A

Project Report

on

PENTACKLES

submitted for partial fulfillment for the award of

BACHELOR OF TECHNOLOGY

DEGREE

In

Computer Science

By

Aakansha Tyagi (2000290120001)
Aakriti Gupta (2000290120002)
Ayush Jaiswal (2000290120052)
Ayush Tyagi (2000290120056)

Under the Supervision of

Mr. Vivek Kumar Sharma
Assistant Professor

Session 2023-24

Department of Computer Science
KIET Group of Institutions, Ghaziabad
Affiliated to
Dr. A.P.J. Abdul Kalam Technical University, Lucknow

May 2024

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature of Student
Name: Aakansha Tyagi
Roll No.:2000290120001

Signature of Student
Name: Aakriti Gupta
Roll No.:2000290120002

Signature of Student
Name: Ayush Jaiswal
Roll No.:2000290120052

Signature of Student
Name: Ayush Tyagi
Roll No.:2000290120056

Date:

CERTIFICATE

This is to certify that Project Report entitled “Pentackles” which is submitted by in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date:

Supervisor Signature

Mr. Vivek Kumar Sharma

Assistant Professor

Department of Computer Science

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Assistant Professor **Mr. Vivek Kumar Sharma** Department of Computer Science, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day. We also take the opportunity to acknowledge the contribution of **Dr. Ajay Kumar Shrivastava**, Head of the Department of Computer Science, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Signature of Student
Name: Aakansha Tyagi
Roll No.:2000290120001

Signature of Student
Name: Aakriti Gupta
Roll No.:2000290120002

Signature of Student
Name: Ayush Jaiswal
Roll No.:2000290120052

Signature of Student
Name: Ayush Tyagi
Roll No.:2000290120056

ABSTRACT

Cardiovascular diseases (CVDs), also known as heart disease, has replaced other diseases like malaria, strokes, cancer, trachea, bronchus which was once considered the most significant diseases worldwide, as the number one killer globally, notably this is true for India as well as other countries [1]. This means that such illnesses should be identified early enough for effective treatment to be undertaken hence there should be such a reliable and appropriate system. Using machine learning on various medical datasets, these approaches have automated large scale and complex data analysis Machine learning is playing a huge role in today's world. From self-driving cars to medical fields, we can find machine learning everywhere. The medical industry generates a huge amount of patient data which can be processed in a lot of ways. So, with the help of machine learning, we have created a Prediction System that can detect heart disease. For the advancement in growth of healthcare our primary goal of this study work is to determine which patient, depending on several medical parameters, is more likely to develop a cardiac condition [2]. To determine whether the patient is likely to receive a heart disease diagnosis or not, we developed a heart disease prediction algorithm. use the patient's medical history The user must enter various parameters of the disease and the system would display the output whether he/she has the disease or not. We employed four Machine learning algorithms, including KNN and logistic regression, support vector machine, random forest algorithm are used to categorize and predict heart disease patients This project can help a lot of people as one can monitor the persons' condition and take the necessary precautions thus increasing the life expectancy.

LIST OF FIGURES

Figure no:	About	Page no:
Fig 4.1	SDLC model diagram	23
Fig 4.2	DFD level 0	24
Fig 4.3	DFD level 1	25
Fig 4.4	DFD level 2 (Login)	25
Fig 4.5	DFD level 2 (User Account Management)	26
Fig 4.6	DFD level 2 (Chatbot Management)	26
Fig 4.7	DFD level 2 (Videos module)	27
Fig 4.8	DFD level 2 (Prediction module process)	27
Fig 4.9	Use Case Diagram	28
Fig 4.10	ER diagram	29
Fig 7.1	Efficiency vs algorithm graph	41

LIST OF TABLES

Table no:	Description	Page no:
Table 2.1	Literature Review	12-15
Table 5.1	Dataset Description	38
Table 6.1	Test Deliverables	39
Table 6.2	Manual Testing	40

LIST OF ABBREVIATIONS

Abbreviations	Stands for
CVD	Cardiovascular disease
KNN	K-Nearest Neighbor
API	Application Program Interface
SVM	Support Vector Machine
ML	Machine Learning
UI	User Interface

TABLE OF CONTENTS

		Page no.
DECLARATION		
CERTIFICATE		
ACKNOWLEDGEMENT		
ABSTRACT		i
LIST OF FIGURES		ii
LIST OF TABLES		iii
LIST OF ABBREVIATIONS		iv
CHAPTER 1 INTRODUCTION		
1.1	Introduction to Project	1
1.2	Project Category	2
1.3	Objectives	2
CHAPTER 2 LITERATURE REVIEW		
2.1	Literature Review	3-15
2.2	Problem Formulation	15-16
CHAPTER 3 PROPOSED SYSTEM		
3.1	Proposed System	17
3.2	Unique Features of The System	17-18
CHAPTER 4 REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION		
4.1	Feasibility Study	19-20
4.2	Software Requirement Specification	
4.2.1	Data Requirement	20
4.2.2	Functional Requirement	20
4.2.3	Performance Requirement	21
4.2.4	Maintainability Requirement	21
4.2.5	Security Requirement	22
4.3	SDLC Model Used	22-24

4.4	System Design	
	4.4.1 Data Flow Diagrams	24-27
	4.4.2 Use Case Diagrams	28
4.5	Database diagram	29
CHAPTER 5	IMPLEMENTATION	
5.1	Introduction Tools and Technologies used	30-37
5.2	Dataset Description	37-38
CHAPTER 6 TESTING, AND MAINTENANCE		
6.1 Testing Techniques and Test Cases Used		39-40
CHAPTER 7 RESULTS AND DISCUSSIONS		
7.1	Presentation of Results	41-42
7.2	Performance Evaluation	42-43
7.3	Key Findings	43-44
7.4	Snapshots of System with Brief Detail of Each	44-52
CHAPTER 8 CONCLUSION AND FUTURE SCOPE		53
REFERENCES		54-55
Research Paper Acceptance Proof		56

CHAPTER 1
INTRODUCTION

1.1 Introduction to project

In this digital world, data is an asset, and enormous data was generated in all the fields. Data in the healthcare industry consists of all the information related to patients.Today, many research workers use diverse machine learning models to help the physicians make a correct cardiac disease diagnosis. The brain has top priorities and the heart comes next in the body of individual. The blood is pumped by the heart throughout the body, feeding each organ with it. It is an important task in the medical field to predict the incidence of heart disease.

In addition to raising health care costs, heart-related illnesses lower personal productivity. Globally, medical organizations gather information on a range of health-related topics. A variety of machine learning techniques can be applied to these data to obtain insightful knowledge. However, the amount of data being gathered is enormous, and it frequently contains a lot of noise.[2] These datasets can be easily explored with a variety of machine learning techniques, even though they are too large for human minds to process. As a result, these algorithms have recently shown to be extremely helpful in accurately predicting the presence or absence of heart-related diseases

Many individuals who would benefit from preventive treatment are not identified by current methods of predicting cardiovascular risk, and others end up receiving needless intervention. Using machine learning to take advantage of intricate relationships between risk factors presents a chance to increase accuracy. We investigated whether machine learning could enhance the prediction of cardiovascular risk.

Our goal in this study is to find out how well different machine learning algorithms predict heart disease. To accomplish this, we used a number of strategies. The study's dataset is accessible to the general public on Kaggle. Python was used for all of the computation, preprocessing, and visualization on Google Collab Up to 94% of heart disease cases have been predicted accurately in the past utilizing machine learning techniques.

1.2 Project Category

For a heart disease prediction system, the project category would be "Application or System Development" since it involves the development of a software application or system aimed at predicting heart disease based on various factors such as medical history, lifestyle habits, and physiological data.

1.3 Objectives

Early Detection:

Pentackles can identify individuals at risk of heart disease at an early stage based on various parameters including medical history

User interface:

Pentackles is an intuitive and user-friendly interface for healthcare professionals to input patient data and interpret the prediction results efficiently.

Improving patient outcomes:

Ultimately, the overarching objective is to improve patient outcomes by enabling early intervention, personalized risk management, and better allocation of healthcare resources in the prevention and treatment of heart disease.

CHAPTER 2
LITERATURE REVIEW

2.1 LITERATURE REVIEW

2.1.1

- **Title**: A Hybrid Cauchy Crazy Particle Swarm Optimization Support Vector Machine Algorithm for Heart disease prediction
- **Author**: Mandakini Priyadarshani Behera et al.
- **Journal Name**: Procedia Computer Science
- **Year of Publishing**: 2022
- **Summary**-

1. Provided an algorithm for heart and liver disease prediction, as well as heart and liver datasets. SVM, PSOSVM, CPSOSVM, and CCPSOSVM are four algorithms used to predict heart and liver disease. The performance of each method has been calculated and assessed in terms of the confusion matrix, classification accuracy, classification error rate, precision, recall, and F1 score.
2. The heart disease dataset was collected from the University of California, Irvine (UCI), machine learning repository [1]. The heart dataset has 13 attributes and 270 instances, the dataset for liver disease is retrieved from the UCI machine learning repository and the dataset has 10 attributes and 583 instances.
3. The study concludes from rigorous experimental investigation that designed CCPSOSVM gives excellent classification results, with the highest classification rate and lowest error rate for heart and liver sickness prediction.

2.1.2

- **Title:** Machine Learning Model Prediction of Mortality in Patients With and Without Heart Failure
- **Authors:** Se Yong Jang et al**.**
- **Journal Name:** JACC: Advances
- **Year of Publishing:** September 2023
- **Summary:**

1. Predicting mortality risk is vital for tailoring medical care to patients, allowing for appropriate monitoring, therapy referral, and end-of-life counseling. Existing risk prediction models tend to be disease-specific, limiting their applicability across different medical conditions [2]. The MARKER-HF model aims to overcome this limitation by providing a reliable indicator of mortality risk across various diseases.
2. The result of this study shows that the average age of the 41,749 patients was 65, with 56.2% being male.
3. MARKER-HF accurately predicted mortality in subgroups with and without cardiovascular disease, as well as in patients with acute coronary syndrome, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, diabetes mellitus, or hypertension. The study concludes that MARKER-HF predicts mortality for patients with HF as well as for patients suffering from a variety of diseases.

