International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Predicting Heart Disease Using Machine Learning Algorithms.

Mihir Patel[1], Rahul Patange[2], Chaitanya Patil[3], Prof. Anuradha Kapoor[4]

[1]*Mihir Patel, Dept of Information Technology Engineering, Atharva College of Engineering*
[2]*Rahul Patange, Dept of Information Technology Engineering, Atharva College of Engineering*
[3]*Chaitanya Patil, Dept of Information Technology Engineering, Atharva College of Engineering*
[4]*Prof. Anuradha Kapoor, Dept of Information Technology Engineering, Atharva College of Engineering*

Abstract - As we all know heart is the most important part of our body other than brain. For having a healthy heart, there are many solutions available in market. Exercise can also play an important role for maintaining heart healthy. Apart from medical treatments, technology can also prove to be very useful in treating any heart disease. Any heart disease is predicted beforehand, then curing it would be not much complex. But predicting it would be a tough task. According to the survey by WHO (World Health Organization), Cardiovascular diseases are the leading cause of death globally. Estimate deaths per year are 17.9 Million. The datasets used are classified in terms of medical parameters. This system evaluates those parameters using data mining classification technique. The datasets are processed in python programming using Machine Learning Algorithm that is Logistic Regression Algorithm which shows the best algorithm in terms of accuracy level of heart disease.

***Key Words*: Heart Diseases, Machine Learning Algorithms, Logistic Regression, Random Forest, Decision Tree.**

1.INTRODUCTION

Heart attacks are the most commonly cause of death among all deadly disorders. Medical professionals perform several surveys on heart disorders in order to acquire information on heart patients, their symptoms, and the development of their disease. These days, heart attack is a common occurrence that can be fatal. Some symptoms foreshadowed the future. Medical science has made excellent use of technological breakthroughs to raise the standard of healthcare These technological developments have opened the path for precise illness diagnosis and prognosis Machine learning might be a great option for you obtain a high level of accuracy when it comes to forecasting heart illnesses As a result, three algorithms will be implemented.

It is an algorithms containing Logistic Regression Decision tree and Random forest. Furthermore, these three methods gives significantly more rapid and consistent outcomes. Predictions are becoming more straightforward as a result of technological advancements. People nowadays live in luxury all over the world, and they work like machines to acquire a lot of money and fame. People forget to take care of their health because of their hectic schedules. As a result, the food they eat and the way they live have changed. Blood pressure, diabetes, and a variety of other ailments develop in young people as a result of strain and stress in their lives. All of these factors contribute to the onset of heart disease.

1.1 Description

In this paper, the heart disease dataset of UCI repository is taken and subjected to various classification and clustering algorithms using python. The main focus is to target all possible combinations of the attributes against various algorithms. Then of all the techniques it is the technique that works the best to predict the heart disease at an early stage is identified.

Implementing 3 algorithms such as Decision tree, Random forest and Logistic regression would make it easier to identify and sort out the disease. Dataset is used to classify and train the model. After training the model, the most accurate and successful algorithm was later used to predict the disease.

1.2 Problem statement

The most difficult aspect of cardiac disease is detecting it. There are tools that can forecast heart disease, but they are either too expensive or too inefficient to quantify the risk of heart disease in humans. Early identification of heart disorders has been shown to reduce mortality and overall consequences. However, it is not possible to correctly monitor patients every day in all circumstances, and consultation with a doctor for 24 hours is not available since it takes more intelligence, time, and competence. We may use various machine learning algorithms to evaluate data for hidden patterns in today's environment since we have a lot of data. Hidden patterns in medical data can be utilised for health diagnosis.

