Mini Project Report on

Heart Disease Predictor

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Engineering in Computer Engineering

by

KARISHMA DATT (13)
VINOD CHOUDHARI (14)
PRIYA PATHAK (15)
MANSI PATIL (16)

Supervisor:

PROF. SUNITA NAIK

Computer Engineering Department

VIVA Institute of Technology

University of Mumbai

2021-2022

CERTIFICATE

This is to certify that the project entitled **“HEART DISEASE DETECTOR”** is a bonafide work of **“Karishma Datt (13), Vinod Choudhari (14), Priya Pathak (15) & Mansi Patil (16)”** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering in Computer Engineering”**

(Name and sign)
Supervisor/Guide

Prof. Ashwini Save
Head of Department

Dr. Arun Kumar
Principal

Mini Project Approval

This project report entitled **Heart Disease Predictor** by **Karishma Datt, Vinod Choudhari, Priya Pathak & Mansi Patil** is approved for the degree of **Bachelor of Engineering in Computer Engineering.**

Examiners

1. ----------

2. ----------

Date:

Place

Abstract

The HealthCare field has a vast amount of data and for processing these data we need certain techniques. Machine learning is one of the techniques often used. Heart disease is the leading cause of death worldwide. The cases of heart disease are increasing rapidly due to unhealthy lifestyle so it has become very important to predict heart disease at an early stage. The diagnosis of heart disease is a difficult task i.e. it should be performed precisely and accurately. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. The proposed system predicts the arising possibilities of Heart Disease. We have used different machine learning models like Logistics Regression, KNN, Decision Tree and XGBoost to predict if the patient is diagnosed with a heart disease or not. The dataset consists of 14 main attributes used for performing the analysis. The correct prediction of heart disease can prevent life threats, and incorrect prediction can prove to be fatal at the same time

Table of Contents

List of Figures

List of Tables

Declaration

We declare that this written submission represents our ideas in my own words and where others' ideas or words have been included. We have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. We understand that any violation of the above will result in disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Karishma Datt

Vinod Choudhari

Priya Pathak

Mansi Patil

Date:

Chapter 1

Introduction

Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine is learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity data produced by the healthcare industry. We have also seen ML new techniques being used in recent developments in different areas and Various studies give only a glimpse into predicting heart disease with ML techniques. According to the World Health Organization more than 10 million die due to Heart diseases every single year around the world. A healthy lifestyle and earliest detection are only ways to prevent heart related diseases. Detection of Heart Disease (HD) by using models of machine learning (ML) is very effective in early stages. The HD treatment and recovery is effective if detected at initial stages. HD identification by machine learning (ML) techniques has been developed to assist the physicians. Kaggle heart disease dataset was for evaluation of the system. The dataset 70% used for training and remaining for validation. The proposed system performances have been measured by using evaluation metrics.

Chapter 2

Literature Survey

2.1 Survey Existing System

Cardiovascular diseases (which often leads to heart failures) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of global deaths. Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity, and harmful use of alcohol using population-wide strategies[1]. However, people with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidemia, or already established disease) need early detection and management wherein a machine learning model can be of great help. This machine learning model could help in estimating the probability of deaths caused by heart failure by to

important features from the dataset and making predictions based on these features. The dataset consists of 12 variables/features, and 1 output variable/target variable. Risk because of heart disease is increasing throughout the world. According to the World Health Organization report, the number of deaths because of heart disease is drastically increasing as compared to other diseases. Multiple factors are responsible for causing heart related issues. Many approaches were suggested for prediction of heart disease, but none of them were satisfactory in clinical terms. Heart disease therapies and operations available are so costly, and following treatment, heart disease is also costly. This chapter provides a comprehensive survey of existing machine learning algorithms and presents comparison in terms of accuracy, and the authors have found that the random forest classifier is the most accurate model; hence, they are using random forest for further processes. Deployment of a machine learning model using web applications was done with the help of flask, HTML, GitHub, and Heroku servers. Webpages take input attributes from the users and give the output regarding the patient's heart condition with accuracy of having coronary heart disease in the next 10 years.

Cardiac disease is one of the most serious problems relating to human life. E- treatment of heart problems has recently been said in a study that has received huge attention in the medical system all over the world In India and other developed countries, about half of all deaths are caused by heart disease[2]. There are many projects and case study papers on heart disease diagnosis using data mining and machine learning methods and RF, SVM, NB, NN, and KNN. model to classify the cardiac disease into normal or abnormal and then results show that RF accomplished the optimal performance. utilized stochastic gradient descent (SGD), KNN, RF, logistic regression (LR), and voting ensemble learning to predict cardiac diseases. Their voting ensemble learning model has achieved the best accuracy of 90%. used a recurrent neural network (RNN), a genetic algorithm, and K-mean etc used for predicting heart diseases. The heart disease dataset is utilized for training and evaluating models. It consists of 1025 records, 13 features, and one target column. -e target column includes two classes: 1 indicates heart disease, and 0 indicates nonheart disease.

