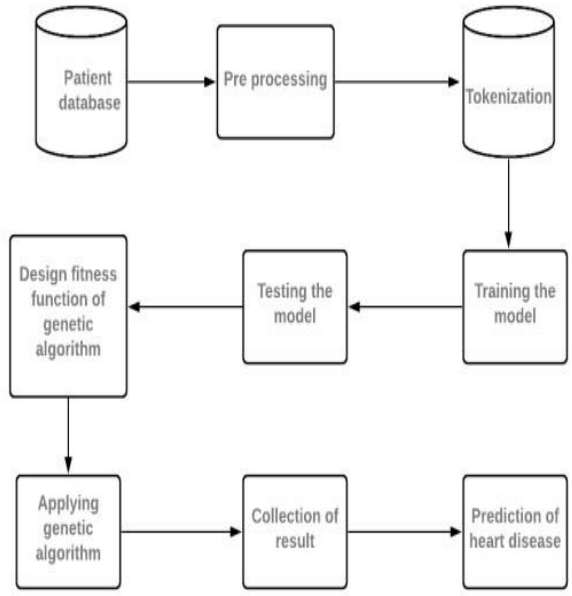

Fig_2: System Architecture

VII. DATASET INFORMATION

(revised data)

Data Set Characteristics:	Multivariate	**Number of Instances:**	303	**Area:**	Life
Attribute Characteristics:	Categorical, Integer, Real	**Number of Attributes:**	75	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	1469955

SOURCE

- Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
- 2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
- 3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
- 4. V.A. Medical Center, Long Beach and Cleveland Clinic.Foundation: Robert Detrano, M.D., Ph.D

CONTEXT – The leading cause of death in the developed world is heart disease. Therefore there needs to be work done to help prevent the risks of having a heart attack or stroke.

CONTENT – Use this dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

ABOUT DATASET – This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with the Cleveland database have concentrated on simply attempting to

distinguish presence (values 1,2,3,4) from absence (value 0). The names and social security numbers of the patients were recently removed from the database, replaced with dummy values. One file has been "processed", that one containing the Cleveland database. All four unprocessed files also exist in this directory. This dataset might be useful for us to perform training and testing of the algorithm that we are going to implement . As the central objective of the project is heart disease detection , we definitely need a heart dataset . So we considered this dataset.

Only	14	Attributes	Full description
1	#3	(age)	Patient Age
2	#4	(sex)	Male/Female
3	#9	(cp)	Chest pain type
4	#10	(trestbps)	Resting blood pressure (in mm Hg on admission to the hospital)
5	#12	(chol)	Serum cholestoral (mg/dl)
6	#16	(fbs)	Fasting blood sugar
7	#19	(restecg)	Resting ECG results
8	#32	(thalach)	Maximum heart rate achieved
9	#38	(exang)	Exercise induced angina
10	#40	(oldpeak)	ST depression included by exercise relative to rest
11	#41	(slope)	Slope of the peak exercise ST segment
12	#44	(ca)	Number of major vessels (0-3) colored by fluoroscopy
13	#51	(thal)	3 = normal, 6 = fixed defect, 7 = reversible defect.
14	#58	(num)	Angiographic disease status (Diagnosis of heart disease)

Fig_3:Dataset Attributes

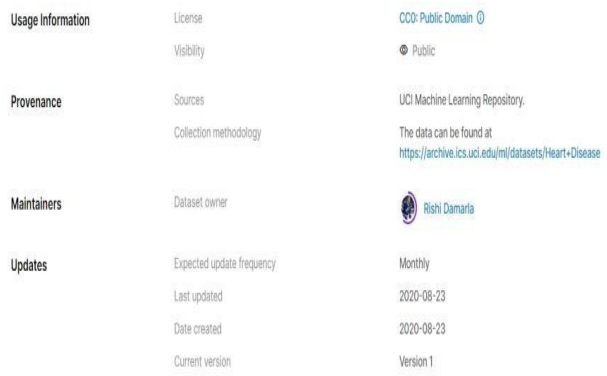

Fig_4:Metadata

VIII. MODULES

Patient database: Generally in hospitals certain patient databases are maintained for future use. We are using that data as our dataset from Kaggle website. This data is the main source of our project and with the help of database we will perform all the other operations in the flow chart.

