



HEART DISEASE PREDICTION AND CONSULTATION

Prof. Haripriya. V¹, Aayushi Shah²

Assistant Professor, Department of CS & IT, Jain (Deemed-to-be) University, Bengaluru, India¹

Student, Department of CS & IT, Jain (Deemed-to-be) University, Bengaluru, India²

Abstract: Since there are more and more cases of heart disease every day, it is important and concerning to anticipate any prospective problems. It takes accuracy and quickness to complete this difficult diagnosis. I will create a system that predict whether or not a patient will be diagnosed with heart disease based just on patient's medical history using a number of machine learning techniques, including such regression models, Random Forest, and KNN. The model is controlled in a very beneficial method to boost the accuracy of myocardial infarction prediction in each individual. Random Forest, KNN, and Logistic Regression are used in the suggested model, which has the advantage of predicting heart disease symptoms in a specific person with greater accuracy than previously used classifiers like naive bayes. These techniques will also show better accuracy than other classifiers like naive bayes. So, a great deal of pressure has been reduced by using the presented model to assess the possibility that the classifier will accurately and consistently detect cardiovascular disease. This experiment teaches us a lot that we can utilise to forecast who will develop heart diseases.

Keywords: Random Forest, KNN, Logistic Regression, Machine Learning, Prediction.

I. INTRODUCTION

Heart disease prediction is very important in the field of research in medical and healthcare fields. There have been many studies in the past few decades aimed at developing accurate and reliable methods for predicting heart disease. Some of the common methods which are used for predicting heart disease that include statistical analysis, algorithms of machine learning, and data mining techniques. The most widely used algorithms in literature include logistic regression, decision trees, random forests, k-nearest neighbors, and support vector machines.

Heart / cardiovascular diseases are pretty common now-a-days, they have a particular range of different conditions related to your heart and which can affect your health. World health organization has estimated that 17.9 million global deaths are due to heart /Cardiovascular diseases.

This is the primary cause of maximum deaths in adults. This project is to help and predict among the people who are much more likely to get diagnosed with a heart disease through the help of their medical health history. It's used to recognize what all symptoms are present and also who all are having symptoms related to heart/ heart disease such as angina, chest pain, High BP, Fasting blood sugar. Once all such symptoms are diagnosed or recognized the patient can take immediate medical help, treatments for that particular disease, so that it can be cured at its earliest.

There is no exception in the healthcare industry . Machine Learning plays an vital role while predicting presence the Heart diseases is present or absent in that particular person . If such information are predicted well in advance an immediate and appropriate treatment can be provided to them.

I. HEART DISEASE PREDICTION

Heart disease prediction involves using many different factors like their health / medical history, lifestyle , living habits and test results for predicting the likelihood of developing cardio diseases in an individual. It is a complex process that involves using statistical models, machine learning algorithms, and clinical decision support systems to analyze and interpret data. The goal of heart disease prediction is to recognize individuals those who are at maximum risk of developing the disease so that they can take proactive measures to reduce their risk and improve their overall heart health. To predict heart disease, several factors are taken into consideration, including:

Demographic information: age, gender, race, heart disease-Family history if any.

Lifestyle Habits: Daily food habits, Physical, Mental, Emotional Habits , smoking, Consumption of Alcohol.

Health conditions: Diabetes, high blood sugar, high cholesterol, past heart attacks, or strokes.



Clinical test results: electrocardiogram (ECG), blood tests, imaging tests like echocardiogram or coronary angiogram.

Based on this information, various predictive models and algorithms are used to estimate a individual's risk of developing a Cardiovascular disease in the future. The results of these predictions can help healthcare providers identify individuals who are at high risk and offer them personalized prevention and treatment strategies to reduce their risk.

It's important to note that heart disease prediction is not a perfect science and there is still much to be learned about the underlying causes of heart disease. However, it remains a valuable tool in the fight against this debilitating condition.

Estimating the effectiveness of science/medical therapy are aided for using Applications of Data Mining . Data mining has the ability to give the action plans that can be successful and are accomplished through the comparison of the treatment's course, causes, and symptoms. Real-world Data Mining applications are exciting because they offer data-miners with a unique set of challenges. Managing databases for heart patients is one of these practical problems.