Heart Disease is one of the major death factors among the various factors. Detection of cardiovascular disease tends to be a little complex because of insufficient knowledge and experience of the medical practitioners concerning warning signs of heart failure[3]. There is a large amount of data in the healthcare sector. With the help of these data and adopting appropriate data mining techniques, Early detection of heart-related diseases can be achieved and also preventing it from occurring. Both the Data Mining (DM) and Machine Learning (ML) techniques prove to be significant and effective in the field of the medical industry.

Table 2.1: Analysis Table

Title	**Summary**	**Advantages**	**Disadvantages**	**Open Challenge**
Early Detection of Heart Syndrome Using Machine Learning Technique[1]	This paper provides a comprehensive survey of existing machine learning algorithms and presents comparison in terms of accuracy, and authors have got the most accuracy in random forest regression	Accuracy is pretty good	The proposed system do not have a GUI	The proposed system should have a proper webApp
Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method[2]	E-Treatment of heart problems has recently been said in a study that has received huge attention in the medical system all over the world	Increased accuracy for effective heart disease diagnosis.	Prediction of cardiovascular disease results is not accurate	Several standard performance metrics such as accuracy, precision and error in classification should be considered.
Survey of Heart Disease Prediction and Identification using Machine Learning Approaches[3]	Comparative performance in datasets is adopted in this paper for predicting heart disease techniques in comparison to the rest of other ML approaches	It reduces the time to retrieve output	The proposed system is not cost-effective	The proposed system can be more user-friendly

2.2 Limitation of Existing System

Application of today's Machine Learning techniques in heart disease prediction have challenges in data size and confidentiality matters. Processing times for statistical models that are core to modern-day machine learning classification algorithms have been greatly reduced thereby allowing for deeper analysis and enhanced validation techniques. We produce an enhanced performance level with an accuracy level of 90% above through the prediction model for heart disease with the random forest(RF). The dataset was 70% used for training and the remaining for validation. The proposed system performances have been measured by using evaluation metrics. The performances of the classifiers are tested on selected feature spaces, selected through various feature selection algorithms mentioned above.

2.3 Problem Statement and Objective

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate the chance of heart disease in humans. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients everyday in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more patience, time and expertise. Since, user have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

2.4 Scope

Today's, world most of the data is computerized, the data is distributed and it is not utilized properly. By Analyzing the available data we can also use unknown patterns. The primary motive of this research is the prediction of heart diseases with a high rate of accuracy. For predicting heart disease we can use logistic regression algorithms, naive bayes, and machine learning. The future scope of the paper is the prediction of heart diseases by using advanced techniques and algorithms in less time complexity.

Chapter 3

Description and Implementation

3.1 Framework and Algorithm

The dataset used for the proposed system was from the kaggle. The dataset contains 303 instances of the patients and 14 attributes like sugar level, cholesterol level, BP level, etc. Now the attributes which are used in this research purpose are described as follows and for what they are used or resemble:

1. age - age in years
2. sex - (1 = male, 0 = female)
3. cp - chest pain type

    0: Typical angina: chest pain related decrease blood supply to the heart

1: Atypical angina: chest pain not related to heart

2: Non-anginal pain: typically esophageal spasms (non heart related)

3: Asymptomatic: chest pain not showing signs of disease

4. trestbps - resting blood pressure (in mm Hg on admission to the hospital) anything above 130-140 is typically cause for concern
5. chol - serum cholestoral in mg/dl

   serum = LDL + HDL + .2 * triglycerides

   above 200 is cause for concern

6. fbs - (fasting blood sugar>120 mg/dl)(1 = true, 0 = False)

   '>126' mg/dL signals diabetes

7. restecg - resting electrocardiographic results

   0: Nothing to note

   1: ST-T Wave abnormality

   - can range from mild symptoms to severe problems
   - signals non-normal heart beat

   2: Possible or definite left ventricular hypertrophy

   - Enlarged heart's main pumping chamber
8. thalach - maximum heart rate achieved
9. exang - exercise induced angina (1 = yes, 0 = no)

10. oldpeak - ST depression induced by exercise relative to rest looks at stress of heart during exercise unhealthy heart will stress more
11. slope - the slope of the peak exercise ST segment

    0: Upsloping: better heart rate with exercise (uncommon)

    1: Flat Sloping: minimal change (typical healthy heart)

    2: Downsloping: signs of unhealthy heart

12. ca - number of major vessels (0-3) colored by fluoroscopy

    colored vessel means the doctor can see the blood passing through

    the more blood movement the better (no clots)