Pre-processing: We have two types of attributes in the database; primary attributes(more important) and secondary attributes(less important). Through preprocessing we will refine the data by separating more important attributes from less one.

Tokenization: It is the process of turning sensitive data into non-sensitive data called "tokens" that can be used in a database or internal system without bringing it into scope. Tokenization can be used to secure sensitive data by replacing the original data with an unrelated value of the same length and format. We will replace the fuzzy values of the data as crisp values and change the data into bit strings so that the data can be easily used in genetic algorithm. Training the model: Training of the model is done by artificial neural network in which we will perform updation of weights with the help of old weights present in database. Then by using threshold value and activation function according to the data obtained we will compare and provide the output and updated weights as results.

Testing the model: Testing the gained results provide the accuracy of the model. We are performing testing through genetic algorithm as the best fitted chromosomes survives and the least fitted will be dead. This mechanism gives the performance of the model. The decision variable 'x' is coded into finite length string and initial population is selected randomly.

Designing fitness of genetic algorithm: Fitness Function (also known as the Evaluation Function) evaluates how close a given solution is to the optimum solution of the desired problem. It determines how fit a solution is. Then 'x' values are decoded for initial population.

Applying genetic algorithm: Here genetic algorithm comes into action. Genetic Algorithm (GA) is a search-based optimization technique based on the principles of Genetics and Natural Selection. It is frequently used to find optimal or nearoptimal solutions to difficult problems which otherwise would take a lifetime to solve. The sub tasks of genetic algorithm like producing child chromosomes from parent chromosomes is done by "crossover" and "mutation" techniques

***Collection of results*:** After crossover and mutation we will get best score of the child chromosomes and matched against their respected parent fitness score. If the child's score is greater than parents then child is best fitted and it can proceed for further survival, otherwise we have to repeat from testing module again till we get the best score.

***Prediction of heart disease**:* With the help of artificial neural network and genetic algorithm we can predict the accuracy of the model. Genetic based neural network is used for training the system. The final weights of the neural network are stored in the weight base and are used for predicting the risk of cardiovascular disease. The classification accuracy obtained using this approach is 81.3%.

IX. IMPLEMENTATION

- Data Import and Pre-processing
- Data Visualizations
- Create Testing and Training Datasets
- Classification Problem
- Performing Genetic Algorithm
- Building and training the Neural Networks.
- Final Results-Output

The implementation has been done in ***"Google Colab"***

X. RESULTS

```
[26]: #Results:
      # generate classification report using predictions for categorical model
      from sklearn.metrics import classification_report, accuracy_score

      categorical_pred = np.argmax(model.predict(X_test), axis=1)

      print('Results for Categorical Model')
      print(accuracy_score(y_test, categorical_pred))
      print(classification_report(y_test, categorical_pred))
```

```
Results for Categorical Model
0.7868852459016393
              precision    recall  f1-score   support

           0       0.83      0.68      0.75        28
           1       0.76      0.88      0.82        33

    accuracy                           0.79        61
   macro avg       0.79      0.78      0.78        61
weighted avg       0.79      0.79      0.78        61
```

Fig_5: Metrics of ANN algorithm for predicting heart disease

```
[27]: # generate classification report using predictions for binary model
      from sklearn.metrics import classification_report, accuracy_score
      # generate classification report using predictions for binary model
      binary_pred = np.round(binary_model.predict(X_test)).astype(int)

      print('Results for Binary Model')
      print(accuracy_score(Y_test_binary, binary_pred))
      print(classification_report(Y_test_binary, binary_pred))
```

```
Results for Binary Model
0.819672131147541
              precision    recall  f1-score   support

           0       0.90      0.68      0.78        28
           1       0.78      0.94      0.85        33

    accuracy                           0.82        61
   macro avg       0.84      0.81      0.81        61
weighted avg       0.83      0.82      0.82        61
```