2.1.3

- **Title:** A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease Patients
- **Author:** Alireza Ghasemieh, Alston Lloyed, Parsa Bahrami, Pooyan Vajar, Rasha Kashef
- **Journal Name:** Decision Analytics Journal

- **Year of published:**2023
- **Summary-**

1. Cardiovascular diseases are a leading cause of mortality worldwide, emphasizing the need for early detection and intervention. Existing machine learning approaches for heart disease detection have limitations, resulting in misdiagnoses and overcrowded medical facilities. This study introduces a novel model for identifying patients at risk of emergency readmission, utilizing behavior-based features to create a new class label for emergency readmission.
2. The proposed model employs a robust Stacking Ensemble Learner (SEL) using ensemble learning, with XGBoost as the meta-learner [3]. The model predicts whether a patient with heart problems requires emergency admission after an initial admission.
3. The study acknowledges the potential limitation of the dataset size, which might affect the generalizability of the model.

2.1.4

- **Title-**Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment
- **Author -**Yue Huang a,1 , YingBo Ren a,1 , Hai Yang a , YiJie Ding b , Yan Liu a , YunChun Yang a , AnQiong Mao a , Tan Yang d , YingZi Wang c , Feng Xiao c , QiZhou He e, Ying Zhang a
- **Journal Name-**Computers in Biology and Medicine
- **Year of publishing-**2022

- **Summary-**

1. Cardiovascular disease, particularly coronary heart disease (CHD), is a leading cause of death worldwide. The study aims to evaluate the use of ML models to predict CHD risk more accurately.
2. Conditions such as hypertension, diabetes, fatty liver, osteoporosis, and hyperlipidemia were considered [4]. Various ML models were selected for prediction, including Random Forest (RF), k-nearest neighbor (KNN), support vector machines (SVM), Kernel Ridge Regression (KRR), and radial basis function neural networks (RBFNN). Performance evaluation of the models included metrics such as accuracy, sensitivity, specificity, and the Matthews correlation coefficient (MCC).
3. RF was found to have the best predictive performance in terms of accuracy, sensitivity, specificity, MCC, and AUC compared to other ML models.

2.1.5

- **Title-**Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome
- **Author Name-** Chun-Ling Lin a, Chin-kun Tseng b, Chien-Jen Wang b, Shu-Hung Chao b, Yuh-Shyan Hwang b, Lih-Jen Kau b
- **Journal Name-** International Measurement Confederation
- **Year of publishing-**2023
- **Summary-**

1. It introduces the concept of a smart cardiovascular measurement system that utilizes ECG and PPG to evaluate sleep deprivation, cold hands and feet, and Shanghuo syndrome (from Traditional Chinese Medicine) and discusses the goals of the study

2. The study involved 83 healthy adults, including 26 males and 57 females. Features were extracted from both ECG and PPG signals. A total of 38 features were selected, including 19 from ECG and 19 from PPG. Various machine learning methods were explored, including ensemble learning, kernel classification, k-nearest neighbor, support vector machine, naive Bayes, neural network, and decision tree classifiers. Bayesian optimization was used to fine-tune model selection and hyperparameter values [5]. Model performance was measured using accuracy, computed using true positives, true negatives, false positives, and false negatives.
3. The study acknowledges limitations, including a relatively small sample size and the potential influence of additional factors not considered.

2.1.6

- **Title:** Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure Patients
- **Authors:** Md. Mamun Ali a , Vian S. Al-Doori b , Nubogh Mirzah c , Asifa Afsari Hemu d , Imran Mahmud a , Sami Azam e , Kusay Faisal Al-tabatabaie f , Kawsar Ahmed g,h,∗ , Francis M. Bui g , Mohammad Ali Moni
- **Journal Name:** Healthcare Analytics
- **Published Year:**2023
- **Summary:**

1. Cardiovascular diseases, including HF, are a major global health concern. This study focuses on HF, a condition where the heart struggles to pump enough blood. The research aims to analyze risk factors and predict survival in HF patients using machine learning methods.

2. The study employs a dataset with 299 instances, including 96 deaths and 203 survivors. Data preprocessing involves handling missing values, feature engineering, and balancing the dataset.
3. Five ML algorithms are applied, including Decision Tree, Random Forest, XGBoost, and Gradient Boosting, to train and test the models [6]. Performance metrics (accuracy, precision, recall, F-measure, log loss) are employed for evaluation. RF shows maximum accuracy in this model.

2.1.7

- **Title:** Machine Learning-Based Predictive Models for Heart Disease Diagnosis
- **Author Name:** Stojanov, Lazarova, Veljkova, Rubartelli, Giacomini
- **Journal Name:** King Saud University
- **Year of Publishing:** 2023
- **Summary:**

1. This study explores the use of machine learning techniques to predict heart disease based on various health-related attributes. Feature selection is employed to identify relevant features, and ten different machine learning algorithms are analyzed. The Decision Tree algorithm stands out with the highest accuracy.
2. This study leverages logistic regression as a core model to predict heart failure against chronic-ischemic heart disease in the elderly population [7]. By analyzing biochemical variables, a subset of variables is identified as excellent for discriminating between heart failure and chronic-ischemic heart disease.
3. The study highlights the predictive potential of individual biochemical parameters. While the study faced limitations due to a small sample size, it

offers the groundwork for the development of precise diagnostic software, which can be enhanced with more patient data.

2.1.8

- **Title:** Heart Diseases Prediction based on Stacking Classifiers Model
- **Author:** Subasish Mohapatra, Sushree Maneesha, Prashanta Kumar Patra, Subhadarshini Mohanty
- **Journal Name:** International Conference on Machine Learning and Data Engineering
- **Year of Publishing:** 2023
- **Summary:**

1. The introduction highlights the importance of machine learning in various domains, including healthcare. It emphasizes the need for efficient analysis of the growing volume of medical data generated by digital systems.
2. Cardiovascular diseases are a significant global health concern, and early detection is crucial for effective treatment [8]. Traditional diagnostic methods are slow and costly, motivating the adoption of machine learning for faster and more accurate diagnoses.
3. The methods section outlines the workflow, data preprocessing steps and the stacking approach used for model building. The stacking involves two levels: base learners and meta-learners. The goal is to harness the strengths of diverse classifiers to improve prediction accuracy.

2.1.9

- **Title:** Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data
- **Authors:** Qi Li, Alina Campan, Ai Ren, Wael E. Eid

- **Journal**: International Journal of Medical Informatics
- **Year of Publishing:** 2022
- **Summary:**

1. Cardiovascular disease is a significant health concern globally, with varying prevalence and mortality rates in different regions. Accurate risk assessment is crucial for effective prevention and intervention strategies [9]. Current risk assessment tools like the PCE Risk Calculator may not provide precise risk estimates for all populations. This study aims to develop an automated CVD risk calculator tailored to a specific population using machine learning and EMR data.
2. The authors collected EMR data from over 100,000 patients from a regional healthcare system, spanning from January 1, 2009 to April 30, 2020. Machine learning techniques were applied to these datasets, considering both cross-sectional (CS) features and a combination of CS and longitudinal (LT) features derived from vital statistics and laboratory values. The study evaluates the performance of various machine learning models and compares them to the PCE Risk Calculator.
3. All machine learning models tested outperformed the PCE Risk Calculator in predicting CVD risk. The random forest machine learning technique applied to the combination of CS and LT features achieved the highest accuracy.

2.1.10

- **Title:** Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset
- **Author Names**: Khandaker Mohammad Mohi Uddin, Rokaiya Ripa, Nilufar Yeasmin, Nitish Biswas, Samrat Kumar Dey
- **Journal Name: Intelligence-**Based Medicine

- **Year of Publishing:** 2023
- **Summary:**

1. Heart disease is one of the most serious ailments, killing the majority of its victims. Heart disease is very difficult to diagnose medically. Early identification of heart disease will reduce the risk of mortality [10]. Predicting heart illness has grown to be one of the most challenging medical tasks in recent years due to the prevalence of cardiac issues.
2. In this work, the existence of cardiac anomalies is detected using Machine Learning (ML) approaches. Several machine learning (ML) algorithm techniques, including Decision Tree (DT), Ada-Boost Classifier (AB), Extra Trees Classifier (ET), Support Vector Machine (SVM), Gradient boost, MLP, extreme gradient boost (XGB), Random Forest (RF), KNN, and LR, are used in the proposed method to predict the likelihood of heart disease and classify the risk level of the patient.
3. The testing findings demonstrate that the Decision Tree method has the best accuracy, when compared to other machine learning techniques.

Table 2.1 - literature review

SR No:	Authors	Title	Output	Year
1	Mandakini Priyadarshani Behera et al.	A Hybrid Cauchy Crazy Particle Swarm Optimization Support Vector Machine Algorithm for Heart disease prediction	The study concludes that designed CCPSOSVM gives excellent classification results.	2022
2	Se Yong Jang et al	Machine Learning Model Prediction of Mortality in Patients with and Without Heart Failure	The study concludes that MARKER-HF predicts mortality for patients with HF as well as for patients suffering from a variety of diseases	2023
3	Alireza Ghasemieh et al	A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease	The model predicts whether a patient with heart problems requires emergency admission after an initial admission.	2023

		Patients		
4	Yue Huang et al	Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment	RF was found to have the best predictive performance in terms of accuracy, sensitivity, specificity, MCC, and AUC compared to other ML models.	2022
5	Chun-Ling Lin et al	Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome	The study acknowledges limitations, including a relatively small sample size and the potential influence of additional factors not considered.	2023
6	Md. Mamun Ali et al	Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure	RF shows maximum accuracy in this model.	2023

		Patients		
7	Stojanov et al	Machine Learning-Based Predictive Models for Heart Disease Diagnosis	This study leverages logistic regression as a core model to predict heart failure against chronic-ischemic heart disease in the elderly population	2023
8	Subasish Mohapatra et al	Heart Diseases Prediction based on Stacking Classifiers Model	The methods section outlines the workflow, data preprocessing steps and the stacking approach used for model building.	2023
9	Qi Li et al	Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data	All machine learning models tested outperformed the PCE Risk Calculator in predicting CVD risk. The random forest machine learning technique	2022

			applied to the combination of CS and LT features achieved the highest accuracy.	
10	Khandaker Mohammad Mohi Uddin et al	Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset	The testing findings demonstrate that the Decision Tree method has the best accuracy, when compared to other machine learning techniques.	2023

2.2 Problem Formulation

Heart disease remains one of the leading causes of mortality worldwide, with early detection crucial for effective intervention and prevention. Existing methods rely heavily on clinical assessment, which may not always capture subtle risk factors or early signs of disease. Therefore, there is a pressing need for a reliable and accurate heart disease prediction system leveraging modern computational techniques.

