Heart Disease Prediction Using AIML

A PROJECT REPORT

Submitted by

Ashi Sharma (20BCS7931)

Reeta Devi (20BCS2109)

Aditya Sharma (20BCS9250)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Chandigarh University

November -2022

BONAFIDE CERTIFICATE

Certified that this project report **Heart Disease Prediction Using AIML** is the bonafide work of **Ashi Sharma, Reeta Devi and Aditya Sharma with UIDs 20BCS7931, 20BCS2109,** and **20BCS9250** respectively who carried out the project under my/our supervision.

SIGNATURE	**SIGNATURE**
Dr. Sandeep Singh Kang	**Er. Tanu Dhiman**
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Computer Science and Engineering Department	**Computer Science and Engineering Department**

Submitted for the project viva-voce examination held on 15th November,2022.

INTERNAL EXAMINER	**EXTERNAL EXAMINER**

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Heart disease, alternatively known as cardiovascular disease, encases various conditions that impact the heart and is the primary basis of death worldwide over the span of the past few decades. It associates many risk factors in heart disease and a need of the time to get accurate, reliable, and sensible approaches to make an early diagnosis to achieve prompt management of the disease. Data mining is a commonly used technique for processing enormous data in the healthcare domain. Researchers apply several data mining and machine learning techniques to analyse huge complex medical data, helping healthcare professionals to predict heart disease. This research paper presents various attributes relatedto heart disease, and the model on basis of supervised learning algorithms as Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithm. It uses the existing dataset from the Cleveland database of UCI repository of heart disease patients.

The paper focuses on the construction of an artificial intelligence-based heart disease detection system using machine learning algorithms. We show how machine learning can help predict whether a person will develop heart disease. In this paper, a python-based application is developed for healthcare research as it is more reliable and helps track and establish different types of health monitoring applications. We present data processing that entails working with categorical variables and conversion of categorical columns. We describe the main phases of application developments: collecting databases, performing logistic regression, and evaluating the dataset's attributes. A random forest classifier algorithm is developed to identify heart diseases with higher accuracy. We then discuss the random forest classifier algorithm, including the experiments and the results, which provide better accuracies for research diagnoses.

Keywords- *research, monitoring applications, data, higher accuracy, scable, dynamic applications*

GRAPHICAL ABSTRACT

Figure A: Graphical Abstract

CHAPTER 1

INTRODUCTION

1.1 Client Identification/Need Identification:

Coronary disease is perhaps the leading cause of death on the world today. A fundamental test in the field of clinical knowledge examination is the expectation of coronary infection. AI (ML) has been demonstratedto be viable in assisting with the choices and demands produced by the medical services sector's massive amount of data. We have saw ML procedures being used in real-time in various areas of the Internet of Things (IoT). Various studies only include a cursory insight at the use of ML methods to predict coronary artery disease. In this research, propose a novel approach to determining vital outcomes using AI approaches in order to improve the precision of cardiovascular infection predictions. For the expectation model, there are many combinations of highlights and a few well-known grouping strategies. We achieve an improved execution level with an accuracy level of 89 percent using the logistic regression machine learning algorithm to predict coronary disease.