13. thal - thalium stress result

    1,3: normal

    6: fixed defect: used to be defect but ok now

    7: reversible defect: no proper blood movement when exercising

14. target - have disease or not (1=yes, 0=no) (= the predicted attribute)

Algorithm

Step 1: Import all the libraries

Step 2: Import the dataset

Step 3: Prepare Training data

Step 4: Perform data preprocessing

Step 5: Prepare the model

Step 6: Train the model

Step 7: Predict the result

Step 8: Connect backend to the frontend

Step 9: Take the image input

Step 10: Display the result

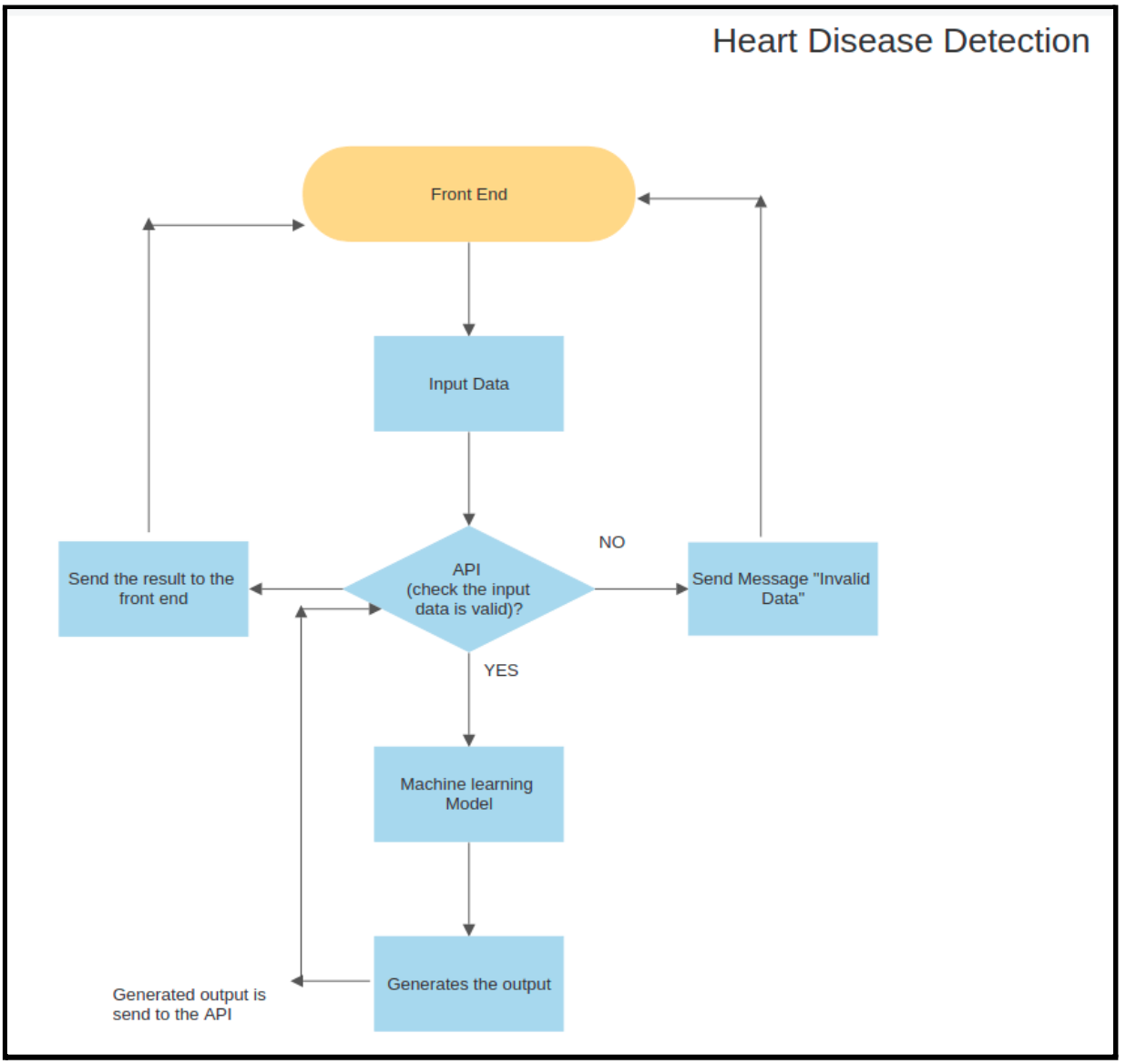

Fig 3.1 FlowChart

Figure 3.1 depicts the picture of the flowchart of the proposed system

3.2 Details of Hardware and Software

Software Requirements:

Python 3.7

Operating System: Windows 7

Python packages: pandas, NumPy, TensorFlow, matplotlib, tqdm, sklearn, xgboost,

Hardware Requirements:

RAM: 4 GB

Processor: Intel i5 9th gen

GPU: 2 GB Nvidia GeForce GTX 1650

3.3 Design Details

3.3.1 Use Case

A use case diagram is a graphical depiction of a user's possible interactions with a system. A use case diagram shows various use cases and different types of users the system has and will often be accompanied by other types of diagrams as well.