The development of data mining applications can aid for the prediction of effective Medical treatment. Data Mining has the ability for delivering effective courses of action that are achieved through the comparison of the causes, symptoms, and treatment options. Real-world data mining applications are fascinating because data-miners are presented with a diverse set of problems. One such real-world issue is dealing with databases pertaining to heart patients.

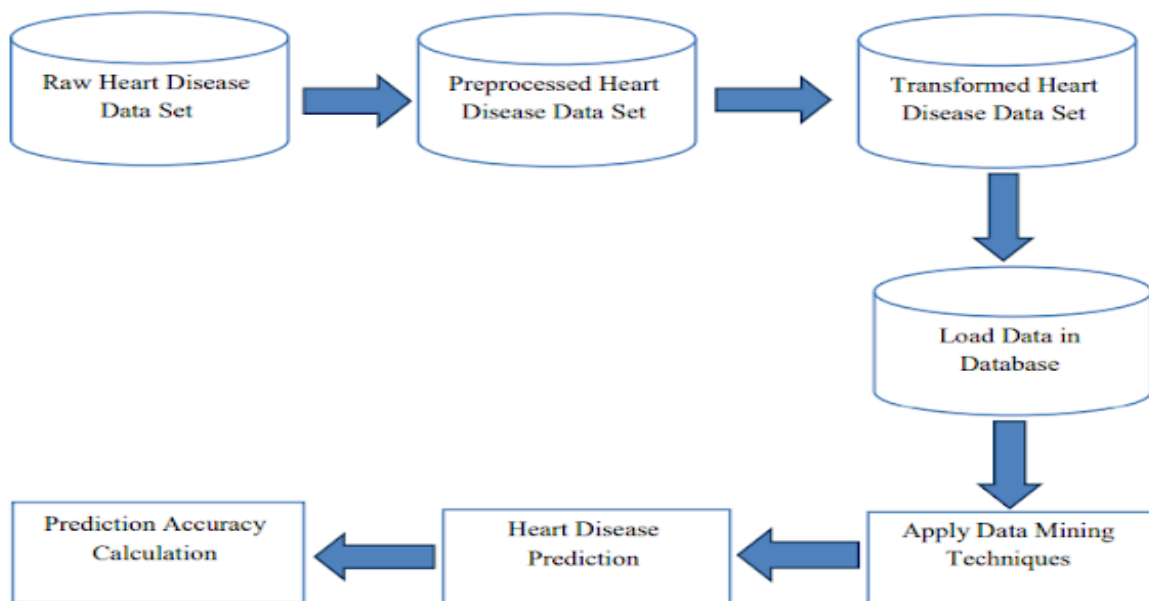


Fig. Prediction Of System

I. LITERATURE REVIEW

Data integration, inconsistent data, missing values, and data inconsistency are presented in “Mohd et al(2013) .'s” discussion of data pre-processing activities including data interpretation. The prepared dataset comprises all the fields needed for the study. Regarding the diagnosis of oral cancer in regard to Demo-graphics, socializing habits, Clinical signs, and lifestyle factors. Following Data, standardization and modification, the research team created an oral cancer dataset with 27 attributes as a contribution to the field. The only two variables which are Age and Id, both are continuous. The remaining variables are category or discrete.

In order to extract relevant patterns, “Intelligent and Effective Heart Attack Prediction (IEHPS) (Patil and Kumaraswamy”, 2009) proposes a methodology. Using heart attack prediction data from the heart disease warehouses. To improve mining efficiency, the data warehouse is first preprocessed. Utilizing the K-means clustering technique, the preprocessed data warehouse is then grouped. MAFA (MAXimal Frequent Itemset Algorithm) is used for frequent item mining (FIM), which is the procedure for extracting organization protocols in the clustered datasets. The patterns necessary for heart attack forecasting are then chosen based on weightage, which is subsequently determined. In order to learn the chosen patterns, a trained neural network. By making use of back-propagation as the training algorithm, Distinct-layer perceptron model is used. Utilizing the importance of weight, significant patterns are extracted greater than the predefined threshold.