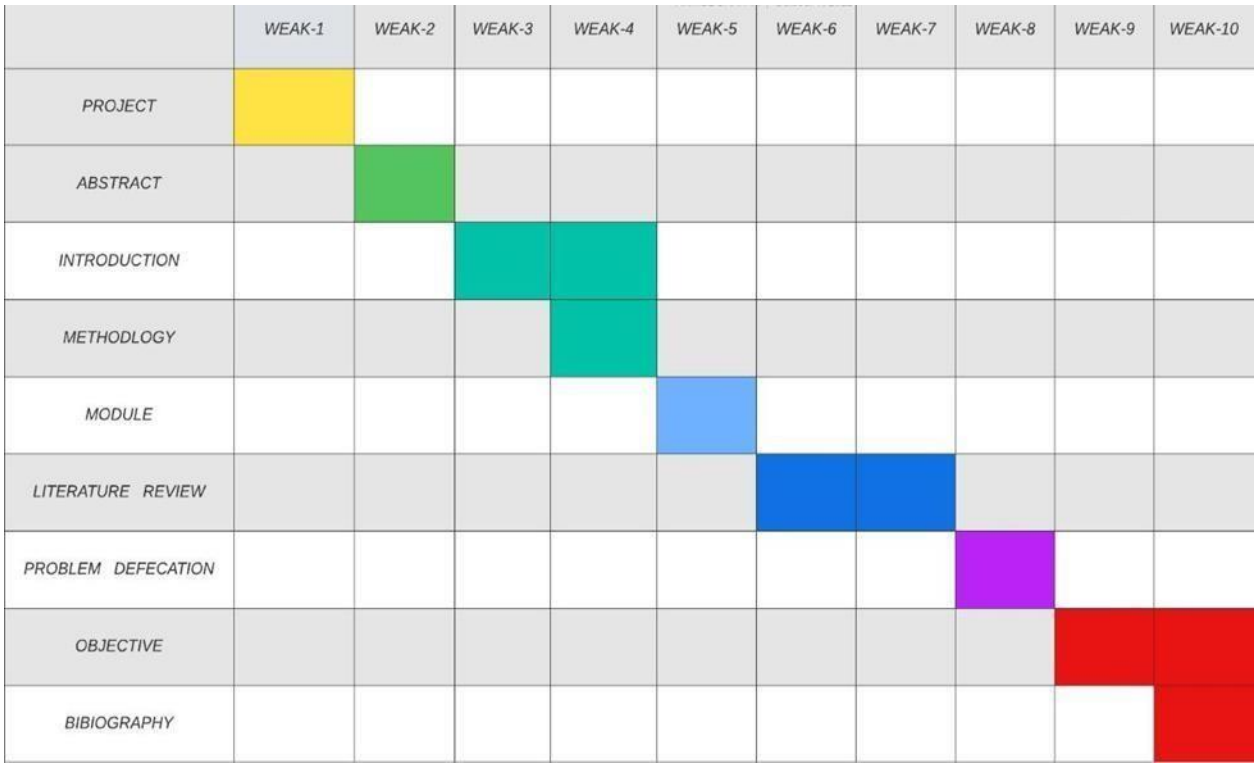
1.2 Identification of Problem

There are many challenges involved in creating a heart disease prediction from scratch. There are currently many disease prediction based on user information, so what should we do if the website does not have enough users? Then we'll work on the accuracy and interface of the system, which is how a system understands the disease. Currently, the health care sector is generating information from several facilities and patients. By applying the best usage of this data, doctors can easily anticipate superior methods for treatment and enhance the complete delivery system of the health care sectors . One of the most important uses is that the python framework can help make sense and encourage computational facilities in extracting valuable insights from the information over the health care sectors.

1.3 Identification of Tasks

To begin the project, we have gathered user requirement for this project and prepare the scopeand objective. The results from this phase are scope and limitation, objectives, cost and benefits, feature of the proposed system and user interface design. In this project, we aimed to develop an Heart Disease Prediction Project which will focus mainly on helping Consumers detect/Predict a Heart Disease to avoidfatalities due Un-Predicted/Lately Diagnosed Heart Diseases

1.4. Timeline



1.5. Organization of the Report

Introduction:
In the current section of the report, we discussed about the Identification of clients and their need, then we discussed about the identification of the problem and various questions related to the problem

Literature Survey:
In this section of the report, will we discuss about the existing works in the field of recommendation systems, their pros and cons and then about the Goals/Objective of Our Project.

Preliminary design/Design Flow:
In this Section of the report, we will be discussing about the Design flow, Design Selection and Implementation Strategy of the project.

Result and Validation:
In this last section of the report, we will be discussing about the outcomes of the project, implementation of the solution and different tools used in the making of this application.

Conclusion and Future Scope:
In this part of the report, we will discuss about the final outcome, conclusion and the advancements that can be done in future

CHAPTER 2
LITERATURE REVIEW

2.1 Timeline of the reported problem

Predictions systems were introduced recently to help people predict most diseases for them from the plethora of options and symptoms related to it. Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases. The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure - a condition that causes around half of all deaths from stroke and heart disease. Heart disease, also known as cardiovascular disease (CVD), encloses a number of conditions that influence the heart – not just heart attacks. Heart disease was the major cause of casualties in the different countries including India. Heart disease kills one person every 34 seconds in the United States. Diagnosis is complicated and important task that needs to be executed accurately and efficiently. The diagnosis is often made, based on doctor's experience & knowledge. This leads to unwanted results & excessive medical costs of treatments provided to patients. Therefore, an automatic medical diagnosis system would be exceedingly beneficial.