The aim of this project is to develop a robust heart disease prediction system capable of accurately assessing an individual's risk of developing various forms of heart disease based on the medical history of patients.

This project addresses a critical need in cardiovascular medicine by leveraging advances in data science and machine learning to enhance risk prediction and preventive care. The resulting prediction system has the potential to revolutionize clinical practice, empower patients, and ultimately reduce the global burden of heart disease.

CHAPTER 3

PROPOSED SYSTEM

3.1 Proposed System

The primary goal of the suggested approach is to forecast the onset of cardiac disease in order to quickly and accurately recognize the condition. To predict heart disease based on some health characteristics, our method uses a variety of data mining approaches and machine learning algorithms, including Support vector machines, K Nearest Neighbor (KNN), Logistic regression, and Random Forest and Google collab is used to evaluate the data. By importing libraries, we are using this open-source software to implement a variety of machine learning techniques.

By using additional data, data analytics can be applied to predict different diseases; therefore, the medical center can forecast. Some data mining and machine learning approaches that may forecast heart disease including logistic regression, K-Nearest neighbor, random forest, and support vector machine and other.

Training and testing data are divided up. 20% of the data are used for testing, while 80% are used for training.

3.2 Unique Features of The System

Personalized risk assessment

“Pentackles” can provide personalized risk assessments for users based on their demographic information, medical history, lifestyle factors, and genetic predispositions. This ensures that recommendations and preventive measures are tailored to individual needs.

Real-time prediction

Utilizing machine learning algorithms, “Pentackles” can provide real-time predictions of a user's likelihood of developing heart disease. This allows users to take timely preventive actions and make informed decisions about their health.

Educational content

“Pentackles” incorporates educational content through YouTube videos, offering users informative and engaging resources on heart disease prevention, healthy lifestyle habits, dietary recommendations, and exercise routines.

Interactive user interface

The website interface offers an interactive user experience where individuals can input their data and receive immediate feedback on their heart disease risk. This includes visually engaging graphics and charts to help users understand their risk factors.

CHAPTER 4
REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

4.1 Feasibility Study

4.1.1 Technical Feasibility-

Data Availability

Evaluate the availability and quality of datasets required for training machine learning models. This includes medical records, demographic information, lifestyle factors, and genetic data.

Machine learning algorithms

Assess the suitability of various machine learning algorithms such as Random Forest, Logistic regression, SVM, KNN for heart disease prediction, considering factors such as accuracy, scalability, and interpretability.

Integration with websites and YouTube

Determine the technical feasibility of integrating the prediction system with websites and YouTube API for data input, visualization, and educational content delivery.

Hardware and Software requirements

Estimate the hardware and software infrastructure needed to support the system, including servers, databases, and development tools.

4.1.2 Operational Feasibility-

User acceptance

Assess the willingness of users, including individuals and healthcare professionals, to adopt and use the heart disease prediction system.

Scalability

Evaluate the system's scalability to accommodate a growing user base and increasing data volume while maintaining performance and reliability.

4.2 Software Requirement Specification

4.2.1 Data Requirement-

Demographic Information

Age, gender, Ethnicity, Socioeconomic status

Clinical data

Medical history (including previous heart conditions, surgeries, and hospitalizations)

Laboratory test results (e.g., lipid profile, fasting blood glucose, hemoglobin A1c)

Biometric Data

Blood pressure readings (systolic and diastolic)

Cholesterol levels (total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides)

4.2.2 Functional Requirement-

User registration and authentication

Allow users to register and create accounts securely. Implement authentication mechanisms to verify user identity and ensure data privacy.

Data input and collection

Provide forms for users to input demographic information (age, gender), medical history (existing conditions, medications) and biometric data (blood pressure, cholesterol levels).

Data processing and Analysis

Apply machine learning algorithms to process and analyze user-provided data. Generate predictive models to assess the user's risk of developing heart disease based on the input data.

4.2.3 Performance Requirement-

Accuracy: The system should have a high level of accuracy in predicting the likelihood of heart disease. It should be able to correctly identify patients with heart disease and those without it.

Sensitivity and specificity: The system should have high sensitivity and specificity. Sensitivity refers to the system's ability to correctly identify patients with heart disease, while specificity refers to the system's ability to correctly identify patients without heart disease.

User-friendliness: The system should be user-friendly and easy to use for healthcare professionals who may not have a background in data science or machine learning.

Scalability: The system should be able to handle a large volume of data and be scalable to accommodate an increasing number of patients.

4.2.4 Maintainability Requirement-

Modularity: The design of system is with modular components that can be easily modified or replaced without affecting the functionality of other parts of the system. This allows for easier updates and enhancements.

Documentation: Provide comprehensive documentation for the system, including code comments, user manuals, and system architecture documentation. This helps developers understand how the system works and facilitates future maintenance tasks.

Standardization: We follow coding standards and best practices to ensure consistency across the codebase. This makes it easier for new developers to understand and work on the system.

4.2.5 Security Requirement-

Data confidentiality: The system should ensure that patient data is kept confidential and protected from unauthorized access, use, or disclosure.

Data integrity: The system should ensure that patient data is accurate and has not been tampered with or altered in any way.

Data availability: The system should ensure that patient data is available when needed by authorized users.

Access control: The system should have appropriate access controls to ensure that only authorized users can access patient data.

Authentication and authorization: The system should require authentication and authorization to ensure that only authorized users can access and modify patient data**.**

4.3 SDLC Model to Be Used

Agile Model is used in this project. Implementing an Agile model in the development of a heart disease prediction system using machine learning (ML) involves breaking down the development process into smaller, manageable iterations or sprints. Each sprint focuses on delivering a functional increment of the system, allowing for continuous feedback and adaptation.

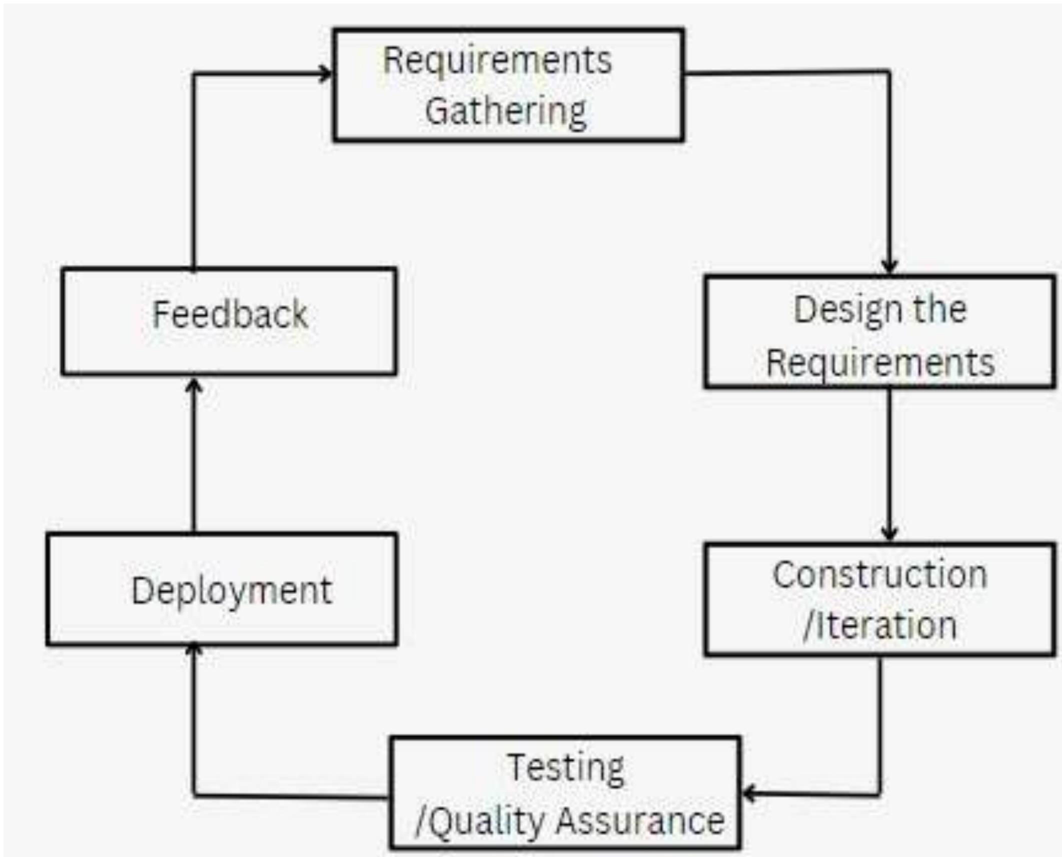


Fig – 4.1

Requirement Gathering- In this step, we gather the requirements, by interaction with the user. We plan the time and effort needed to build the project. Based on this information we evaluate technical and economical feasibility.

Design the Requirements- In this step, we use user-flow-diagram or UML diagrams to show the working of the new features and show how we apply to the existing software. Designing user interfaces are done in this phase.

Construction / Iteration- In this step, we the members of development team start working on the project, which aims to deploy a working product.