“The Multiple Regression Model for Predicting Heart Disease by K. Polaraju et al “. demonstrated that Multiple to estimate the likelihood of developing heart disease, linear regression is suitable. Data-sets used for training, contain 3000 instances/examples of 13 distinct characteristics mentioned earlier, is used in this work. Data Sets are split in half, for like 30% data being utilized to test and 70% being used to train. Then the results show that the “Regression method” has a higher classification accuracy compared to other algorithms.

Using 29 clinical data from patients or individuals, “Chen et al. (2011)”, helped to improve Heart diseases and Prediction of a Heart Disease method to aid medical practitioners in anticipating heart diseases locations. There was increased usage of Prediction in Heart diseases. System would employ the input clinical conditions and data portion, the ROC curve display segment and the prediction performance segment to bring it all together in different qualities. Projected concepts might be used for predicting patients heart conditions with accuracy. This method was a cutting-edge approach for categorizing heart diseases.

“A decision support method was put forth by Kantar et al. (2014)” to automatically identify the normal cycle of sinus or different types of diseases. Algorithms which are improved help to assist physician with Applications in education. Thirteen functions were used in the major code to mechanically identify eight different ECG pathologies. With the current study goal, better achievement was anticipated to predict in the future in the existing techniques.

“Kavitha and Kannan included a better structure for simple cardiac diseases prediction(2016)”. Principal Component Analysis (PCA) was used to develop a better structure in order to mine characteristics and mathematics pattern that helps to choose the related restraint. Proposed work had aided to improve efficacy, precision, and process efficiency. This could be useful in applications such as information retrieval, image processing, and pattern matching.

“Wijaya and Prihatmanto (2013) used machine learning to develop heart disease prediction models”. Tools like smart mirror, Smart Phones, mouse, Smart chair . To Obtain Heart rates data over the internet, which are stored on the server . Learning System was used for a year to collect enough data for predictions. Potential heart diseases prediction has the course of years increased a individuals's awareness of heart diseases. Our Machine Learning algorithm with Web-application is proposed in such a way that it reduces heart disease-related patient deaths.

“KNN Algorithm was used by “Rajathi and Radhamani ” efficiency in the combination of Ant Colony Optimization (ACO) for predicting heart diseases. In two stages the investigation was divided. Acute Rheumatic Fever is caused by Streptococcus Pyogenes bacteria. An algorithm named KNNACO, was implemented for the current approach and its precision and error rate were evaluated.

Medical Diagnosis System was Proposed and Predicted by “Amma (2012)”. Proposed system was built with the help of a genetic algorithm and a neural network. It included 303 cases of heart disease with 14 class label attributes. To train the system, a genetic neural network was used. The final neural network weights were weighted in the weight base and used to predict cardiovascular disease. This concept had a classification precision of 94.17%.

I. PREDICTIVE AND ADAPTIVE ALGORITHM

Machine Learning Algorithms: By examining numerous medical and age, gender, blood pressure, cholesterol levels, and so forth comes under demographic data, to forecast heart diseases machine learning algorithms are used. The program can find patterns and connections between different variables and the existence of heart disease after being trained on a sizable dataset of patient records. Based on the medical data of new patients, the trained model can subsequently be used to estimate their risk of developing heart disease. This strategy has the potential to increase the precision of heart disease diagnosis and enable quicker and more efficient treatment.

A] Random Forest Algorithm –

One algorithm for supervised learning is Random Forest. In order to improve Decision Tree performance, bagging has been included into machine learning classifiers. It combines tree predictors, and trees rely on a randomly selected vector taken from each person. There is a uniform distribution of trees everywhere. For learning using Random Forests, the worst-case temporal complexity is $O(M(n \log n))$, where M is the number of growing trees, n is the number of occurrences, and d is the data dimension

The following are the steps to finish the task:

- Randomly choose samples from a dataset.



- To make predictions by building a Decision Tree for each sample.
- Cast a vote for each potential result.
- Choose the result that received the most votes as the final forecast.