1.3 Proposed Approach

People all across the globe now live luxury lives, and they work like machines to gain a lot of money and recognition. People fail to care on their health because of their hectic schedules. As a result, the food they eat has changed, as has their way of life. Because of the strain and stress in their

lives, they develop high blood pressure, diabetes, and a variety of other disorders at an early age. All of these factors contribute to the onset of heart disease. We are employing the Logistic Regression approach to develop an effective heart attack prediction system in this system. We can provide input to the system in the form of a CSV file or by entering it manually. Logistic Regression is an algorithm that is applied after accepting input. The operation is carried out after accessing the data set, and an effective heart attack level is created. The suggested method would incorporate some more criteria relevant to heart attacks, such as weight, age, and priority levels, after consultation with specialist doctors and medical professionals. The heart attack prediction system is intended to assist in identifying different risk levels of heart attack, such as normal, low, or high, as well as providing prescription data based on the expected outcome. The user can also register and login or else user can quickly predict the heart disease by just one click, but after quickly predicting the result the user's data will not be stored and neither user can access his/her previous record. After Successful registration and login the user can feed the values on the website and click on predict, on the very next screen the user can see the result . It will show the probability of heart disease in the percentage. If the percentage is greater than 60%, it will display "Your chance of having heart disease is high" and will provide prevention and symptoms as well as a doctor's contact information for further assistance; if the percentage is less than 60%, it will display "Your chance of having heart disease is low." The user data is saved in the database using their name and username, and any doctor who wants to add his/her details can do so using the admin panel. Our major goal is to save the prior record so that the user may see how much progress has been made by the user.

2. LITERATURE SURVEY

According to [1] numerous readings have been carried out to produce a prediction model using not only distinct techniques but also by relating two or more techniques. The dataset with a radial basis function network (RBFN) is used for classification, where 70% of the data is used for training and the remaining 30% is used for classification. The prediction model is introduced with different combinations of features and several known classification techniques. It produced an enhanced performance level of 88.7% through the prediction model for heart disease with Hybrid Random Forest with Linear Model (HRFLM).

Sakshi Goel, Abhinav Deep Shilpa Srivastava, Aparna Tripathi [2] has proposed a number of models for predicting heart diseases using different technologies such as artificial neural networks, machine learning, data mining, etc. This paper analyses the work done by various researchers on the accuracy of heart disease prediction through the different approaches. A detail literature review has been provided in the study. The analysis has also been presented on the basis on technology used. Data mining gives maximum accuracy of 92.1% with SVM algorithm and minimum accuracy of 89.6% with Decision Tree algorithm.

Mohini Chakarverti, Saumya Yadav, Rajiv Rajan [3] approached as the useful data is extract from raw dataset with the help of data mining approach. The alike and unlike information is clustered after measuring a resemblance among input dataset. Both alike and unlike data types are classified using KNN classifier. The KNN is the classifier which can classify the data based on the nearest neighbor. The execution time of the KNN classifier is also low as compared to SVM classifier. Here KNN has given the highest accuracy of 83% as compared to SVM. In future, the proposed approach will be further enhanced to design hybrid classifier for the prediction of heart diseases.

M.Preethi and Dr.J.Selvakumar [4] reviewed a literature survey that delivers the concept of various techniques has been studied for diagnosing the cardiovascular disease. Use of big data, machine learning along with data mining can provide promising results to bring the most effective accuracy in analyzing the prediction model .The main aim of this paper diagnosing the cardiovascular disease or the heart disease and using different methods and many approaches to get prediction.

3. COMPARISON ANALYSIS

For the comparison analysis, we have compared different machine learning algorithms like Decision tree, Logistic Regression, Random forest. The overall result was 77% for decision tree, 84% for random forest and 92% for logistic regression. From the comparison, we get to know that Logistic Regression is preferable as it gives better classification and higher prediction results for the detection of Heart disease diseases.