Fig_6: Metrics of binary classification algorithm for predicting heart disease

Models	Accuracy
ANN Algorithm	0.79
Binary Classification Algorithm	0.82

XI. CONCLUSION

The Cardiac Muscle in the heart is one of the hardest working muscle groups in the entire human body. Beating over 72 times in a minute and more than 3 billion times in a lifetime, a salutary heart can sustain various biological functions. Prediction of 10-year risk of contracting Coronary Heart Diseases (CHDs) is therefore crucial for a prolonged life. The proposed system identifies the best set of Binary Classification Algorithm to achieve an accuracy of over 82%. ANN Algorithm has an accuracy over 79%. Although these models consider all the attributes of the dataset for building the model, a refined Genetic Algorithm approach for selecting the set of features that have the greatest influence on the target variable is achieved through selection, crossover, and mutation and elitism operations.

XII. REFERENCES

[1] Reddi, Sivaranjani & srinivasa naresh, Vankamamidi & Murthy, Nistala V.E.S.. (2019). Coronary Heart Disease prediction using genetic algorithm based decision tree. 10.1515/9783110621105-004. https://www.researchgate.net/publication/338273928_Coronary_Heart_Disease_prediction_using_genetic_algorithm_based_decision_tree

[2] Gadekallu, Thippa & Reddy, Praveen & Lakshman, Kuruva & Rajput, Dharmendra & Kaluri, Rajesh & Srivastava, Gautam. (2020). Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evolutionary Intelligence. 13. 10.1007/s12065-019-00327https://www.researchgate.net/publication/337552031_Hybrid_genetic_algorithm_and_a_fuzzy_logic_classifier_for_heart_disease_diagnosis

[3] Bano, Shaikh. (2019). Heart Disease Prediction System using Genetic Algorithm. International Journal for Research in Applied Science and Engineering Technology. 7. 2178- 2182. 10.22214/ijraset.2019.6366. https://www.ijraset.com/fileserve.php?FID=35495

[4] Kumkum Chaudhary, Radhika Naidu, Rhea Rai, Narendra Gawai. (2019). Hybrid Architecture of Heart Disease Prediction System using Genetic Neural Network. V6/i5/IRJETV6I5857 https://1library.net/document/y4we57vq-hybridarchitecture-heart-disease-prediction-genetic-neuralnetwork.html

[5] Uyar, Kaan & Ilhan, Ahmet. (2017). Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. Procedia Computer Science. 120. 588-593. 10.1016/j.procs.2017.11.283. https://www.researchgate.net/publication/321818545_Diagnosis_of_heart_disease_using_genetic_algorithm_based_trained_recurrent_fuzzy_neural_networks

[6] Awan, Shahid & Riaz, Muhammad & Khan, Abdul. (2018). PREDICTION OF HEART DISEASE USING ARTIFICIAL NEURAL NETWORK. 13. 102-112. https://www.semanticscholar.org/paper/Predictionof-Heart-Disease-using-Artificial-Neural-Awan-Riaz/ad1bd6004d6fbf9f13a9b24bde756c8b2a002c89

[7] Prasadgouda B Patil, Dr. P Mallikarjun Shastry, Dr Ashokumar P S. (2020). A Novel Approach for Prediction of Cardio Vascular Disease: An Improved Genetic Algorithm Approach Using Classifiers. International Journal of Advanced Science and Technology, 29(7s), 4493 – 4504 https://www.researchgate.net/publication/348391058_A_Novel_Approach_for_Prediction_of_Cardio_Vascular_DiseasAn_Improved_Genetic_Algorithm_Approach_Using_Classifiers