B) LOGISTIC REGRESSION ALGORITHM

The most often used machine learning algorithm in the autonomous learning category is logistic regression. The dependent variable variable is forecasted using a predetermined array of independent variables. When a dependent variable is categorical, the output can be estimated using logistic regression.

A distinct or categorical value thus has to be the result. Instead of giving a precise value in the range of 0 & 1, it gives probabilistic values between 0 and 1. The calculated outcomes include True, False, Yes, No, 0 or 1, etc.

C) KNN ALGORITHM

For correlations between predictors and values, KNN looks through the dataset. Because no precise parameters have been identified for any single functional form, KNN uses a non-parametric technique. It makes no inferences whatsoever about the characteristics or findings of the dataset. KNN is sometimes referred to as a lazy classifier as it memorizes the training data rather than accurately learning and setting the weights. As a result, the majority of the computer effort is completed during the classification phase as opposed to the training phase. Often, KNN just adds a new feature to the class it thinks it is most similar to after making no attempt to identify a class that it is closest to.

The selection of a K value is the first step in the execution of the KNN algorithm. Now let's try to measure the distance. Let's concentrate on the Euclidean distance in this case. Get the Euclidean distance between the k neighbors. The new point we have supplied is now compared to all of its neighbors to determine which is closest to our point. Now, we essentially search for k-nearest. We may now look at which class had the highest score. After choosing the greatest number, we assign that class our new point.

II. APPLICATION OF HEART DISEASE PREDICTION AND CONSULTATION

Early detection and prevention: Before any symptoms appear, heart disease prediction algorithms can assist in identifying people who are very susceptible to acquiring the condition. This can make it possible for early intervention and prevention measures to lower the risk of developing heart disease, such as lifestyle adjustments or medication management.

Customized Treatment Planning: Heart disease prediction algorithms can assist clinicians in creating individualized treatment plans by reviewing a patient's medical history, family history, lifestyle factors, and genetic information. According to the patient's unique needs, this can include suggestions for medication, lifestyle changes, or even surgical procedures.

Remote Monitoring: Patients with heart disease can be monitored remotely and given real-time information by clinicians thanks to wearable technology and remote monitoring systems. health status. This can help identify potential issues early and intervene before they become more severe.

Consultations using telemedicine: Patients can now consult with their healthcare practitioner from the comfort of their own homes thanks to the availability of telemedicine. Telemedicine platforms can incorporate algorithms for heart disease prediction, enabling clinicians to offer individualized advice and treatment regimens over the internet.

Research: To help guide research efforts, patterns and trends in patient data can be found using heart disease prediction algorithms. This could lead to the discovery of new risk factors or therapeutic possibilities, which would ultimately enhance the healthcare for heart disease patients.

By assisting medical professionals in identifying patients who are at risk, offering individualized treatment suggestions, enhancing access to care, and promoting lifestyle modifications, heart disease prediction and consultation can, overall, have a major impact on healthcare.

I. PROBLEM STATEMENT

Predicting the risk that a person will acquire heart disease based on their demographic, lifestyle, and medical history is the aim of the challenge of heart disease prediction. This can aid medical practitioners in identifying high-risk patients and putting preventative measures in place to lessen the chance that heart disease will develop.



Predicting the chance that a person may acquire heart disease in the future involves using a variety of medical and demographic data. In order to find patterns in the data that are indicative of cardiac disease, this challenge is often solved using machine learning techniques. To do this, a model is trained on a dataset of labelled samples. Cardio-vascular disease identification are significant problem. While there are methods for predicting coronary artery disease, they tend to be too inexpensive or inefficient to reliably estimate the probability of heart illness among humans. Early identification of heart disorders can lower mortality and overall consequences. Unfortunately, it's impossible sometimes to correctly look after patients regularly, and a Dr. can't counsel a patient for 24/7 as this could consume more hours, patience, and skill. Today's world has a lot of data, so we may use a variety of machine learning methods for searching for concealed patterns. The discovered knowledge in patient records may be employed to diagnose illnesses.