2.2 Proposed solutions and Bibliometric Analysis

In this system we are implementing effective heart attack prediction system using Naïve Bayes algorithm. We can give the input as in CSV file or manual entry to the system. After taking input the algorithms apply on that input that is Naïve Bayes. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

2.3 REVIEW SUMMARY

Clinical decisions are often made based on doctor's insight and experience rather than on the knowledge rich data hidden in the dataset. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. 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. There are voluminous records in medical data domain and because of this, it has become necessary to use data mining techniques to help in decision support and prediction in the field of healthcare. Therefore, medical data mining contributes to business intelligence which is useful for diagnosing of disease.

2.4 Problem Definition

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients

2.5 Goals/Objectives

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases. The research aims are to detect heart disease using the python programming language.

The objectives of the study are as follows:

- To critically analyze the ways python language is used to detect heart disease.
- To critically investigate the previous activities and apply a suitable methodological approach for superscribing the identified problem.
- To critically apply data interpretation strategies in python language for health problem detection.

CHAPTER 3.
DESIGN FLOW/PROCESS

3.1 Evaluation & Selection of Features

The diagnosis of heart disease is a challenging task, which can offer automated prediction about the heart condition of patient so that further treatment can be made effective. The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. Selection of features is based on the requirements and the constraints of the previously available applications and the features are as follows:

1. Show preview of the dataset
2. Ability to recognize symptoms
3. Subscription based on PayPal integration.
4. Disease classified into different groups

3.2 Design Constraints

There are no such design constraints like Regulations, Environmental, Health, etc. as this application is just a software application that doesn't require any sort regulations andis not going to affect the environment in any way. But it had a few constraints which include economic constraints and data collection. Economic constraints hit us ina way that we are not able to design and optimize the application in the best possible way. Talking about the data collection it should bear characteristics of a song in its metadata and secondly it should have considerable number of datasets available. RNN models suffer from a problem known as vanishing gradient problem which basically means that they might not be able to connect information from several previous steps to the present step.

3.3 Analysis and Feature finalization subject to constraints

After going through the evaluation and selection of features and the constraints to that, we came up with the final decisions of what we aregoing to add as the features, so the final features are as follows:

i. Show preview of the dataset

ii. Ability to recognize symptoms

iii. Subscription based on PayPal integration.

iv. Disease classified into different groups

3.4 Design Flow

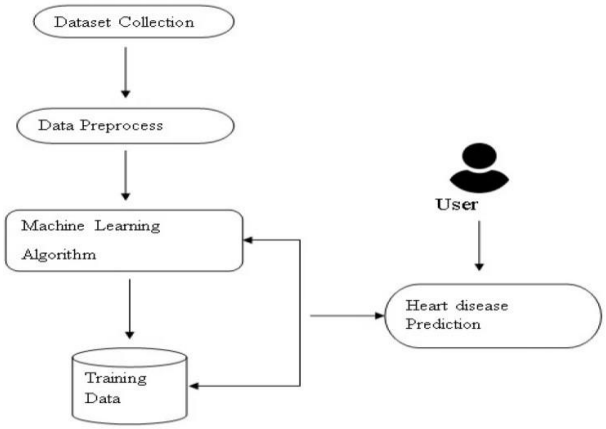
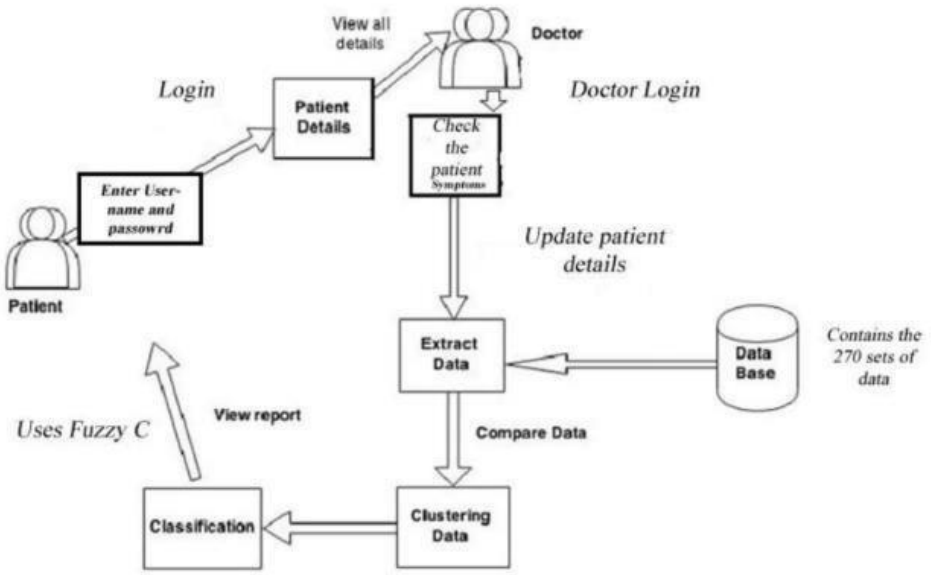