Testing / Quality Assurance- Testing involves Unit Testing, Integration Testing, and System Testing.

Deployment- In this step, we deploy the working project to end users.

Feedback- This is the last step of the Agile Model. In this, we receive feedback about the product and works on correcting bugs based on feedback provided by the user.

The Agile model's central principle is delivering an increment to the customer after each iteration.

4.4 System Design

4.4.1 Data flow diagram

DFD level 0

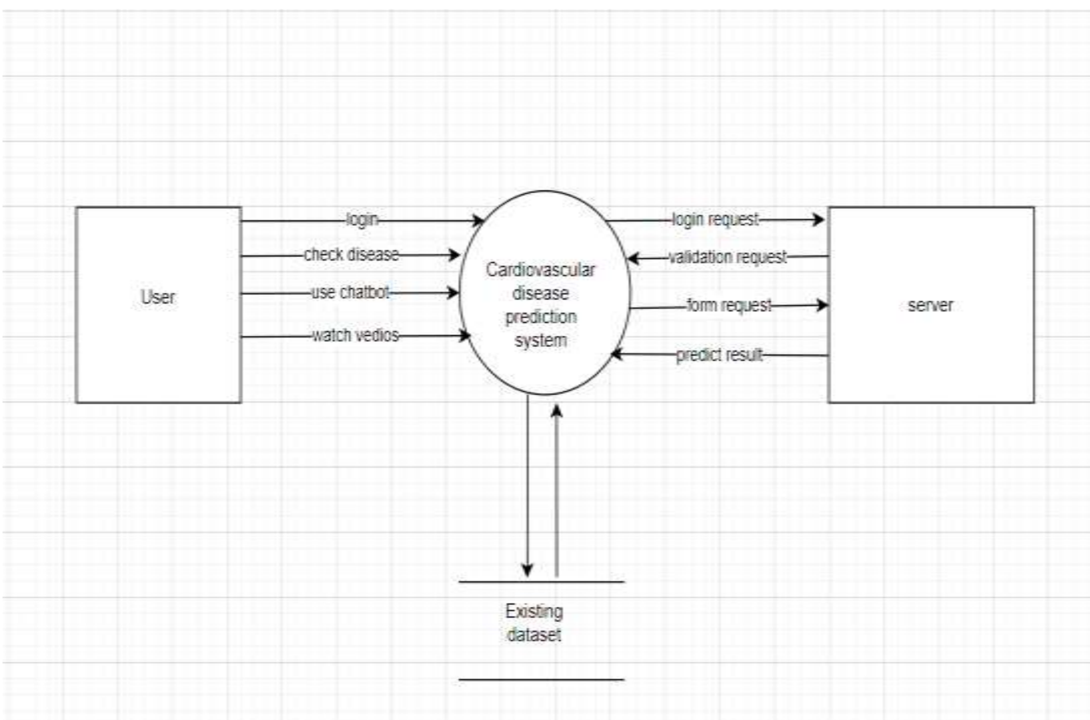


Fig 4.2

DFD level 1

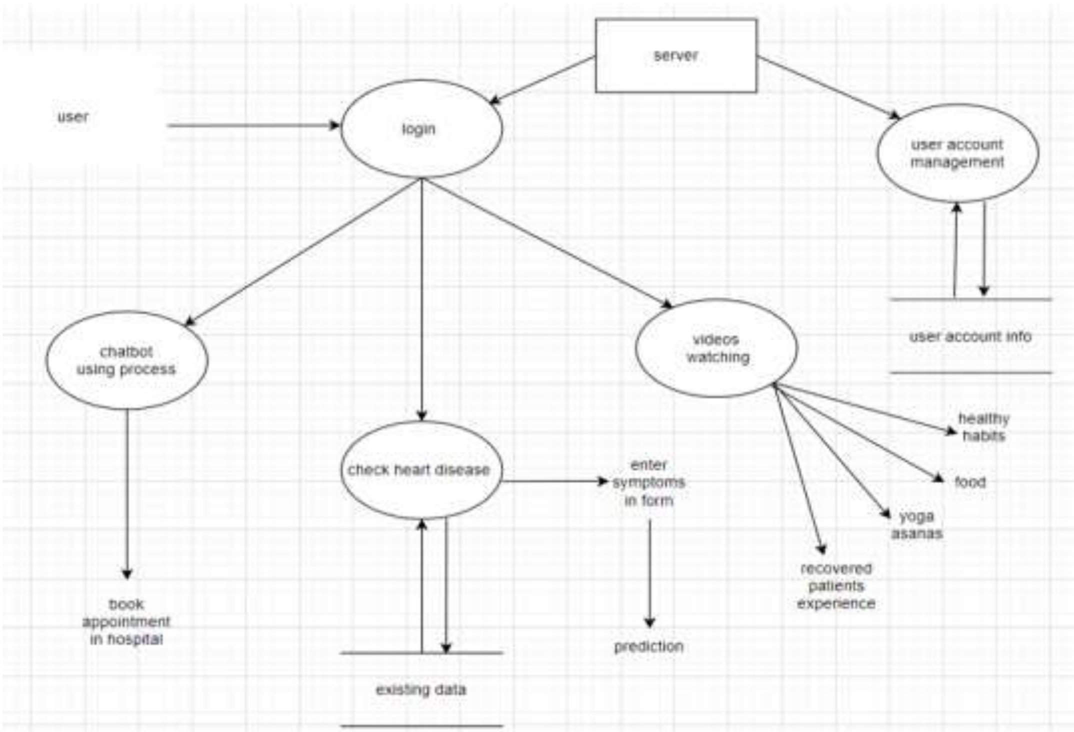


Fig-4.3

DFD level 2

Login