VII . PROPOSED METHODOLOGY

The gathering of data and the decision-making process for the most crucial attributes make up the first stage of system operation.. The data is re-created by pre-processing the patients data which is available. The dataset is then separated into data for testing and training. By using algorithms and train the model, the model is trained.. System's correctness is assessed by running tests on test data.

Error detection, locating pertinent data, and verifying the correlation between exploratory analysis variables are all accomplished using exploratory data analysis (EDA). The risk factors associated with heart disease are taken into account in this work, as well as the heart attack prediction. Logistic regression, Random Forest, and KNN are the machine learning classifiers used for the research.

1. The first phase is data acquisition, often known as data collection. This involved assessing the physical conditions and taking into account the numerical data by converting the samples that the computer will use to alter.
2. Pre-processing is the next step, when we work to clean up the dataset by addressing problems with the data such missing values, outlier detection, and redundancy reduction. The uniform environment has been subjected to predictive analysis.
3. "Integration" is the third step, when libraries and other subsets are merged by importing independent Python modules and joining them to carry out required tests.
4. The fourth phase is "analysis," which involved using EDA to determine how various data attributes relate to one another.
5. The fifth phase involved "intervention" to enter decision-making procedures, i.e., a search strategy for comprehending earlier experimental investigations to ascertain whether it is practical to use models to solve practical issues.
6. Making predictions using machine learning algorithms is the sixth phase.