Fig 3.1 HYBRID METHOD

Architecture Of The System

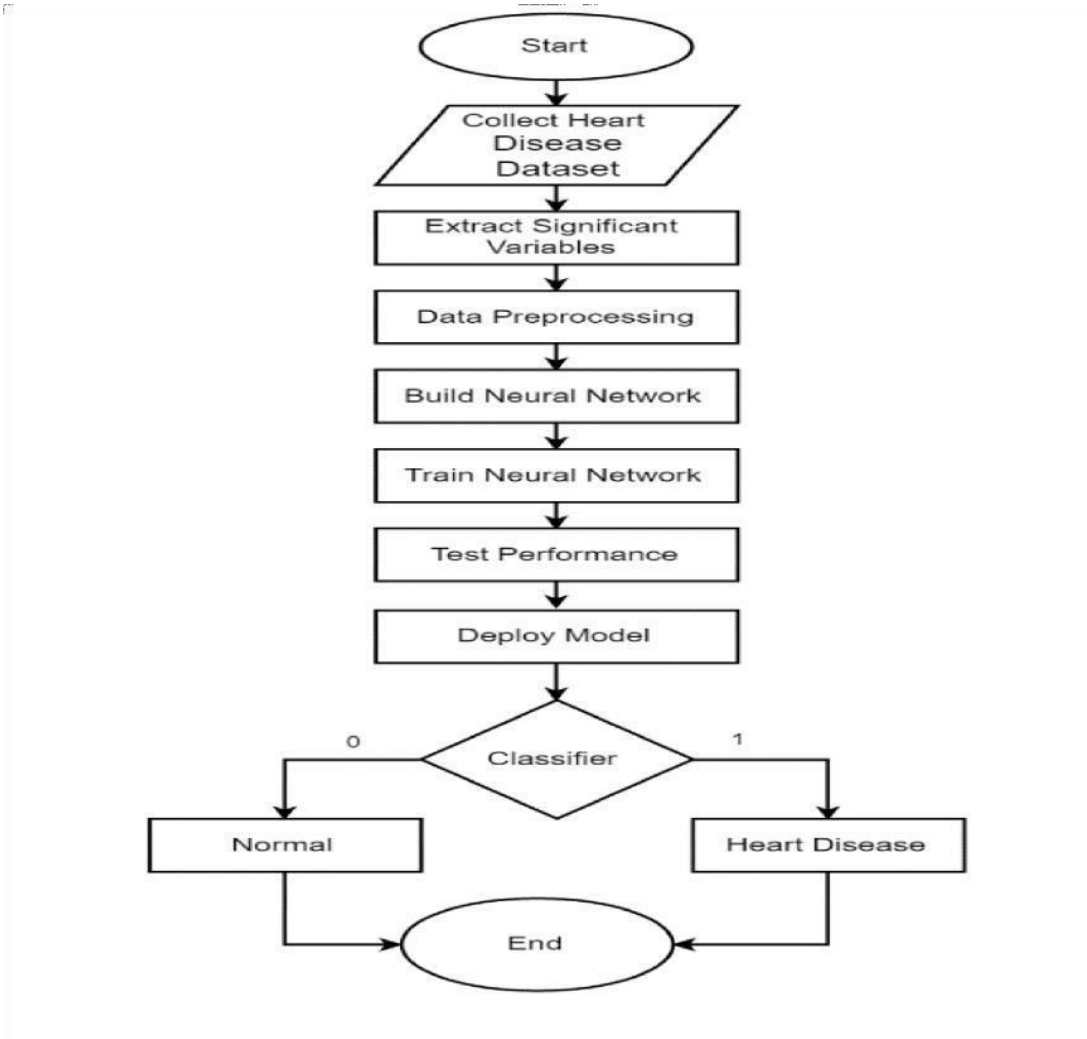


3.5 Design selection

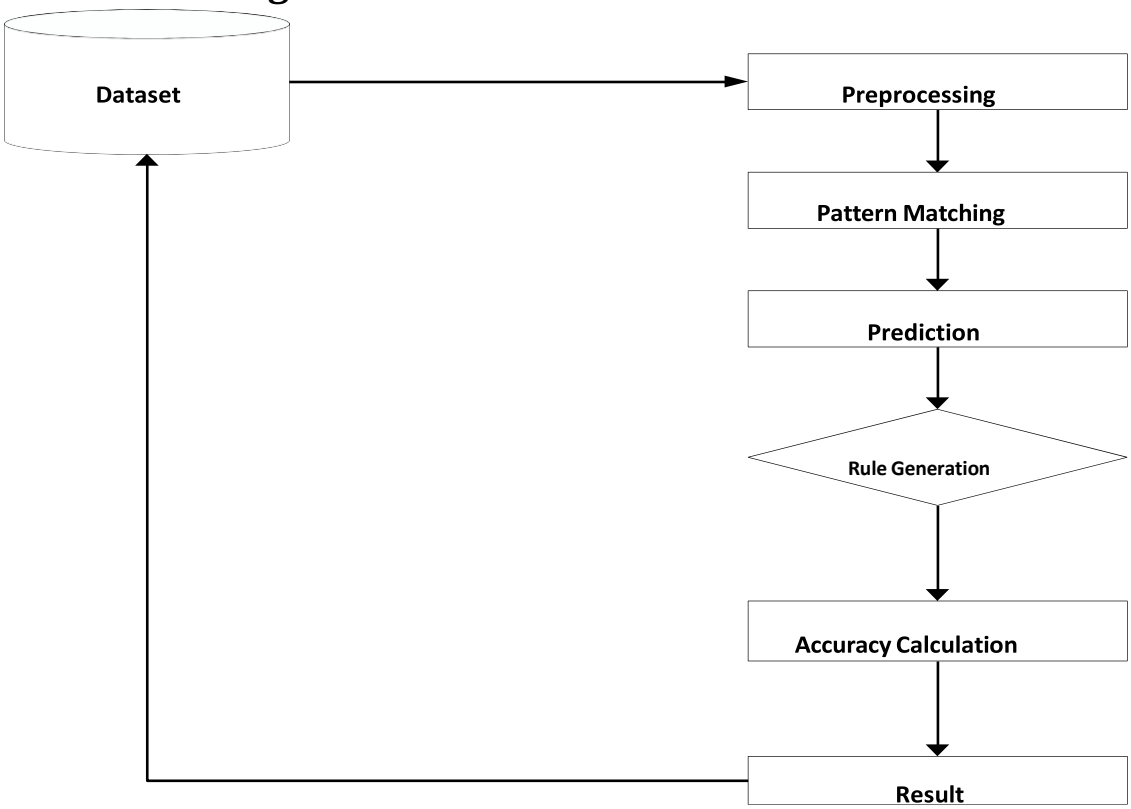
The above architectural diagram and the hybrid model's diagram basically explains the overall flow of how a prediction system works and based on both of these we have designed our system, where both the content-based filtering and the collaborative filtering will be used and hence the quality of prediction will be increased, and asmachine learning is used in the system, it automatically increases the chances of best possible outcome.

3.6 Implementation plan/methodology

DESIGN FLOW DIAGRAM



Data Flow Diagram



CHAPTER 4.