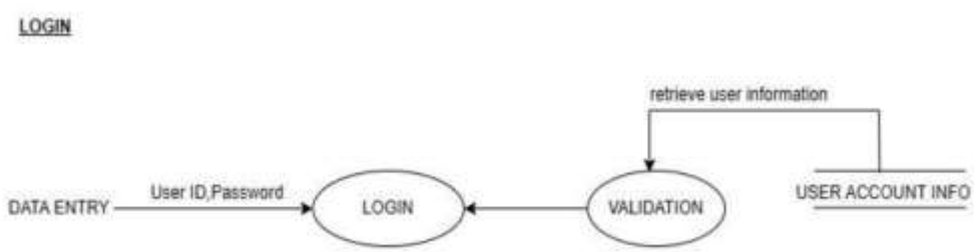


Fig-4.4

User Account Management

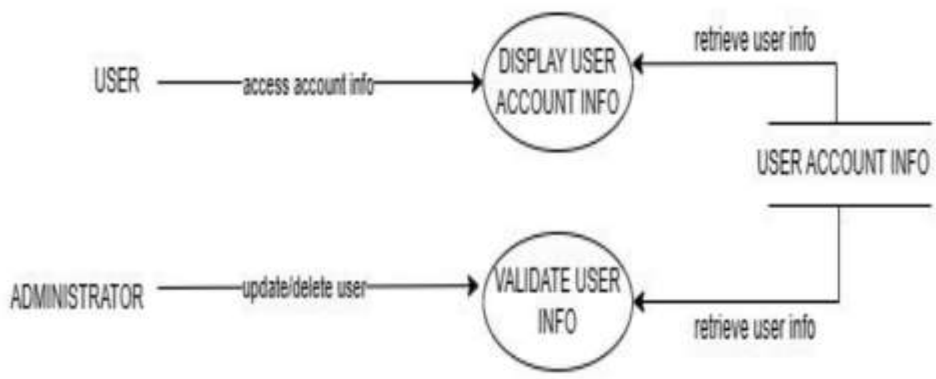


Fig -4.5

Chatbot module management

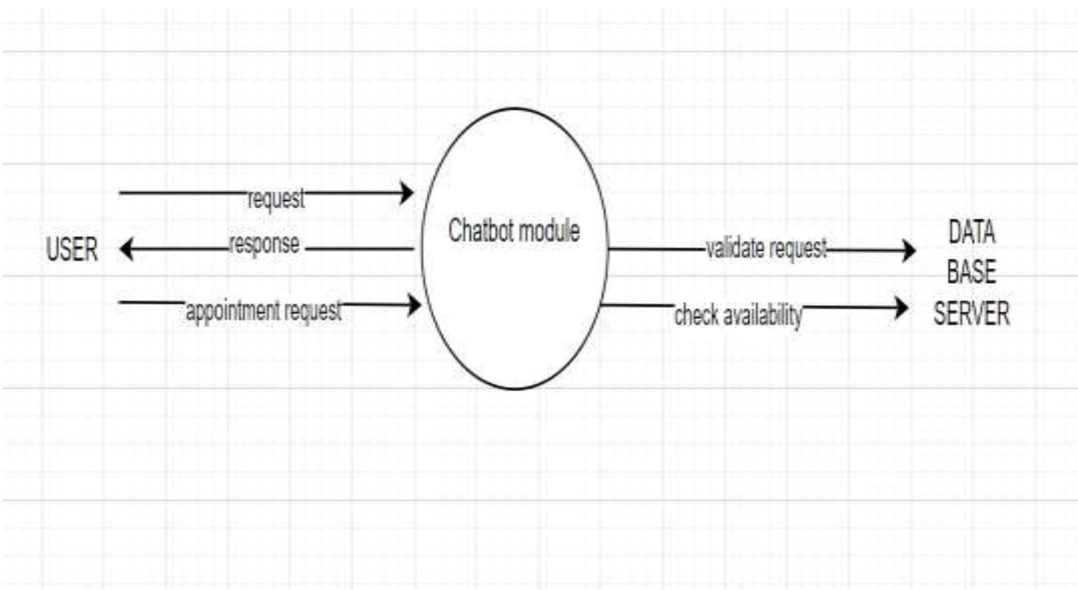


Fig -4.6

Videos module

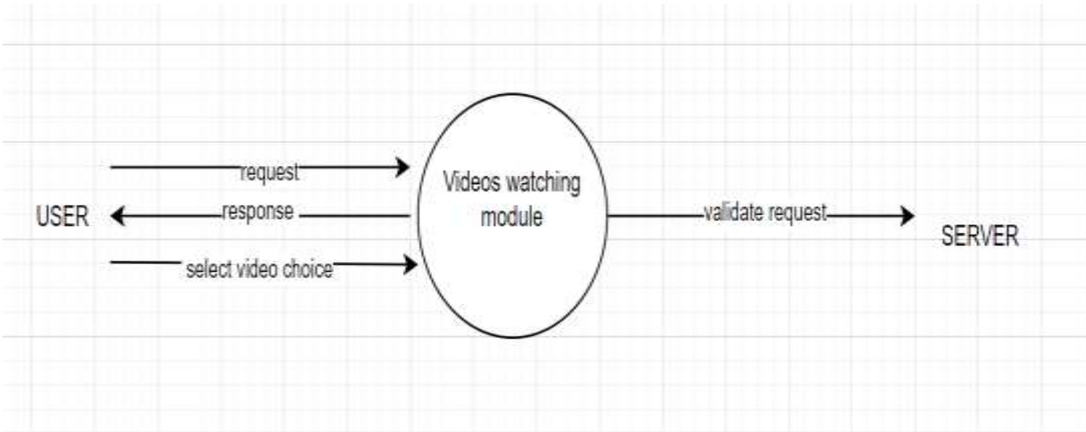


Fig-4.7

Prediction module process

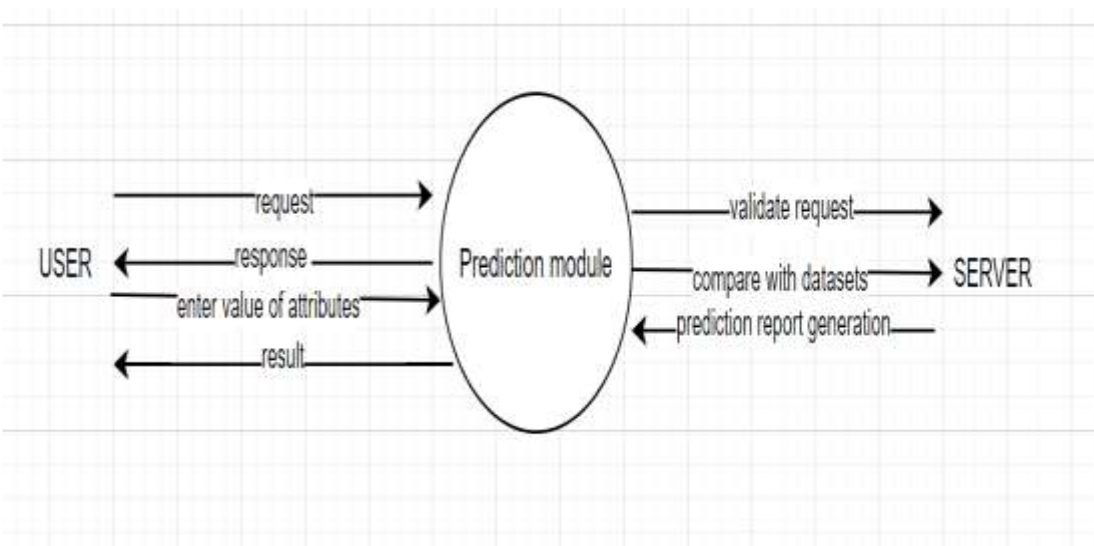


Fig-4.8

4.4.2 Use Case Diagram

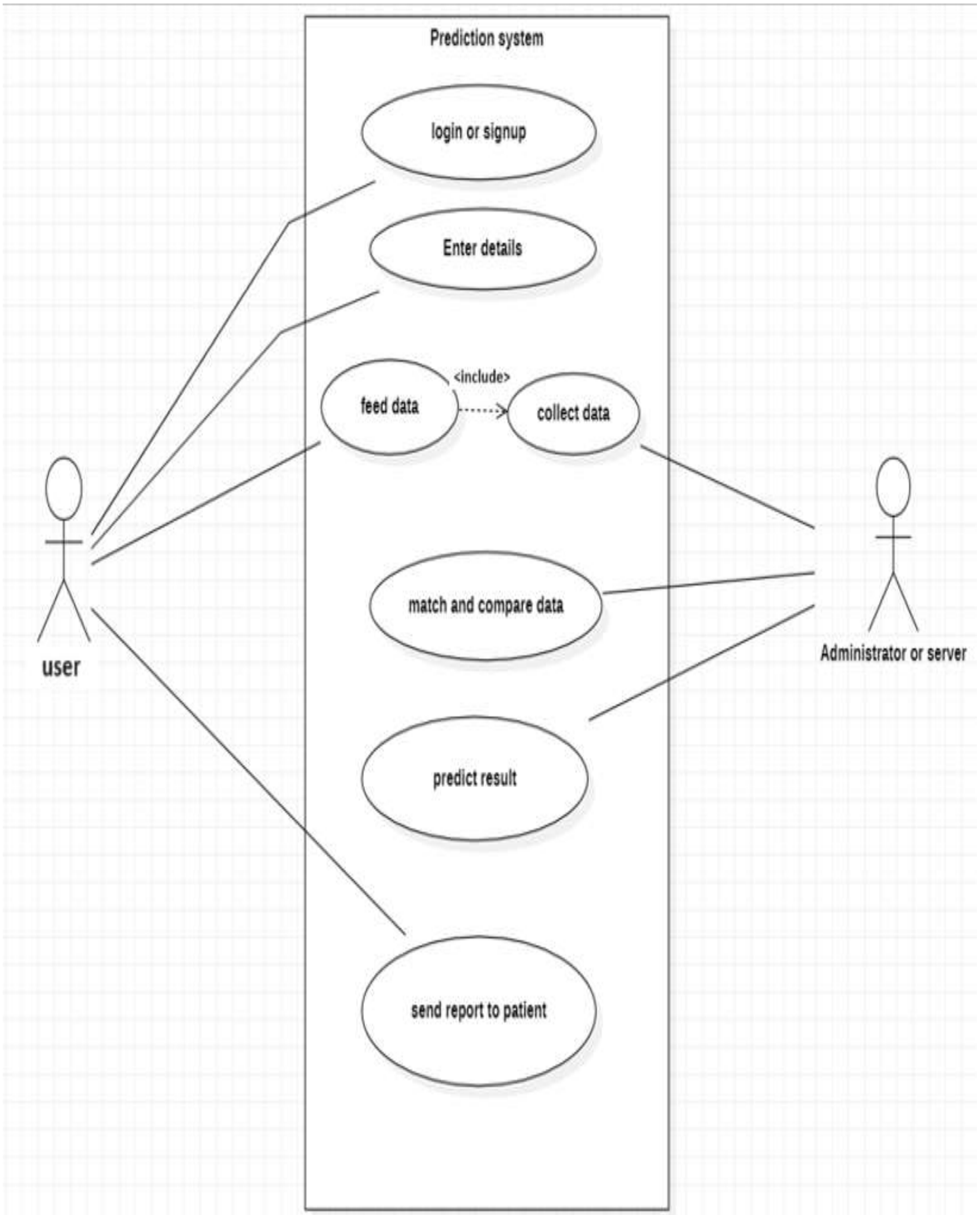


Fig- 4.9

4.5 Database design

ER diagram

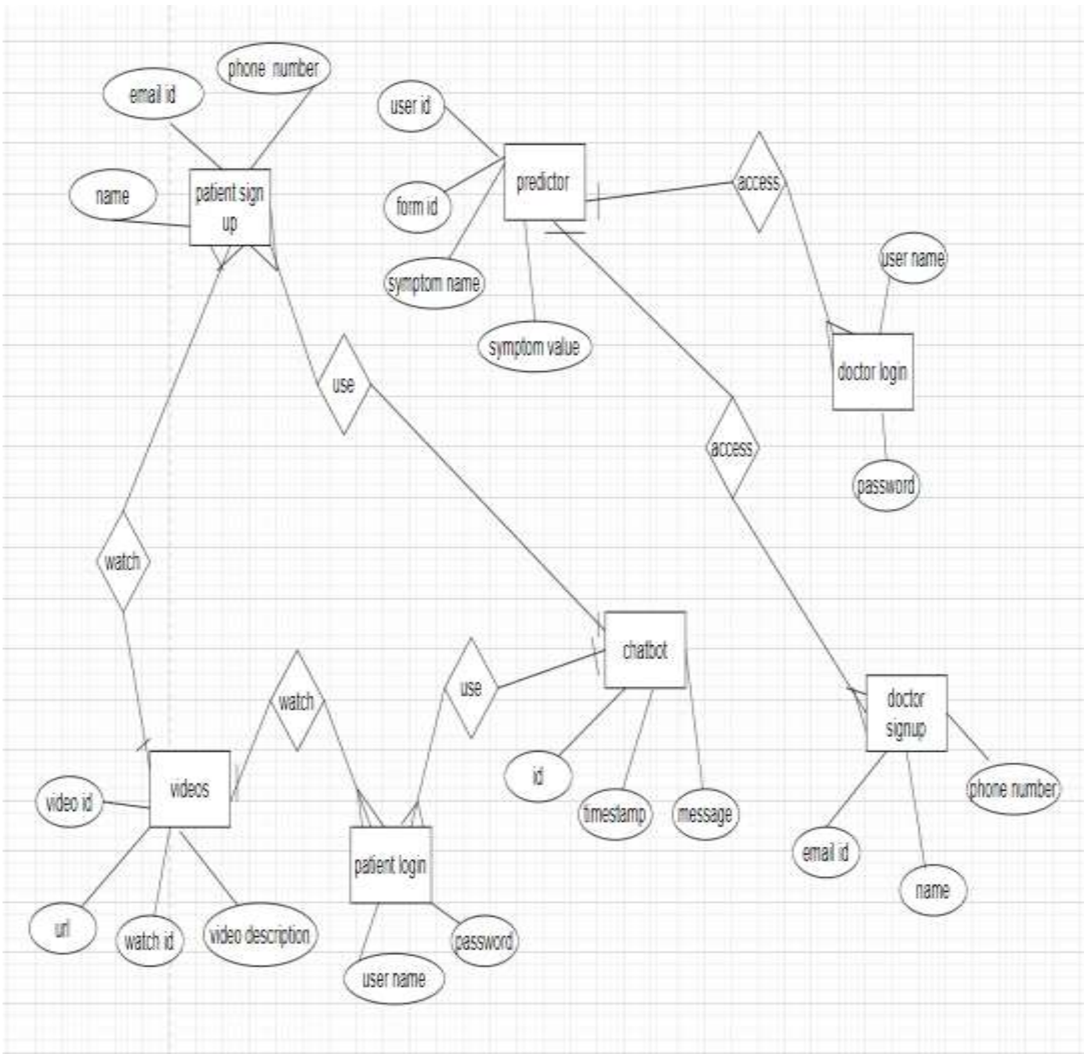


Fig 4.10

CHAPTER 5
IMPLEMENTATION

5.1 Introduction to Languages, Tools and Technologies Used for Implementation

Frontend -

Hypertext Markup language (HTML)