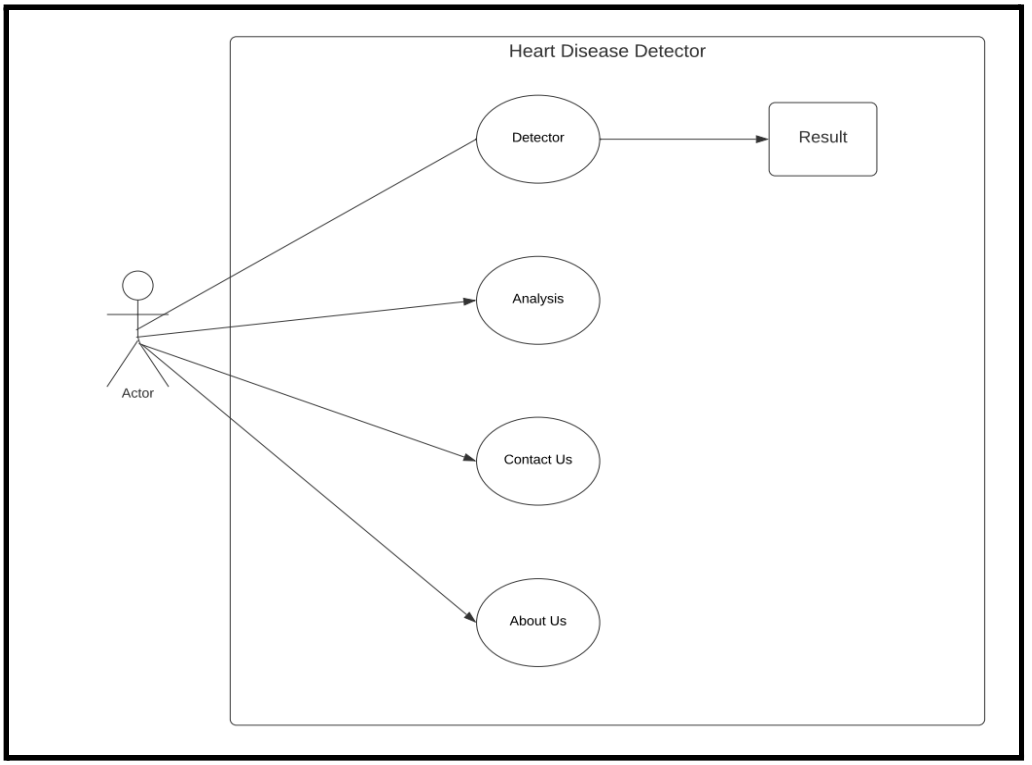

Fig 3.2 Use Case

Figure 3.1 depicts the picture of the use case diagram of our proposed system.

3.3.2 Data Flow Diagram

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It can be manual, automated, or a combination of both.