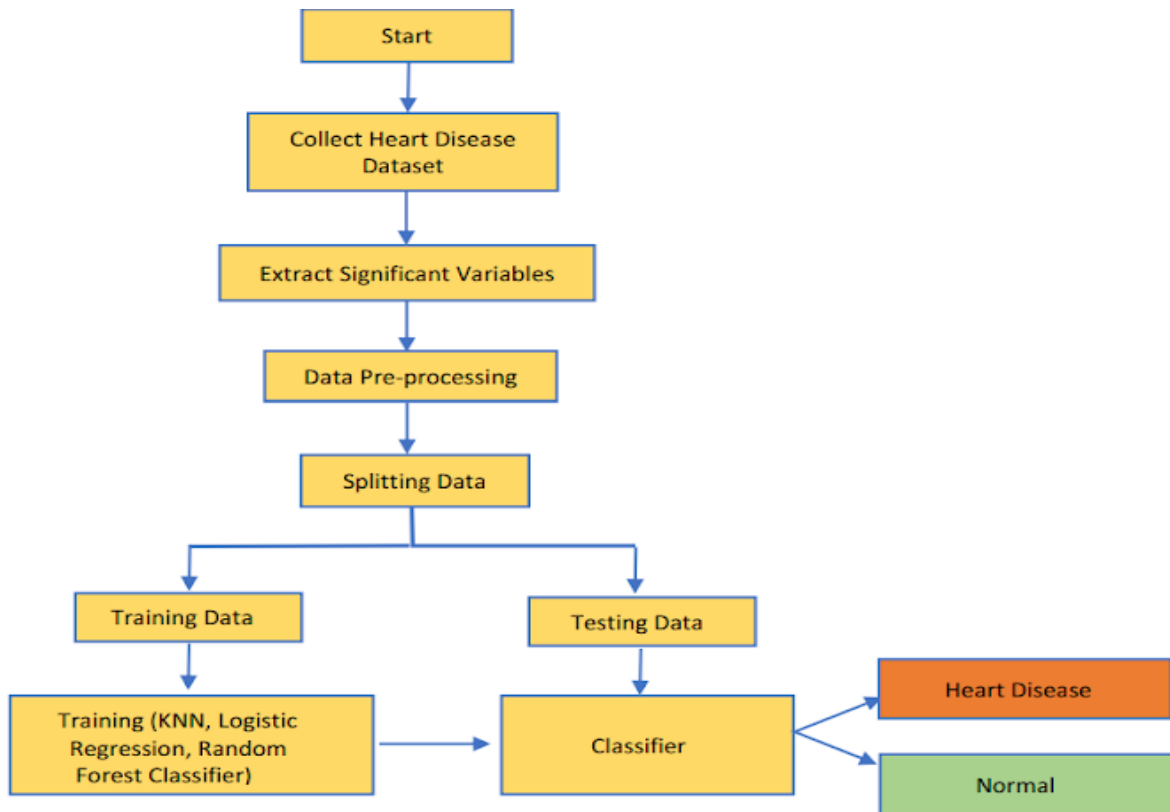


Fig. Proposed Model

VIII. IMPLEMENTATION

- **Data gathering:** The first stage is to gather pertinent data about heart disease patients. This can contain information about their medical background, way of life, test findings, and any other pertinent facts.
- **Preprocessing of data:** The subsequent stage involves preprocessing the gathered data. This entails preparing the data for machine learning algorithms by cleaning it, eliminating any unnecessary or missing data, and transforming it to the proper format.
- **Feature Selection :** Accuracy of the predicted model could be increased by choosing the pertinent factors from the preprocessed data's. This entails determining the key elements, such as Person's age, sex, blood pressure, Cholesterol levels, smoking behaviors, and family history, that affect the risk of heart disease.
- **Model selection:** The correct machine learning algorithm must be used in order to create an accurate prediction model. Examples of regularly used algorithms include logistic regression, decision trees, support vector machines, and neural networks.
- **Training and testing:** The model must first be selected, then it must be trained and tested using the preprocessed data. Although the forecast model's accuracy is evaluated using testing data, the algorithm is taught how to recognise patterns using training data.

- (a) The doctor can access the system by logging in or, if they are new, by signing up.
- (b) The physician has the option to view the patients and approve or deny the requests.
- (c) If the doctor agrees to the request, he or she can read the reports and make comments to them on the dashboard, which the patient can see as well.



The Django web framework, which supports quick development of the programmed, is used to build the web application. The framework facilitates a straightforward and understandable user interface while also assisting in the integration of the various text recognition, extraction, and interpretation modules. Two main modules make up the user interface.

Patient side: The patient side logs in, uploads reports, and views report analysis using the user interface. Patients can enter the values from their reports, and machine learning techniques like Random Forest, KNN, and Logistic Regression algorithms are used to forecast the results accurately, giving the patients access to the results as soon as they are known. The reports are uploaded, analyzed, and stored on the SQLite database in an encrypted fashion. After they have been decrypted, the user can retrieve the solutions and related analytics for the reports from the database and view them.

Doctors/Physicians' side: A patient may request a consultation with a doctor when they register and log in. Patients may be accepted or rejected by the doctor. The doctor can see the reports after approving requests and add comments depending on them.

Admin Side: With his ID and Password, the admin can access the system. Admin can add Doctor: The database administrator can add new doctor information. Admins have the ability to add dataset files to databases. He can also View Doctor: The admin can view a variety of doctors' personal information, View Disease: The administrator can view the database's entry on numerous diseases, View Patient: The administrator can see different patient records that have logged into the system. As well as View the User Feedback: Admin may view user feedback sent by different users.

In general, the application of machine learning algorithms can enhance the accuracy and efficacy of cardiac disease diagnosis and counselling, enhancing patient outcomes and reducing medical costs.

IX. ADVANTAGES

There are several advantages in heart disease prediction and consultation, including:

- **Early detection:** To find trends and forecast the possibility of cardiac disease, AI and ML systems can evaluate enormous volumes of patient data, including medical records, imaging investigations, and genetic data. As a result, cardiac disease may be identified earlier and treated more successfully.
- **Treatment plans that are specifically tailored to each patient's needs** can be created by doctors with the use of AI and ML. Medications, lifestyle modifications, and other therapies that lower the risk of heart disease might all fall under this category.
- **Enhanced accuracy:** AI and ML algorithms are better at swiftly and accurately analyzing complicated data sets than people, which results in more accurate diagnoses and treatment suggestions.
- **Remote monitoring:** AI and ML can make it possible for clinicians to monitor and treat patients with heart disease remotely.
- **Cost-effective:** By simplifying diagnosis and treatment and obviating the need for pointless tests and procedures, AI and ML can lower the cost of healthcare.