RESULTS ANALYSIS AND VALIDATION

4.1 DESIGN DRAWINGS

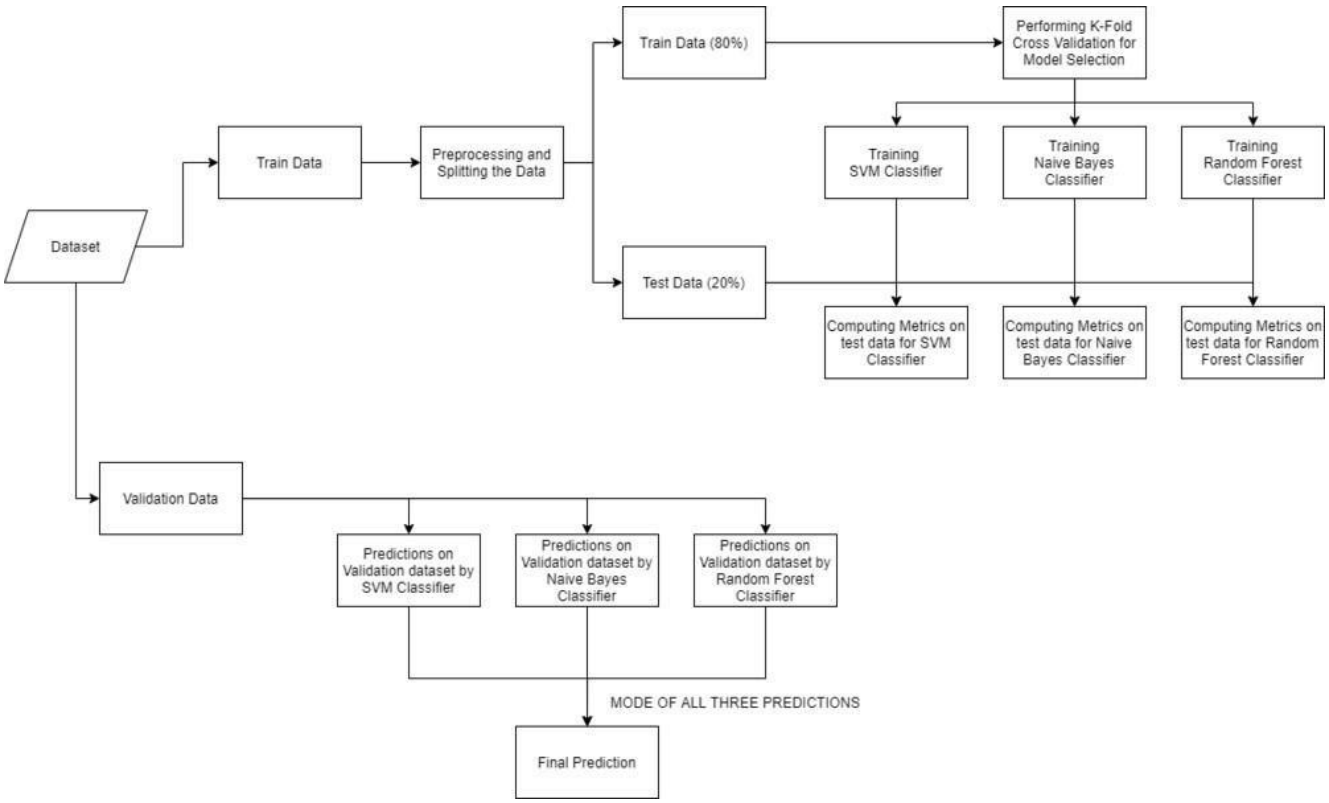


FIG 4.1 Data Flow Diagram

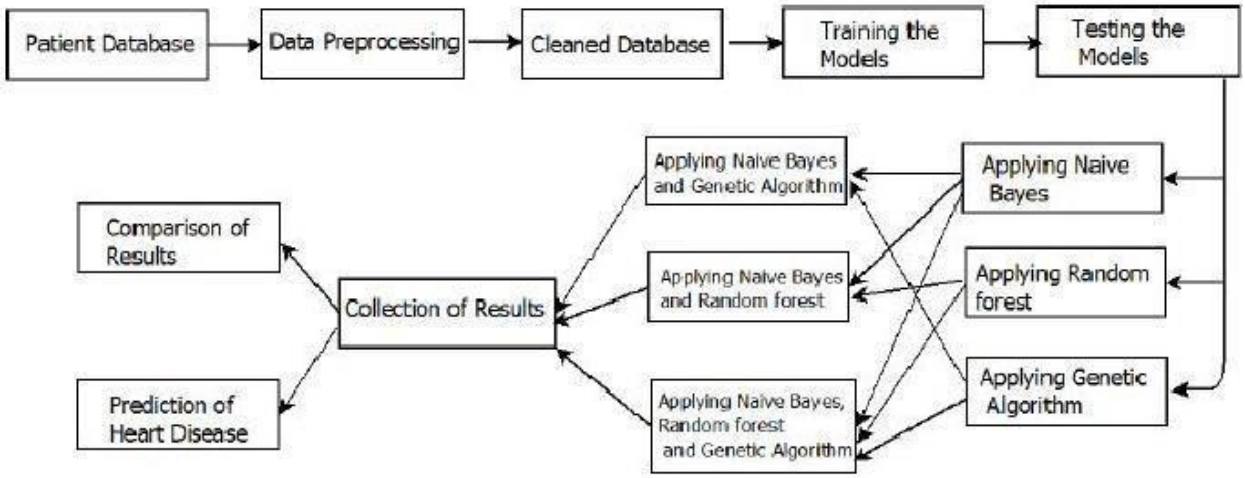


FIG 4.2 ER DIAGRAM

4.2 RESULTS

In [29]:
```
input_data =(41,0,1,130,204,0,0,172,0,1.4,2,0,2)

#changing input data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

#reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = regressor.predict(input_data_reshaped)
print(prediction)

if(prediction[0]==0):
  print('The Person is Healthy ')
else :
  print('The person is not Healthy ')
```

```
[1]
The person is not Healthy

C:\Users\91788\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have vali
d feature names, but LogisticRegression was fitted with feature names
  warnings.warn(
```

In [28]:
```
input_data =(57,0,0,140,241,0,1,123,1,0.2,1,0,3)

#changing input data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

#reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = regressor.predict(input_data_reshaped)
print(prediction)

if(prediction[0]==0):
  print('The Person is Healthy ')
else :
  print('The person is not Healthy ')
```

```
[0]
The Person is Healthy

C:\Users\91788\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have vali
d feature names, but LogisticRegression was fitted with feature names
  warnings.warn(
```

CHAPTER-5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. This project resolved the feature selection i.e. backward elimination and RFECV behind the models and successfully predict the heart disease, with 85% accuracy. The model used was Logistic Regression. Further for its enhancement, we can train on models and predict the types of cardiovascular diseases providing recommendations to the users, and also use more enhanced models

5.2 FUTURE WORK