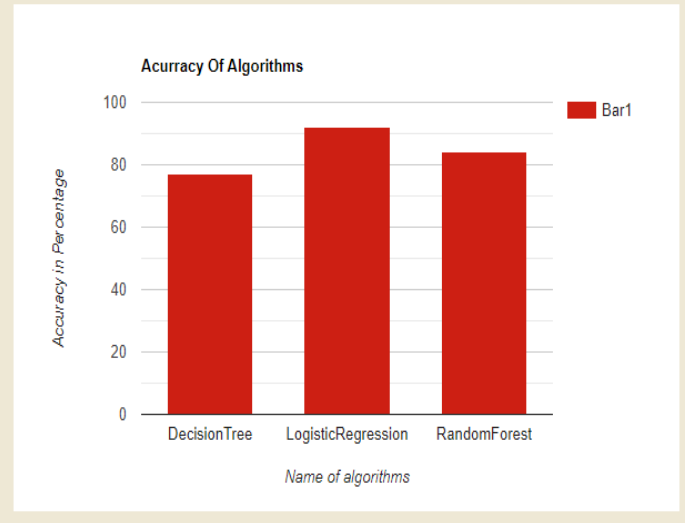
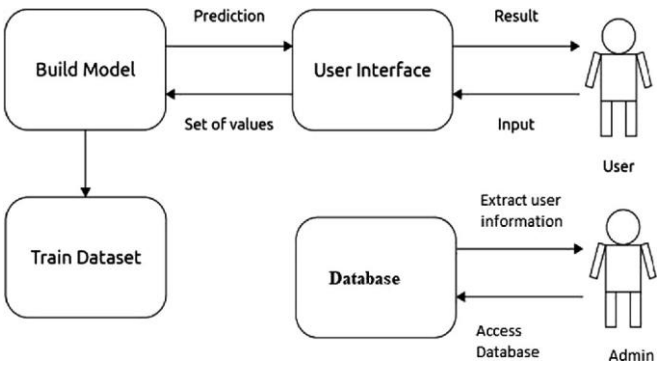


Chart 1- Chart showing prediction accuracy percentage for different algorithms.

4. ARCHITECHTURE



4.1 Block Diagram

In our project we have created a UI with the database, where user can register and login to predict their heart disease if user did not want to register and login than they can directly predict the heart disease on clicking on quickly predict button. After that the user have to insert the 13 attributes like age , gender, cholesterol level, etc. And after entering the details on basis of user details the model will be build and the dataset will be trained we have divided Test into 25% and Train into 75%. The build model and the percentage of heart disease risk will be displayed on the screen, if the percentage is more than 60% than there is the risk of heart disease and on our UI the user can get prevention and symptoms option. The user can get their previous records on our website if the user has login from same credentials if they have predicted earlier. If the model predicted the heart disease then the user have the option to contact directly via email or phone. All the user information is stored in the database to access whenever they required.

We have compared three algorithms which are Decision Tree, Logistic Regression and Random Forest. After comparison we have found that Logistic Regression has given the highest accuracy of 92% followed by Random Forest given accuracy of 84% and Decision Tree of 77%. So we have used Logistic Regression model in our project because of highest accuracy.

5. METHODOLOGY

A. Data Source

The dataset used here for predicting heart disease is taken from UCI Machine learning repository. UCI is a collection of databases that are used for implement machine learning algorithms. The dataset used here is real dataset. The dataset consists of 300 instances of data with the appropriate 14 clinical parameters. The clinical parameter of dataset is about tests which are taken related to the heart disease as like blood pressure level, chest pain type, electrocardiographic result and etc.

OBSERVATIONS	DESCRIPTION
Age	Age in years
Sex	Male/Female
Cp	Chest Pain
Trestbps	Resting Blood Pressure
Chol	Serum Cholestrol
FBS	Fasting Blood Sugar
Restecg	Resting Electrocardiograph
Thalach	Maximum heart rate achieved
Exang	Exercise Induced Angina
Oldpeak	Depression when workout compared to the amount of rest taken
Slope	Slope of peak Exercise
Ca	Gives the number of major vessels coloured by fluoroscopy
Thal	Defect type
Target	Heart disease present or not present

5.1 Parameters

B. DESCRIPTION OF METHODS AND ALGORITHMS.

a.) Decision Tree Classifier

A decision tree structure comprises root, internal and leaf nodes. A decision tree model depends on various collections of rows from within a dataset. The decision tree classification technique is performed in two phases: tree building and tree pruning. Tree building is done in a top-down manner concerning the target variable. The decision tree model analyses the data based on three nodes, namely

- Root node - this primary node, based on this node, all others perform its function.
- Interior node - this node handles the condition of dependent variables.
- Leaf node - the final result is carried on a leaf node.

b.) Logistic Regression

Logistic regression is a classification technique that uses supervised learning to estimate the likelihood of a target variable. Because the nature of the target or dependent variable is dichotomous, there are only two viable classes. The dependent variable is binary, with data represented as either 1 (for success/yes) or 0 (for failure/no). A logistic regression model predicts P(Y=1) as a function of X mathematically. It is one of the most basic ML techniques that may be used to solve various classification issues such as spam detection, diabetes prediction, cancer diagnosis, and many moreBelow is an example logistic regression equation:
y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x))