In the development of a heart disease prediction system, HTML serves as the backbone for structuring the user interface (UI) and facilitating data input from users. Through HTML, we design the layout and content of web pages, creating forms where individuals can input crucial health-related information. These forms typically include fields for variables such as age, gender, blood pressure, cholesterol levels, and lifestyle habits, all of which are essential factors in predicting the risk of heart disease. By structuring the UI with HTML, we ensure that users can easily navigate the system and provide the necessary data for accurate prediction. Additionally, HTML's support for semantic elements enhances the accessibility and readability of the interface, promoting a user-friendly experience for individuals seeking to assess their risk of heart disease.

Cascading Style Sheet (CSS)

In the realm of heart disease prediction systems, Cascading Style Sheets (CSS) play a pivotal role in enhancing the visual appeal and usability of the user interface (UI). CSS allows us to customize the appearance of HTML elements, thereby optimizing the presentation of critical health-related information. By employing CSS, we can define the layout, color scheme, typography, and overall aesthetics of the prediction system's UI. For instance, CSS styles can be utilized to create a clean and organized layout for input forms, ensuring that users can easily locate and interact with the necessary fields. Moreover, CSS enables us to establish a cohesive visual identity for the prediction system, aligning it with established design principles and branding guidelines. Through the use of appropriate styling techniques, such as responsive design principles, CSS ensures that the UI remains accessible and functional across various devices and screen sizes. Ultimately, CSS empowers developers to create a visually engaging and user-friendly interface that effectively communicates crucial health information, thereby enhancing the overall user experience of the heart disease prediction system.

JavaScript (JS)

JavaScript (JS) serves as the dynamic powerhouse behind the heart disease prediction system, providing interactivity and functionality that significantly enhance user experience and system performance. JS is instrumental in handling various aspects of the prediction process, from validating user input to dynamically updating the interface with prediction results. One of its key roles is in form validation, ensuring that users provide accurate and appropriate data for analysis. Through JS, we implement client-side validation logic to instantly provide feedback to users if any errors or inconsistencies are detected in their input, thus improving data accuracy and system robustness. Additionally, JS facilitates seamless communication between the user interface and the server-side prediction engine. It enables asynchronous requests to the server, allowing the prediction system to retrieve and process data without reloading the entire page, leading to a smoother and more responsive user experience. Furthermore, JS enables real-time updates of the UI, dynamically displaying prediction results as soon as they are available, thereby keeping users informed and engaged throughout the prediction process. By leveraging the power of JS, developers can create a highly interactive and efficient heart disease prediction system that empowers users to make informed decisions about their cardiovascular health.

Bootstrap

Bootstrap plays a crucial role in streamlining the development process of a heart disease prediction system by providing a comprehensive framework for creating responsive and visually appealing user interfaces. By utilizing Bootstrap's pre-designed components and styles, we efficiently design and implement key elements of the prediction system's interface, such as input forms, navigation menus, and result displays. One of Bootstrap's key advantages is its built-in support for responsive design, ensuring that the prediction system's UI remains accessible and functional across a wide range of devices and screen sizes. This responsiveness is particularly important for a healthcare application like a heart disease prediction system, as it allows users to access and interact with the system seamlessly, whether they are using a desktop computer, tablet, or smartphone. Additionally, Bootstrap offers a wide range of UI components and customization options, allowing developers to tailor the appearance and functionality of the prediction system to meet specific design requirements and branding guidelines. By leveraging Bootstrap's capabilities, we accelerate the development process, maintain consistency across different parts of the UI, and deliver a polished and user-friendly heart disease prediction system that effectively serves the needs of both healthcare professionals and patients.

Backend -

Flask

Flask, a lightweight and versatile web framework for Python, serves as a powerful tool in the development of a heart disease prediction system, facilitating the creation of robust and scalable web applications with ease. Flask's simplicity and flexibility make it well-suited for building the backend infrastructure necessary to support the prediction system's functionality. One of Flask's primary functions is handling HTTP requests and responses, allowing developers to define routes and endpoints through which data can be transmitted between the client-side interface and the server-side prediction engine. In the context of a heart disease prediction system, Flask enables seamless communication between the user interface, where users input their health data, and the backend prediction model, which analyzes this data to generate risk assessments. Furthermore, Flask integrates seamlessly with popular Python libraries and tools for data processing and machine learning, enabling developers to efficiently implement the prediction model within the web application. Flask's modular design also facilitates the implementation of additional features, such as user authentication, data logging, and result visualization, enhancing the overall functionality and usability of the prediction system. Overall, Flask empowers developers to build a reliable, scalable, and feature-rich heart disease prediction system that delivers accurate and actionable insights to healthcare professionals and patients alike.

Python (NumPy, Pandas, Seaborn, Matplotlib)

Python, along with libraries like NumPy, Pandas, Seaborn, and Matplotlib, forms the backbone of the data processing, analysis, and visualization pipeline in a heart disease prediction system. NumPy provides essential functionalities for numerical computing, such as handling arrays and mathematical operations, making it indispensable for data manipulation tasks. Pandas, on the other hand, offers powerful data structures and tools for data wrangling, enabling efficient handling of structured datasets, including preprocessing and feature engineering tasks. These libraries together empower developers to prepare and preprocess raw patient data, extracting relevant features and formatting them in a suitable manner for input into the prediction model.

Once the data is prepared, Seaborn and Matplotlib come into play for data visualization, allowing developers to explore and understand patterns within the dataset and gain insights into the relationships between different variables. Seaborn offers a high-level interface for creating informative and visually appealing statistical graphics, while Matplotlib provides more fine-grained control over plot customization and layout. Through the use of these visualization tools, developers can identify potential correlations and trends in the data, which may inform the selection of features for inclusion in the prediction model.

Moreover, these Python libraries seamlessly integrate with machine learning algorithms for model training and evaluation. With the processed data, developers can build predictive models using machine learning techniques such as logistic regression, decision trees, or neural networks. The predictive performance of these models can then be assessed using techniques like cross-validation or holdout validation, with evaluation metrics provided by libraries like scikit-learn.

Machine Learning-

Support Vector Machine (SVM)

Support Vector Machine (SVM) algorithms are a powerful tool in the development of heart disease prediction systems, offering robust performance in both classification and regression tasks. SVMs excel in identifying complex patterns and relationships within data, making them well-suited for analyzing the intricate interplay of various risk factors associated with heart disease.

In the context of a heart disease prediction system, SVMs can be leveraged to classify patients into different risk categories based on their health data. By training an SVM model on a dataset containing features such as age, gender, blood pressure, cholesterol levels, and lifestyle habits, developers can create a predictive model capable of accurately assessing an individual's risk of developing heart disease. One of the key advantages of SVMs is their ability to handle high-dimensional data effectively, making them particularly suitable for scenarios where the number of features exceeds the number of samples, as is often the case in healthcare datasets.

Random Forest

Random Forest is a versatile and powerful machine learning algorithm widely used in heart disease prediction systems due to its ability to handle complex datasets and produce accurate predictions. In the context of heart disease prediction, Random Forests offer several advantages, making them a popular choice among data scientists and healthcare professionals.

One of the key strengths of Random Forests is their ability to handle high-dimensional data with a large number of features, which is common in healthcare datasets containing various physiological and lifestyle variables. Random Forests work by constructing an ensemble of decision trees, where each tree is trained on a random subset of the features and samples from the dataset. By aggregating the predictions of multiple trees, Random Forests reduce the risk of overfitting and improve generalization performance, resulting in more robust and reliable predictions.

Logistic Regression

Integrating logistic regression into a heart disease prediction system offers a pragmatic and interpretable solution for assessing the likelihood of heart disease occurrence. Initially, a comprehensive dataset encompassing pertinent patient attributes like age, gender, cholesterol levels, and blood pressure, alongside corresponding heart disease labels, is collected and preprocessed to ensure data quality. Feature selection and engineering techniques are employed to identify and refine predictive features, thereby enhancing the model's efficacy.

Through rigorous training on a partitioned dataset, the logistic regression model optimizes parameters to minimize loss, subsequently evaluated using metrics like accuracy, precision, and recall. Post-training, the model is seamlessly integrated into the prediction system, ensuring smooth interoperability with other components. Continuous monitoring and periodic retraining maintain model performance over time, crucial for adapting to evolving patient demographics and healthcare practices. Leveraging the inherent interpretability of logistic regression, the system can provide transparent insights into the factors influencing heart disease risk, fostering user understanding and trust. This iterative process of deployment, monitoring, and maintenance underpins the reliability and utility of the heart disease prediction system, empowering healthcare professionals in early intervention and personalized patient care.