The proposed system is GUI-based, user-friendly, scalable, reliable and an expandable system. The proposed working model can also help in reducing treatment costs by providing Initial diagnostics in time. The model can also serve the purpose of training tool for medical students and will be a soft diagnostic tool available for physician and cardiologist. General physicians can utilize this tool for initial diagnosis of cardio-patients. There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosiscan be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research. In DM warehouse, generally, the dimensionality of the heart database is high, so identification and selection of significant attributes for better diagnosis of heart disease are very challenging tasks for future research.

REFERENCES

- Mathur P.**Overview of machine learning in healthcare** Machine Learning Applications using Python, A Press, Berkeley, CA (2019), pp. 1-1 CrossRefGoogle Scholar
- Guleria P., Sood M.**Intelligent learning analytics in healthcare sector using machine learning**
- Machine Learning with Health Care Perspective, Springer,Cham (2020), pp. 39-55
- CrossRefGoogle Scholar [4] Kumar V.V. **Healthcare Analytics Made Simple: Techniques in Healthcare Computing using Machine Learning and Python** Packt Publishing Ltd. (2018)
- BelItSoft **Pythonin healthcare** BelItSoft(2017)Available at: https://belitsoft.com/custom-applicationdevelopment-services/healthcare-software-development/python-healthcare,
- https://link.springer.com/chapter/10.1007/978-1-4842-3787-8_1
- https://aip.scitation.org/doi/abs/10.1063/5.0081782
- Mayfield H., Smith C., Gallagher M., Hockings M. **Use of freely available datasets andmachine learning methods in predicting deforestation** Environ. Model. Softw., 87 (2017), pp. 17-28 ArticleDownload PDFView Record in ScopusGoogle Scholar[12]Barot P.
- BoTreeTechnologies. (2020)Available at: https://www.botreetechnologies.com/blog/python- inhealthcare-application/, Google Scholar [13] Bhanot K.
- **Predicting presence of heart diseases using machine learning** Towards Data Sci. (2019) Available at: https://towardsdatascience.com/predicting-presence-of-heartdiseases- using-machine-learning-36f00f3edb2c, [Accessed on 5th March, 2021] Google Scholar [14] Zaidi J.