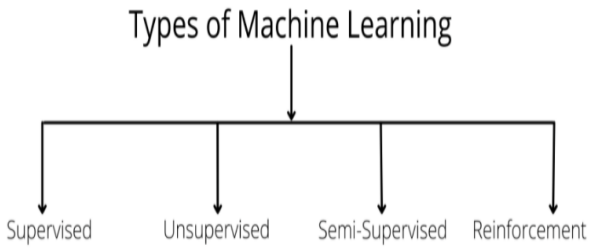
c.) Random Forest

Random forest is a supervised classification machine learning algorithm that uses the ensemble method. A random forest is made up of numerous decision trees and helps to tackle the problem of over-fitting in decision trees. These decision trees are randomly constructed by selecting random features from the given dataset.

Random forest arrives at a decision or prediction based on the maximum number of votes received from the decision trees. The outcome, which is arrived at a maximum number of times through the numerous decision trees, is considered the outcome by the random forest.

d.) Machine Learning

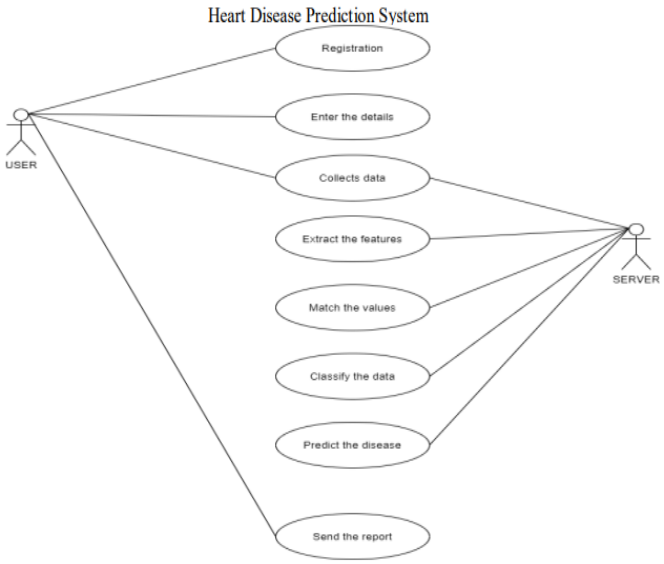
Machine Learning is an efficient technology based on two terms: testing and training. A system takes training directly from data and experience, and based on that training, a test should be applied to various types of needs according to the algorithm. Machine learning algorithms are divided into four categories:



6. IMPLEMENTATION

Heart disease prediction system is a website which is developed using HTML, CSS, JAVASCRIPT for the front end and Django for backend. In this website the user can make its own account. Using registration page the user can view his previous records as it gets store in the data base. The data which is entered in the parameter page where we can predict the heart disease, the data gets compared with the existing data in the data set used. As per that the system predicts the heart disease. The result gets stored in the backend with the user name. Once the result is displayed, user can see some pre entered suggestions on the prediction page, even doctors name and address, number, email is also displayed for further use. For the prediction of heart disease, 3 algorithms were used. After comparing all the three algorithms which are, Logistic regression, Random forest, Decision tree, Logistic regression was giving best accuracy in compared to other twos. Logistic regression gave 92% accuracy.

7. SYSTEM DESIGN



7.1 UML Diagram

The above diagram justifies that the user can access the Registration page, where-in user can enter his details. Further he can enter his data which gets stored in server. The server extracts the features of used algorithm and hence it matches the value with the data which is used to traun the model. After that the server classifies the data which helps in predicting disease. At the end the server displays the report to the user which is calculated by the algorithm used. This report is about whether the user is having heart disease or not.

8. FUTURE SCOPE & FEASIBILITY

A.) Future scope

The machine learning model will eventually employ a bigger training dataset, maybe more than a million individual data points stored in an electronic health record system. Although it would be a tremendous jump in terms of computer power and software sophistication, an artificial intelligence-based system might allow the medical practitioner to choose the optimal therapy for the worried patient as soon as feasible

B.) Feasibility

The project's scope is that integrating clinical decision assistance with computer-based patient records might reduce medical mistakes, improve patient safety, eliminate unnecessary practise variation, and improve patient outcomes. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions. We

have to use algorithms like logistic regression for creating a model that is accurate and less error percentage.