K Nearest Neighbor(KNN)

Integrating the K Nearest Neighbors (KNN) algorithm into a heart disease prediction system offers a versatile and intuitive approach to classification tasks. Initially, a comprehensive dataset comprising relevant patient attributes, including age, gender, cholesterol levels, and blood pressure, along with corresponding heart disease labels, is collected and preprocessed to ensure data quality. Feature engineering techniques may be applied to enhance the discriminative power of the features. Through model training on partitioned data, the KNN algorithm learns to classify patients by identifying the most similar instances in the training set based on their feature vectors. Model evaluation, conducted using metrics like accuracy, precision, recall, and AUC-ROC, provides insight into the model's predictive performance. Once validated, the trained KNN model is seamlessly deployed into the heart disease prediction system, facilitating seamless interaction with other system components. Continuous monitoring and periodic retraining of the model ensure its adaptability to evolving patient demographics and healthcare practices. Leveraging the simplicity and transparency of the KNN algorithm, the system can offer interpretable insights into heart disease risk factors, aiding clinicians in decision-making and patient care. This iterative process of deployment, monitoring, and maintenance underpins the reliability and effectiveness of the heart disease prediction system, empowering healthcare professionals in early detection and intervention strategies.

5.2 Dataset Description

The ".csv" dataset from Kaggle is a comprehensive dataset related to heart health. These attributes include demographic information such as age, sex, and chest pain type, as well as physiological measurements like cholesterol levels, resting blood pressure, and maximum heart rate achieved during exercise.

For my heart disease prediction project, we preprocessed the data by handling missing values, encoding categorical variables, and scaling numerical features. Then, we employed various machine learning algorithms such as logistic regression, decision trees, random forests, and support vector machines to train predictive models.

Table 5.1 Dataset Description

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
42	1	0	140	226	0	1	178	0	0	2	0	2	1
61	1	2	150	243	1	1	137	1	1	1	0	2	1
40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
71	0	1	160	302	0	1	162	0	0.4	2	2	2	1

After evaluating the performance of these models using metrics such as accuracy, precision, recall, and F1 score, we selected the most promising model and fine-tuned its hyperparameters using techniques like grid search or random search. Finally, we deployed the optimized model to predict the likelihood of heart disease in new individuals based on their input features.

CHAPTER 6

TESTING, AND MAINTENANCE

6.1 Testing Techniques and Test Cases Used

Test Environment-

The following software is required in addition to client-specific software.

Development Environment: Google colab, numpy, Pandas, Seaborn, Matplotlib, VS Code (Visual Studio Code) for scripting and coding related to data analysis for predicting heart disease.

Data Format: Ensure that the input data files are in the required format, such as CSV, Excel, or any other applicable format for predicting heart disease.

Testing Tool-

No Tools is used for testing because manual Testing is done.

Table 6.1-Test Deliverables

Test case ID	Username	Password	Expected Outcome	Actual Outcome	Pass/Fail
1	aakriti	Aa_1@	Login Successfull	Login Successfull	Pass
2	aniket	aA24ftr	Login Successfull	Login Successfull	Pass
3	avi	Ak34@	Error	Error	Pass
4	anmol	Hhuy5@	Login Successfull	Login Successfull	Pass
5	dgryj	Gjyrb@7	Login Successfull	Login Successfull	Pass
6	dggfbd	Fh4@o	Login Successfull	Login Successfull	Pass
7	werct	mbR3@	Login Successfull	Login Successfull	Pass
8	anaisha123@gmail.com	ghgt	Error	Error	Pass
9	sanvi@kiet.edu	Ag3@jk	Login Successfull	Login Successfull	Pass
10	an1	Adsu%2	Error	Error	Pass

Tabel 6.2-Manual Testing

Case Id	Input Parameters							Expected output	Actual Output	Status
	Age	BP	Blood Sugar	Heart rate	Cholestrol	ST depression	Vessel colour flouroscopy			
1	0	0	0	0	0	0	0	Invalid data	Invalid data	Pass
2	20	80	70	100	128	3	1	Valid	valid	Pass
3	58	0	80	0	138	2	1	Invalid data	Invalid data	Pass
4	89	98	0	90	0	7	1	Invalid data	Invalid data	Pass
5	29	3	95	98	0	8	1	Invalid data	Invalid data	Pass
6	46	109	98	0	100	9	1	Invalid data	Invalid data	Pass
7	53	111	100	102	102	5	1	Valid	valid	Pass
8	5	0	0	89	135	3	1	Invalid data	Invalid data	Pass

CHAPTER 7

RESULTS AND DISCUSSIONS

7.1 Presentation of Results

SVM demonstrates the highest accuracy and superior performance across multiple evaluation metrics compared to LR, RF, and KNN.

The robustness and discriminative ability of SVM make it a promising choice for heart disease prediction applications**.**

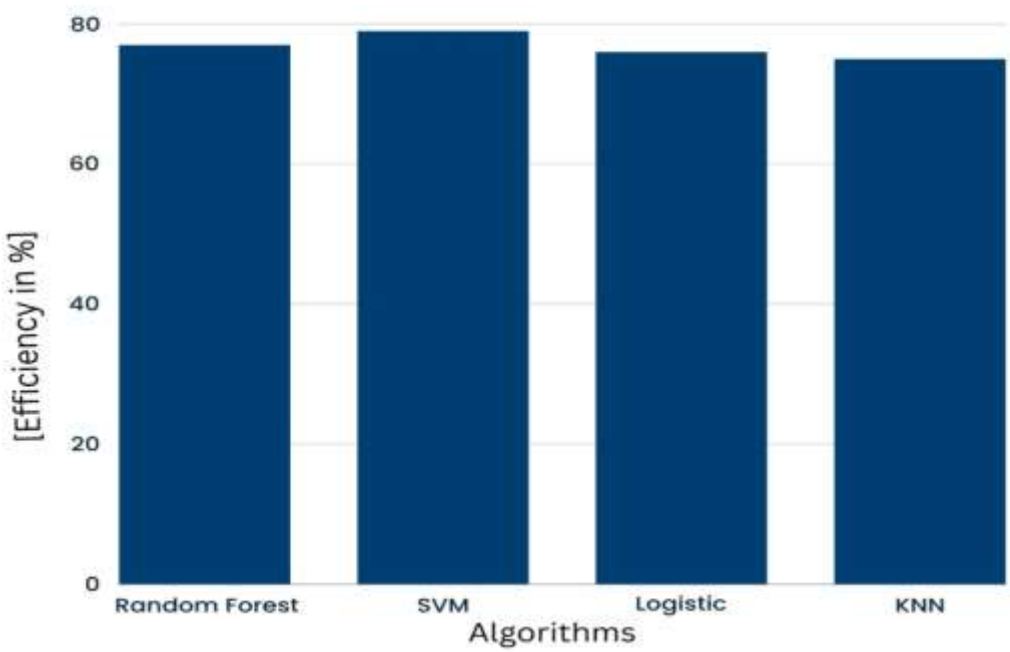


Fig -7.1

Accuracy Comparison:

Support vector machine(SVM) achieved the highest accuracy of 79%, followed by Random forest, K nearest neighbor, logistic regression and support vector machine.

Precision-Recall Analysis:

Support vector machine (SVM) exhibits higher precision and recall compared to other models, indicating its effectiveness in correctly identifying positive cases (patients with heart disease) while minimizing false positives.

F1 Score Comparison:

Support vector machine (SVM) achieved the highest F1 Score ,indicating a balanced performance in terms of precision and recall, followed by Random forest, K nearest neighbor, logistic regression and support vector machine.

Confusion Matrix:

Detailed breakdown of model predictions, highlighting SVM's better performance in accurately classifying both positive and negative cases.

This presentation format provides a structured overview of the heart disease prediction system's performance, highlighting the strengths of SVM while also acknowledging areas for future research and improvement.

7.2 Performance Evaluation

In evaluating the performance of a heart disease prediction system where Support Vector Machine (SVM) achieved the highest accuracy among Logistic Regression (LR), SVM, Random Forest (RF), and K-Nearest Neighbors (KNN), several key metrics and techniques are employed:

Accuracy: Accuracy measures the proportion of correctly classified instances over the total number of instances in the dataset. It provides an overall assessment of the model's performance and is particularly useful when the classes are balanced.

Precision and Recall: Precision represents the ratio of correctly predicted positive observations to the total predicted positives, while recall measures the ratio of correctly predicted positive observations to the all observations in actual class. These metrics are especially relevant in medical applications like heart disease prediction, where correctly identifying individuals at risk (high recall) while minimizing false alarms (high precision) is crucial.

F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balanced assessment of a classifier's performance. It's useful when there's an imbalance between the classes or when both false positives and false negatives are important.

Confusion Matrix: A confusion matrix provides a detailed breakdown of the classifier's predictions, showing the number of true positives, true negatives, false positives, and false negatives. It's valuable for understanding where the classifier is making errors and which classes are being confused.

7.3 Key Findings

The key finding of a heart disease prediction system indicating the highest accuracy for Support Vector Machine (SVM) among Logistic Regression (LR), SVM, Random Forest (RF), and K-Nearest Neighbors (KNN) could suggest that SVM is the most effective classifier for this particular dataset and problem. SVM is known for its ability to handle high-dimensional data well and find complex decision boundaries. Its performance might outshine the others due to its capability to separate classes efficiently, especially in cases where the data might not be linearly separable.

Non-linearity: SVM can effectively handle non-linear decision boundaries through the use of kernel functions. In heart disease prediction, where relationships between risk factors and disease occurrence might be complex and non-linear, SVM's ability to capture these intricate patterns can be advantageous.

Robustness to overfitting: SVM tends to generalize well even with limited training data, making it less prone to overfitting compared to models like RF and KNN, which can be sensitive to noisy data.

Optimized margin: SVM aims to maximize the margin between classes, which can lead to better generalization performance by reducing the risk of misclassification. This optimization strategy can be particularly beneficial in scenarios where the distribution of classes is imbalanced or overlapping.