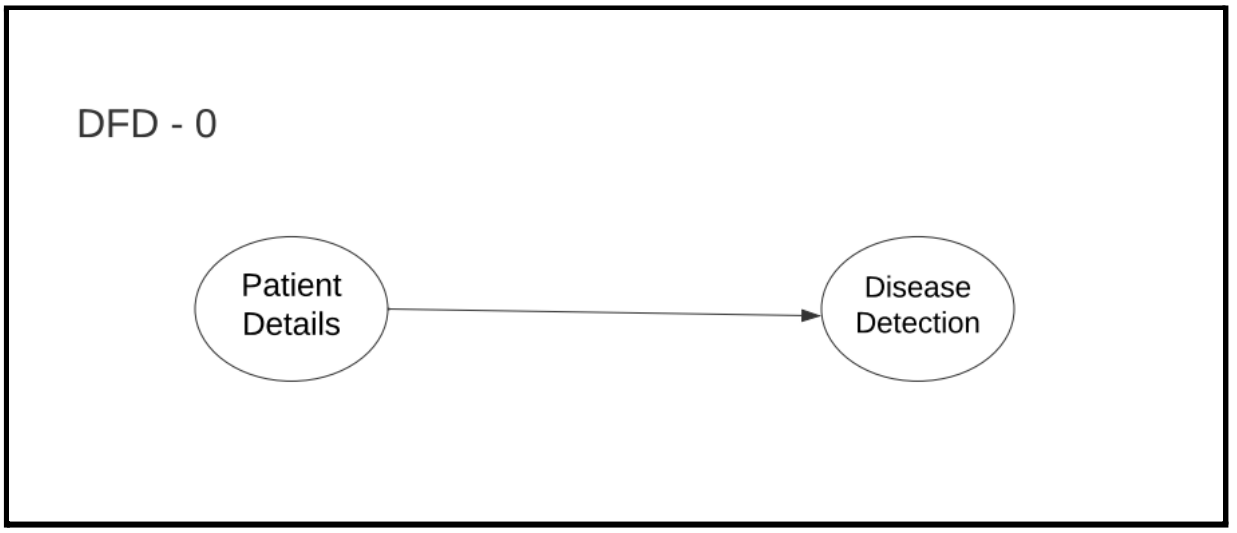

Fig 3.3 DFD Level 0

Figure 3.3 depicts the level 0 Data Flow Diagram of our Proposed System. Users have to enter the details and the disease will be detected based on the data provided by the user.

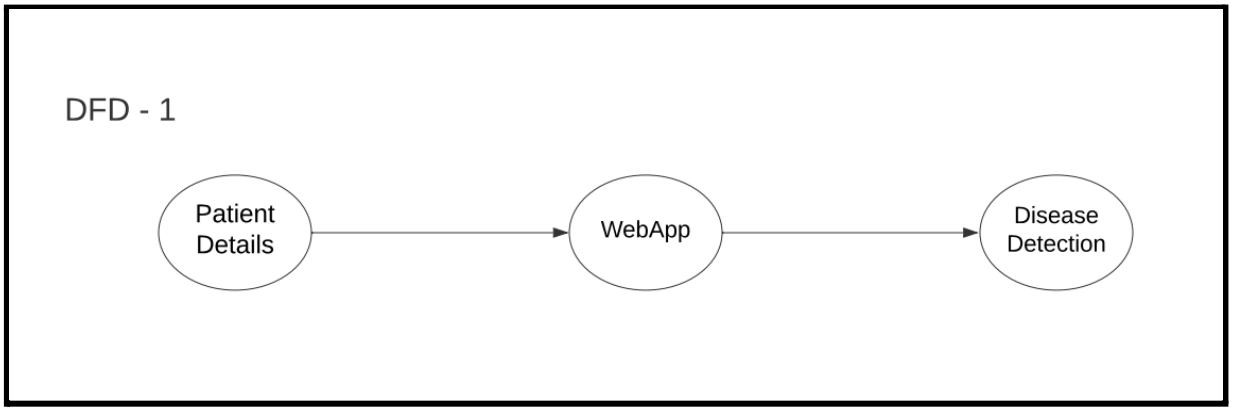

Fig 3.4 DFD Level 1

Figure 3.4 depicts the level 1 Data Flow Diagram of our Proposed System. Users have to enter the details via WebApp and the user will get the predicted results.

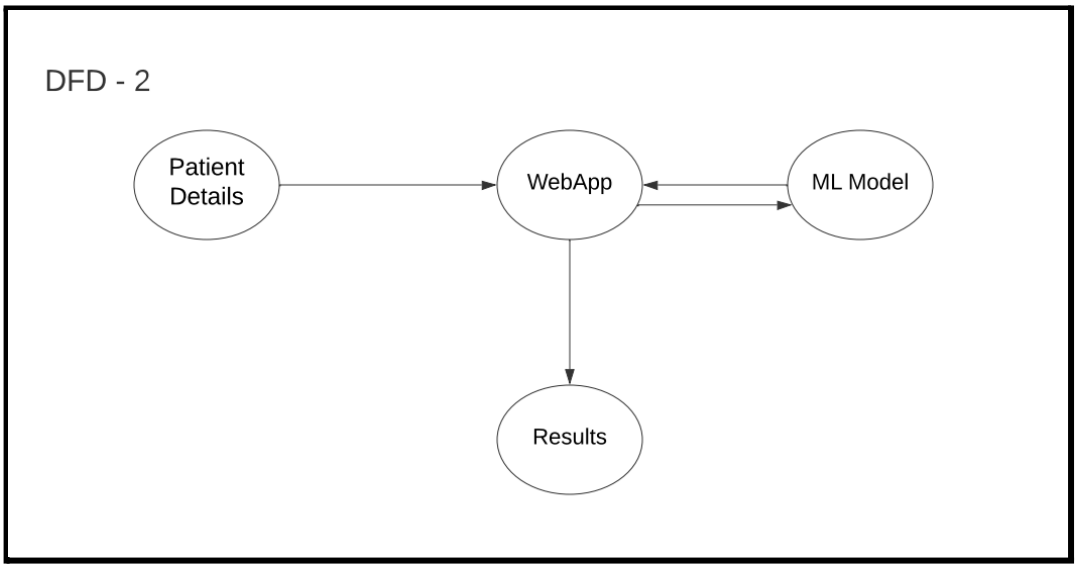

Fig 3.5 DFD Level 2

Figure 3.5 depicts the level 2 Data Flow Diagram of our Proposed System. Users have to enter the details via WebApp and the data will be passed as input to the ML model and then the model will predict the result. The result will be displayed via WebApp.