9. RESULT AND DISCUSSION

The aim of this project is to know whether the patient has heart disease or not. So we have created a website where patient can easily predict the heart disease

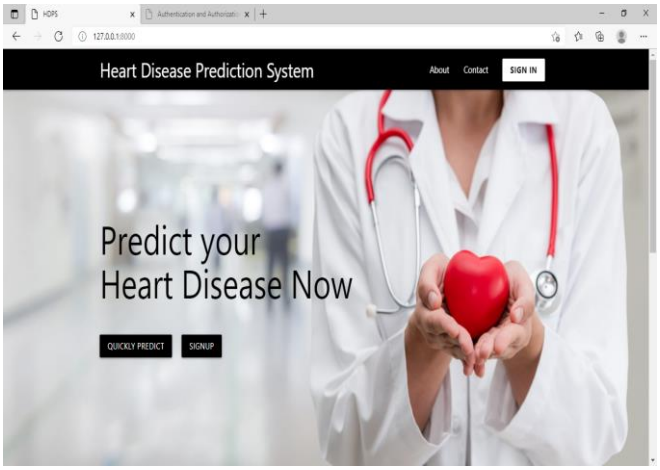


Fig 9.1 Index Page

In Fig 9.1 we can see that this is the main page of website where quickly predict option is available for the patient or they can register and login and then predict so for that sign up and sign in is included on website as well as about and contact section is also available.

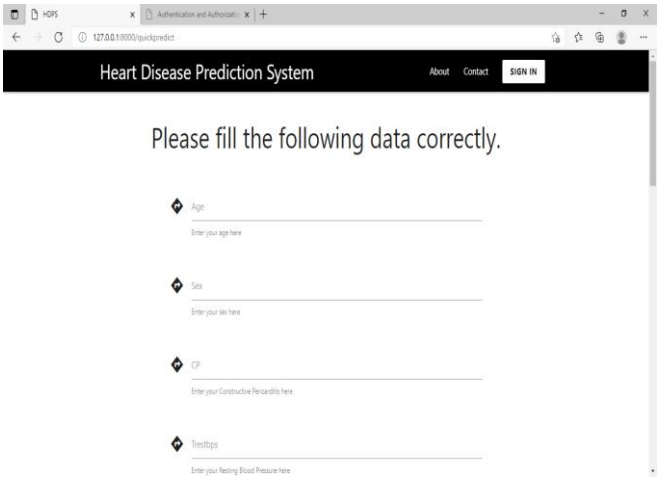


Fig 9.2 Prediction Page

In fig 9.2, the patient can predict the heart disease by filling all the 13 attributes like age, gender, Cholesterol level, etc. The user should enter the following data correctly otherwise it can give wrong prediction.

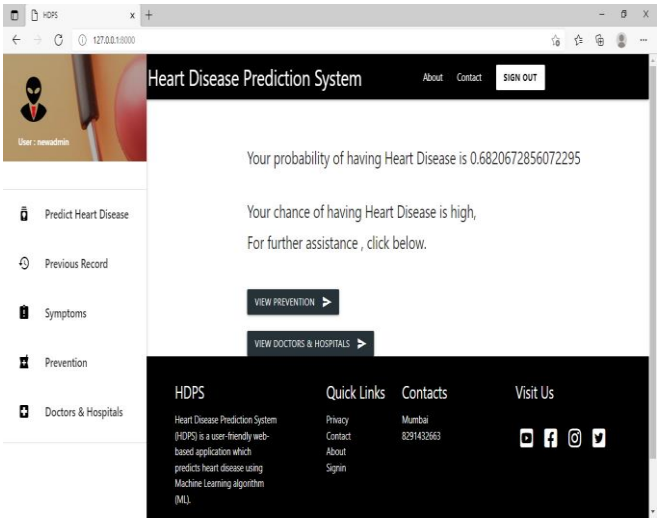


Fig 9.3 Output Page

In Fig 9.3, the prediction given by model is 68 % so the chance of having heart disease is high and if the prediction percentage is less than 60% than there is a low chance of having heart disease. The patient csn view the prevention and can further assist to Doctor or Hospital. The user can get their previous records if they have predict their results earlier using same credentials for Login.