[8] Srikanth Meda1 , Raveendra Babu Bhogapathi 2 1 Research scholar, Acharya Nagarjuna University, Guntur & Associate Professor in the Department of Computer Science and Engineering at R.V.R. & J.C. College of Engineering, Guntur 522019, India 2 Professor, Department of Computer Science and Engineering, R.V.R & J.C College of Engineering, Guntur 552019, India 53. http://cse.rvrjcce.ac.in/faculty/brb.php

[9] Akruti Dave, Prof. Gayatri Pandi , Master of Engineering Student, Head of Department Department of Computer Engineering ,L. J Institute of Engineering &Technology (Gujarat Technological University), Ahmedabad, India. https://www.researchgate.net/profile/Gayatri-PandiJain

[10] Kasbe, Tanmay & Pippal, Ravi. (2017). Design of heart disease diagnosis system using fuzzy logic. 3183-3187. 10.1109/ICECDS.2017.8390044. https://www.researchgate.net/publication/325981519_Design_of_heart_disease_diagnosis_system_using_fuzzy_logic

[11] Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard , Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm , Computer Methods and Programs in Bio medicine , Volume 141,2017,Pages 19-26,ISSN 0169-2607. https://pubmed.ncbi.nlm.nih.gov/28241964/

[12] Zabeen, Ashiya & Utsav, Ankur & Lal, Kanhaiya. (2018). Detection of Heart Disease Applying Fuzzy Logics and Its Comparison with Neural Networks. 461-467. 10.1109/RTEICT42901.2018.9012315. https://www.researchgate.net/publication/339560617_Detection_of_Heart_Disease_Applying_Fuzzy_Logics_and_Its_Comparison_with_Neural_Networks

[13] Kumar, S., & Sahoo, G. (2018). Enhanced decision tree algorithm using genetic algorithm for heart disease prediction. International Journal of Bioinformatics Research and Applications, 14(1-2), 49-69. https://www.inderscienceonline.com/doi/abs/10.1504/IJBRA.2018.089199

[14] Nikam, S., Shukla, P., & Shah, M. (2017). Cardiovascular disease prediction using genetic algorithm and neuro-fuzzy system. http://dspace.sfit.co.in:8004/jspui/bitstream/123456789/909/1/Cardiovascular%20disease%20prediction%20using%20genetic%20algorithm%20and%20neuro%20fuzzy%20system%20full%20text.pdf

[15] Y.Harimadhava Reddy1 , Dr..K.Vijaya Bhaskar Reddy2 , D. Gireesh Kumar3 Assistant Professor, Dept. of EEE, Universal College of Engineering &Technology, Perecherla, Guntur, A.P, India Professor , Dept. of EEE , B V Raju Institute of Engineering &Technology, Narsapur, TS, India Assistant Professor, Dept. of EEE, B V Raju Institute of Engineering & Technology, Narsapur, TS, India http://www.ijareeie.com/upload/2016/rapideet/20_PP_RE1145_NEW.pdf

[16] Kumar, P. S., Anand, D., Kumar, V. U., Bhattacharyya, D., & Kim, T. H. (2016). A computational intelligence method for effective diagnosis of heart disease using genetic algorithm. International Journal of Bio-Science and Bio-Technology, 8(2), 363-372. https://www.semanticscholar.org/paper/AComputational-Intelligence-Method-for-Effective-Kumar-Anand/f28ef0312492bef246cc6d468371959d52c4edcc

[17] Salem, T. (2018). Study and analysis of prediction model for heart disease: an optimization approach using genetic algorithm. International Journal of Pure and Applied Mathematics, 119(16), 5323-5336. https://www.acadpubl.eu/hub/2018-119-16/2/543.pdf

Santhanam, T., & Ephzibah, E. P. (2015). Heart disease prediction using hybrid genetic fuzzy model. Indian Journal of Science and Technology, 8(9), 797. https://indjst.org/articles/heart-disease-predictionusing-hybrid-genetic-fuzzy-model