X. DISADVANTAGES

- **Bias:** AI and ML algorithms may take into account biases from the data they are trained on, which might result in predictions that are unreliable or unfair. For patients from other racial or ethnic groups, the algorithm might not be as accurate if it was developed on data from primarily white patients.
- **Lack of interpretability:** Because ML and AI algorithms can be sophisticated and tricky to grasp, it can be difficult for medical professionals to comprehend how a specific prediction was generated. This can make it challenging to convey to patients the logic behind a diagnosis or therapy prescription.
- **Data caliber:** In order to provide reliable predictions, AI and ML systems depend on high-caliber data. The algorithm's predictions might be off if the data is unreliable or wrong.
- **Regulatory issues:** AI and ML in healthcare are subject to regulatory scrutiny, and there may be issues with law and ethics if they are used to anticipate and provide advice on heart disease.



- Opposition to change: Using AI and ML in healthcare may be met with resistance from physicians and patients, especially when it comes to delicate topics like the diagnosis and treatment of heart disease.
- Thus, although AI and ML have the potential to enhance cardiac disease prediction and consultation, these technologies must be employed cautiously to ensure they are accurate, fair, and advantageous for patients.

XI. RESULTS AND DISCUSSIONS

The accuracy of the diagnostic tests, the severity and stage of the disease, the patient's general health and medical history, and the efficacy of the treatment plan are just a few of the variables that will affect the expected outcome in heart disease prediction and consultation.

The goal of cardiac disease prediction and consultation is to give patients a better understanding of their situation and a clear strategy for controlling it. Healthcare professionals can help patients feel more in control of their health and lessen the burden of heart disease on their life by identifying risk factors, establishing a diagnosis, recommending treatments or preventive measures, and offering emotional support. A healthcare professional could occasionally advise a patient to get additional testing or treatment, including a cardiac catheterization or heart surgery.

XII. CONCLUSION

Working on this intriguing and hard project has been a genuine pleasure for me. This project was beneficial to me since it gave me actual experience with other web-based Sqli and Python programming languages. The most recent client server technology, which will be in high demand in the future, is also covered. This technology is utilised to create web-enabled apps. Thus, future independent project development will have greater prospects and direction.

REFERENCES

- [1] <https://shodhganga.inflibnet.ac.in:8443/jspui/handle/10603/252675#>
- [2] <https://shodhganga.inflibnet.ac.in/handle/10603/341145>
- [3] <https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012072/>
- [4] https://www.researchgate.net/publication/331589020_Heart_Disease_Prediction_System
- [5] <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>
- [6] Peyman Mohammadi, Abdolreza Hatamlou & Mohammad Masdari 2013, 'A Comparative Study on Remote Tracking of Parkinson's Disease Progression Using Data Mining Methods', International Journal in Foundations of Computer Science & Technology, pp. 71-83
- [7] Shantakumar, B Patil, & Kumaraswamy, YS 2009, 'Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network', European Journal of Scientific Research, pp. 642-656
- [8] Chen, AH, Huang, SY, Hong, PS, Cheng, CH & Lin, EJ 2011, 'HDPS: Heart disease prediction system. In Computing in Cardiology', IEEE, pp. 557-560.
- [9] Kantar, T, Koseoglu, O & Erdamar, A 2014, 'Analysis of heart diseases from ECG signal', In Biomedical Engineering Meeting (BIYOMUT), 2014 18th National IEEE, pp. 1-4.
- [10] Kavitha, R & Kannan, E 2016, 'An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining', In Emerging Trends in Engineering, Technology and Science (ICEETTS), International Conference on IEEE, pp.1-5.
- [11] Wijaya, R & Prihatmanto, AS 2013, 'Preliminary design of estimation heart disease by using machine learning ANN within one year', In Rural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T), 2013 Joint International Conference on IEEE, pp. 1-4.
- [12] Rajathi, S & Radhamani, G 2016, 'Prediction and analysis of Rheumatic heart disease using kNN classification with ACO', In Data Mining and Advanced Computing (SAPIENCE), International Conference on IEEE, pp. 68-73.
- [13] Amma, NB 2012, 'Cardiovascular disease prediction system using genetic algorithm and neural network', In Computing, Communication and Applications (ICCCA), 2012 International Conference on IEEE, pp. 1-5.