10. ADVANTAGES

1.) Improved accuracy in diagnosing cardiac disease.
2.) Handles the most difficult (enormous) quantity of data utilizing the Logistic regression technique and feature selection.
3.) Reduce physicians' time complexity.
4.) Patient-friendly pricing.

11. CONCLUSION

In this paper, the heart disease dataset of UCI repository is taken and subjected to various classification and clustering algorithms using python. The main focus is to target all possible combinations of the attributes against various algorithms. Then of all the techniques it is the technique that works the best to predict the heart disease at an early stage is identified. The Heart Disease Prediction System helps to predict the disease with the 13 attributes and can be very helpful in the crucial period. We have compared three classification algorithms such as Logistic Regression, Random Forest and Decision Tree in which Logistic Regression has given highest accuracy of 92%. If the probability of the prediction is more than 60% than the risk is high and if the prediction is less than 60 % than the risk is low. Quickly predict is the good feature in case of emergency . The system we are proposing is more advanced and cost effective than the existing ones as we have added a previous records and doctor and hospitals details features on our website.

ACKNOWLEDGEMENT

We owe sincere thanks to our college Atharva College of Engineering for giving us a platform to prepare a project on the topic "Predicting Heart disease using machine learning algorithms" and would like to thank our principal Dr. Shrikant Kallurkar for giving us the opportunities and time to conduct and research on the subject. We are sincerely grateful for Prof. Anuradha Kapoor as our guide and Prof. Deepali Maste, Head of Information Technology Department, and our project coordinator Prof. Renuka Nagpure for providing help during our research, which would have seemed difficult without their motivation, constant support, and valuable suggestion. Moreover, the complication of this research paper would have been impossible without the co-operation, suggestion, and help of our friends and family.

REFERENCES

[1] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction using Hybrid Machine Learning Techniques", IEEE Access 2019.

[2] Sakshi Goel, Abhinav Deep, Shilpa Srivastava, Aparna Tripathi, "Comparative Analysis of Various Techniques for Heart Disease Prediction System", ISCON 2019.

[3] Mohini Chakarverti, Saumya Yadav, Rajiv Rajan, "Classification Techniques for Heart Disease Prediction in Data Mining", IEEE 2019.

[4] Dr.J.Selvakumar, M.Preethi, "A Literature Survey Of Predicting Heart Disease", IRJET 2020.

[5] Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data

mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.

[6]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[7]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150-154). IEEE.

[8]. Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "Analysis of data mining techniques for heart disease prediction." In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-5. IEEE, 2016.

[9].Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[10].Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[11]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer Science & Information Technology Journal (2014): 53-59.

[12]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In 2015 2nd International Conference on Computing for Sustainable Global Development

(INDIACom), pp. 704-706. IEEE, 2015.

[13] G. Purusothaman, and P. Krishnakumari, June 2015,"A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology, Vol. 8(12), DOI:10.17485/ijst/2015/v8i12/58385, pp. 1-5.

[14] Ashish Chhabbi,Lakhan Ahuja,Sahil Ahir, and Y. K. Sharma,19 March 2016,"Heart Disease Prediction Using Data Mining Techniques", International Journal of Research in Advent Technology,E-ISSN:2321-9637,Special Issue National Conference "NCPC-2016", pp. 104-106.

[15] Boshra Bahrami, and Mirsaeid Hosseini Shirvani,February 2015,"Prediction and Diagnosis of Heart Disease by Data Mining Techniques",Journal of Multidisciplinary Engineering Science and Technology(JMEST), ISSN:3159- 0040, Vol. 2, Issue 2, pp. 164-168.

[15] Siddharth Mundra, Kiran Manjrekar, Nimit Lalwani, Nilesh Rathod. Review on prediction of heart disease using data mining, International Journal of Advance Research, Ideas and Innovations in Technology5.6 (2019), www.IJARIIT.com.

[16] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.

[17] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277- 281

[18] Sharan Monica.L, Sathees Kumar.B, "Analysis of CardioVasular Disease Prediction using Data Mining Techniques", International Journal of Modern Computer Science, vol.4, 1 February 2016, pp.55-58.