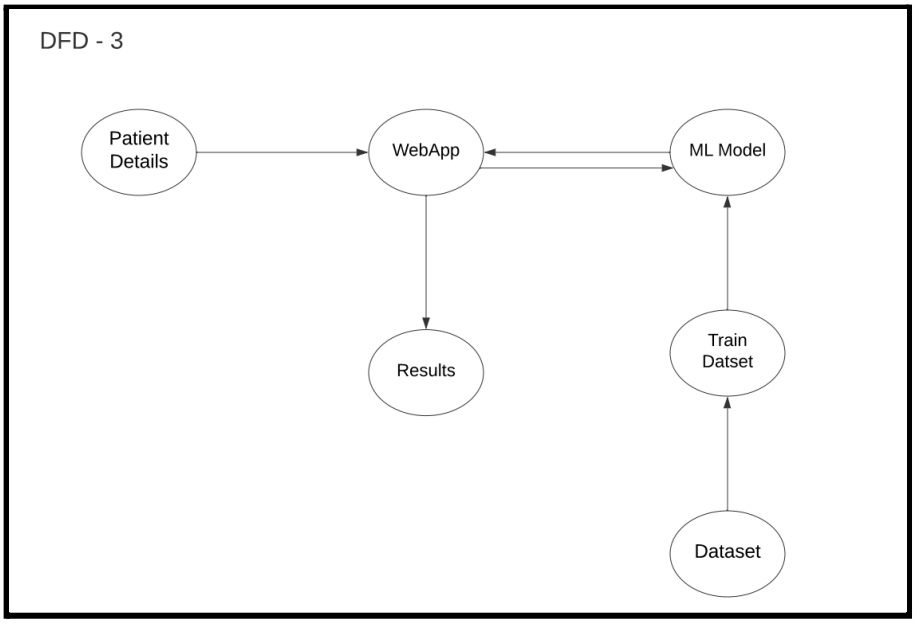

Fig 3.6 DFD Level 3

Figure 3.6 depicts the level 3 Data Flow Diagram of our Proposed System

3.4 Results

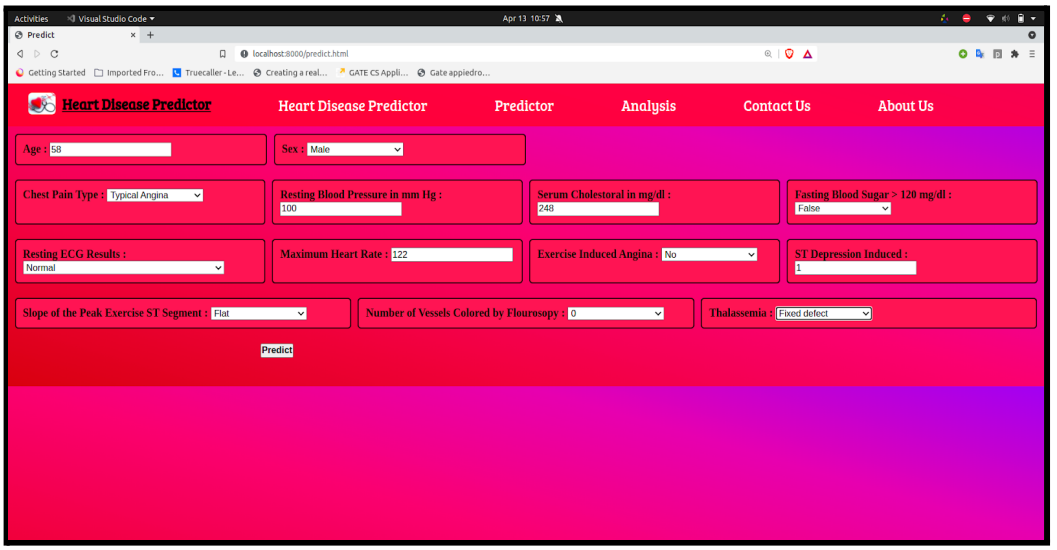

Fig 3.7 Input 1

Figure 3.7 depicts the pictures where user has entered the details of the patient who is having any heart disease.

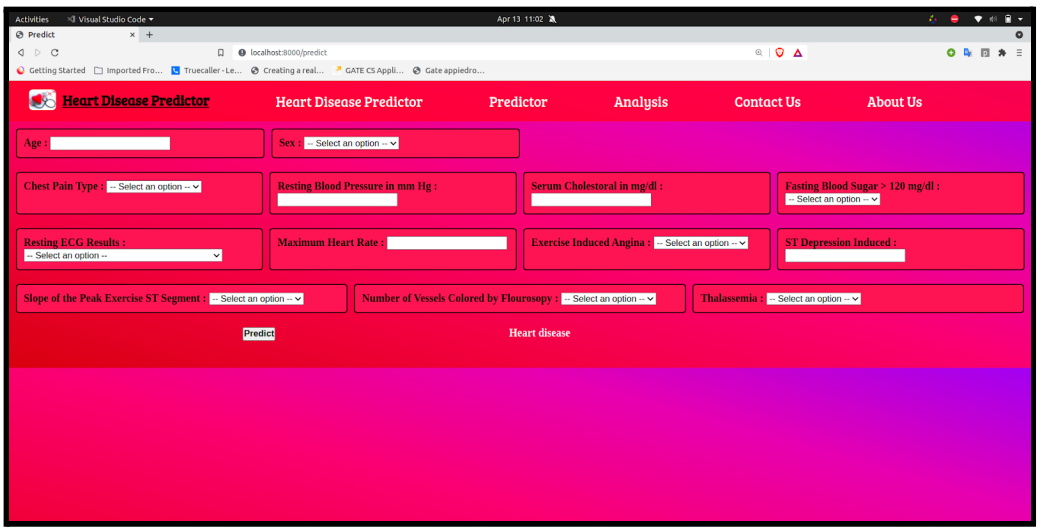

Fig 3.8 Output 1

Figure 3.8 depicts the picture of output where the user gets the results as “ Heart Disease”.

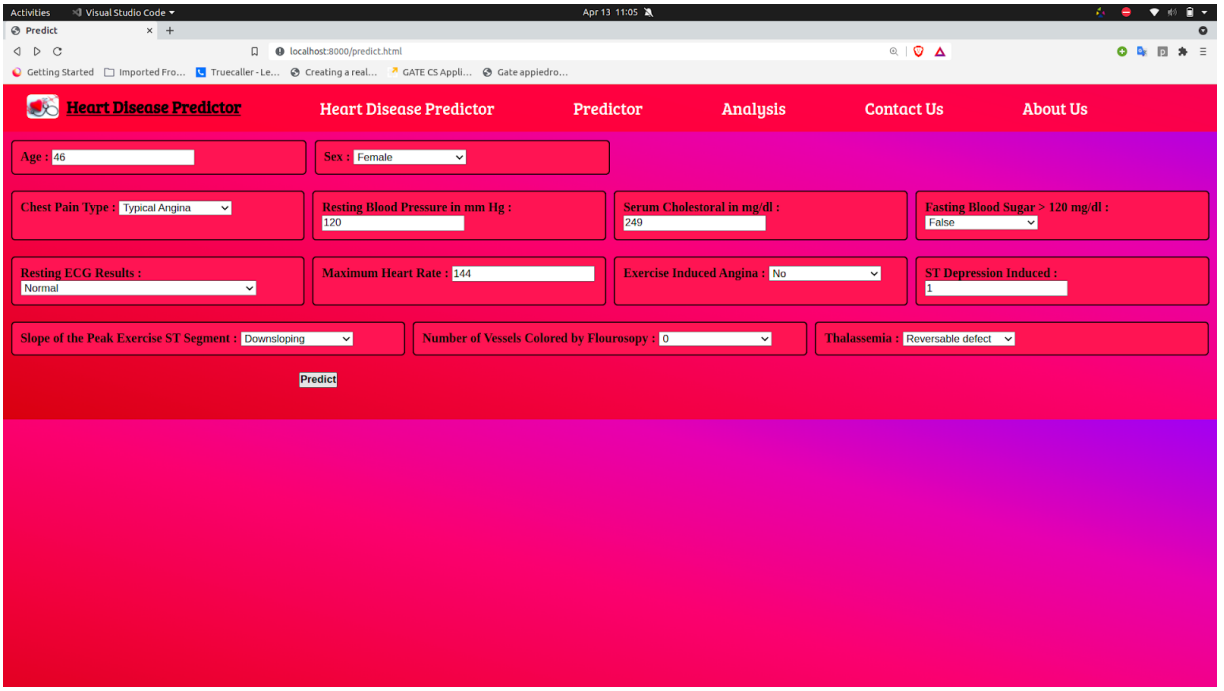

Fig 3.9 Input 2

Figure 3.9 depicts the pictures where the user has entered the details of the patient who is not having any heart disease

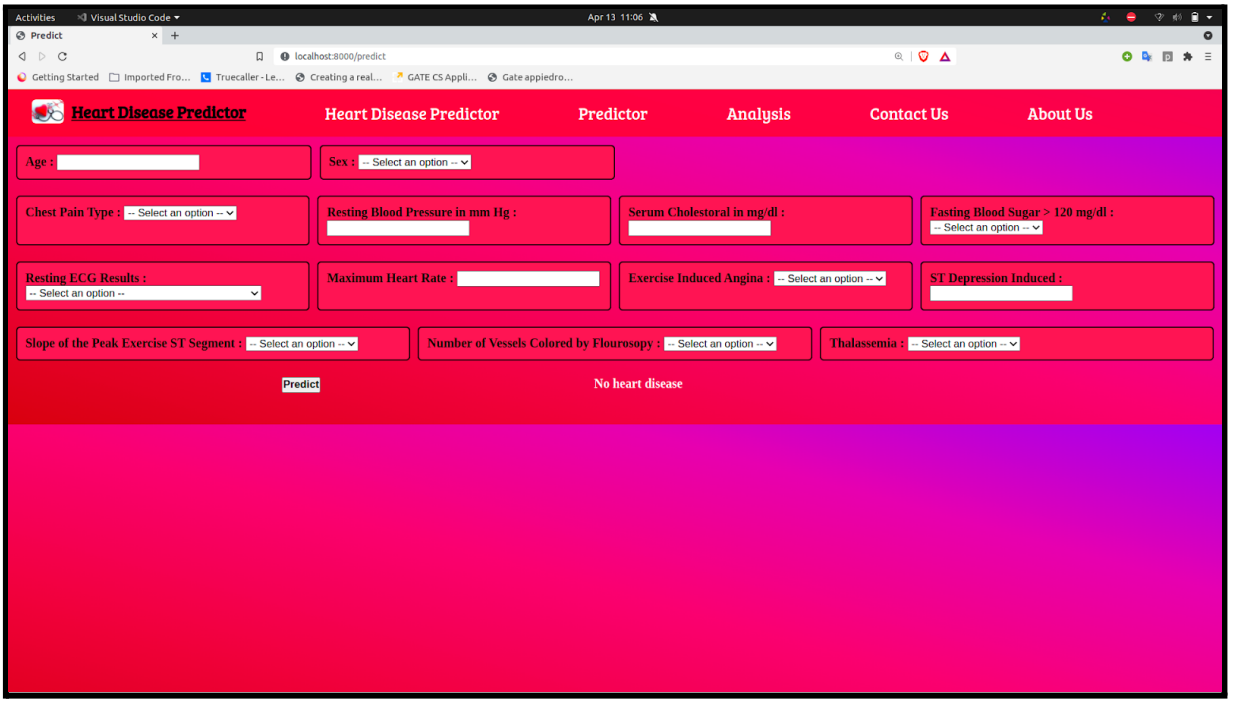

Fig 3.10 Output 2

Figure 3.10 depicts the picture of the output where a user has entered the details of a patient who is not having heart disease and got the correct output as “No Heart Disease”.

Chapter 4

Conclusion

Heart Disease is one of the major concerns for society today. Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias), and heart defects you're born with (congenital heart defects), among others. Heart disease is very fatal and it should not be taken lightly. Heart disease happens more in males than females. It is difficult to manually determine the odds of getting heart disease based on risk factors. However, machine learning techniques are useful to predict the output from existing data. The dataset used for the proposed system was from the kaggle. The dataset contains 303 instances of the patients and 14 attributes like sugar level, cholesterol level, BP level, etc. We have trained four Machine learning models out of which Random forest regression has the highest accuracy, so the proposed system is giving the results based on Random forest Regression.

References:

1. Basha et al, “Early Detection of Heart Syndrome Using Machine Learning Technique” IEEE, 2010.
2. T. E. Chatham and B. Wegbreit, “Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method” IEEE, 2020.
3. Ramya, G. Franklin and Dr. B. Muthukumar, “Survey of Heart Disease Prediction and Identification using Machine Learning Approaches” IEEE, 2019.
4. R. Chen, N. Sun, X. Chen, M. Yang, and Q. Wu, “Supervised feature selection with a stratified feature weighting method,” *IEEE, 2018*.http://en.wikipedia.org/wiki/Data_cleansing, last accessed on: 9/4/2022.
5. https://www.hindawi.com/journals/cin/2021/8387680/, last accessed on: 05/4/2022.
6. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction, last accessed on: 3/3/2022.
7. https://towardsdatascience.com/heart-disease-prediction-73468d63cfc, last accessed on: 5/3/2022.
8. https://github.com/nadaalay/Heart-Disease-Classifier-Web-App, last accessed on: 13/3/2022.

Acknowledgment

We would like to express a deep sense of gratitude towards our guide Prof. ABC, Computer Engineering Department for her/his/their constant encouragement and valuable suggestions. The work that we are able to present is possible because of her/his timely guidance.

We would like to thank Head of the Computer Engineering Department, Prof. Ashwini Save for her support and coordination.

We are also grateful to the teaching and non-teaching staff of the Computer Engineering Department who lend their helping hands in providing continuous support.

Karishma Datt

Vinod Choudhari

Priya Pathak

Mansi Patil