Effective with high-dimensional data: In heart disease prediction, datasets often contain numerous features (e.g., patient demographics, medical history, lab results). SVM's effectiveness in high-dimensional spaces allows it to efficiently handle such datasets without sacrificing performance.

Tuning parameters: SVM offers flexibility in parameter tuning, such as the choice of kernel type and regularization parameters. Properly tuning these parameters can significantly impact the model's performance, potentially leading to superior accuracy compared to LR, RF, and KNN.

Handling outliers: SVM is less sensitive to outliers compared to LR and KNN. Given that outliers can significantly affect the performance of predictive models, especially in medical datasets where data quality might vary, SVM's robustness to outliers can be advantageous.

Interpretability: While SVM models themselves are not as interpretable as LR, they can still provide insights into feature importance through examining support vectors. This can aid in understanding the underlying factors contributing to heart disease prediction.

7.4 Snapshots of System with Brief Detail of Each

Sign In/Sign Up page

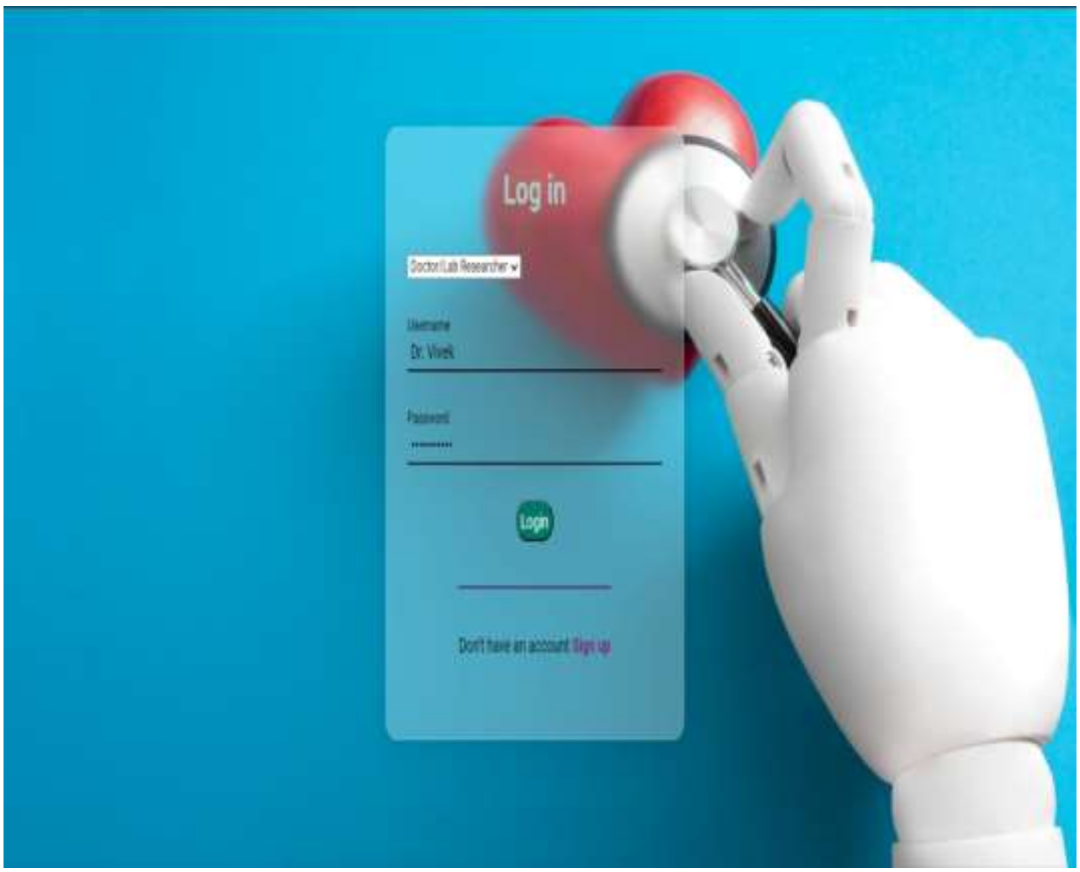
The system's Sign In/Sign Up page provides a secure entry point for users. For existing users, the Sign In section prompts entry of credentials for personalized access. New users can easily navigate to the Sign-Up section, where they input necessary information to create an account. This initial interaction ensures user authentication and sets the stage for a personalized and secure heart disease prediction experience.

Sign-Up: Separate forms for patients and doctors, including fields for username, password, and user type selection.

Login: Single form with fields for username, password, and user type selection.

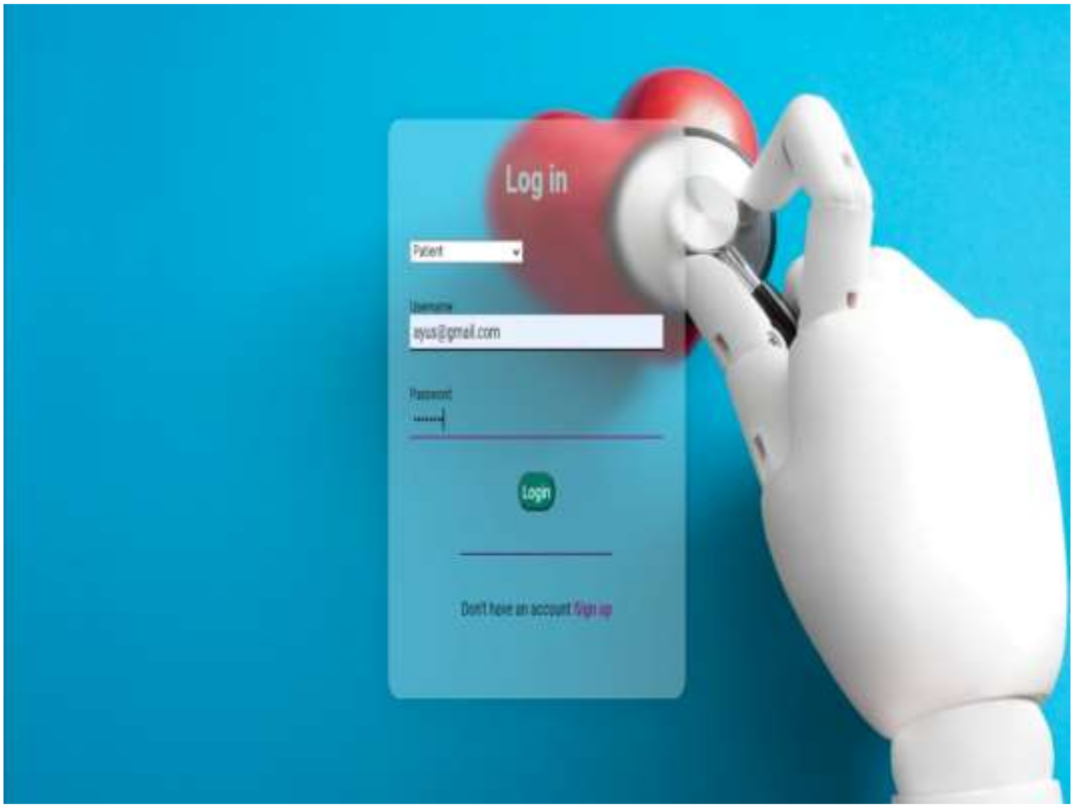
Here the user can choose whether he want to login as patient or doctor/lab researcher.

At this user choos as doctor/lab researcher for login and then go through it.



Login/Sign up for Doctor/Lab Researcher

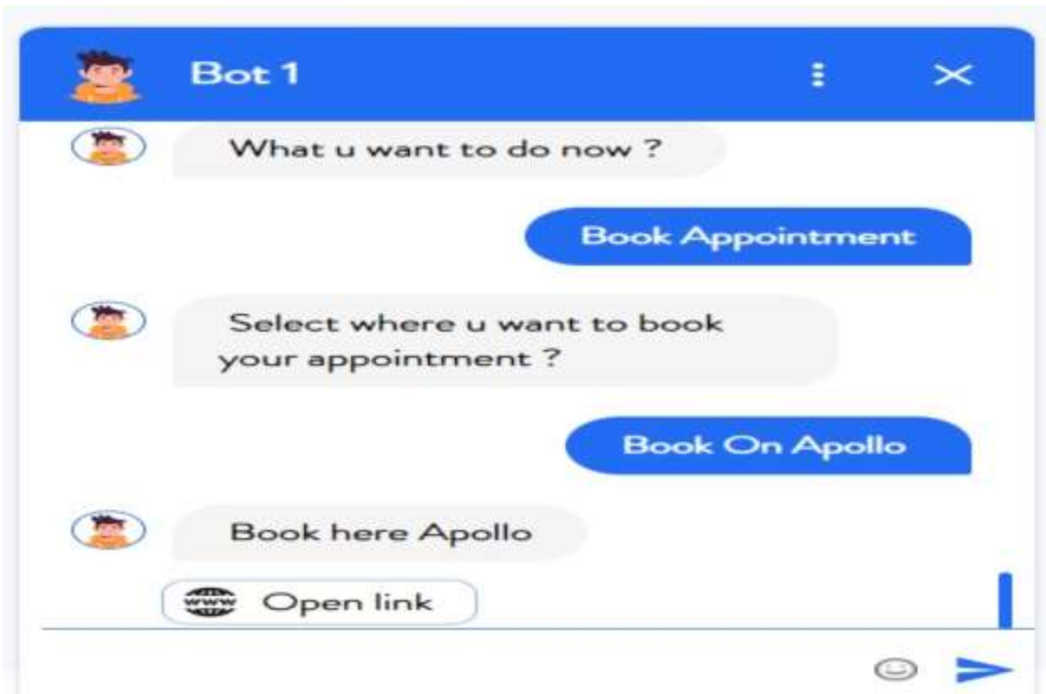
At this user choos as patient for login and then go through it.



Login/Sign up for Patient

ChatBot

The chatbot's primary functions include interacting with users to collect personal information, facilitate booking appointments, and provide heart disease risk predictions. This implementation aims to improve user engagement, streamline appointment scheduling, and offer accessible health predictions.

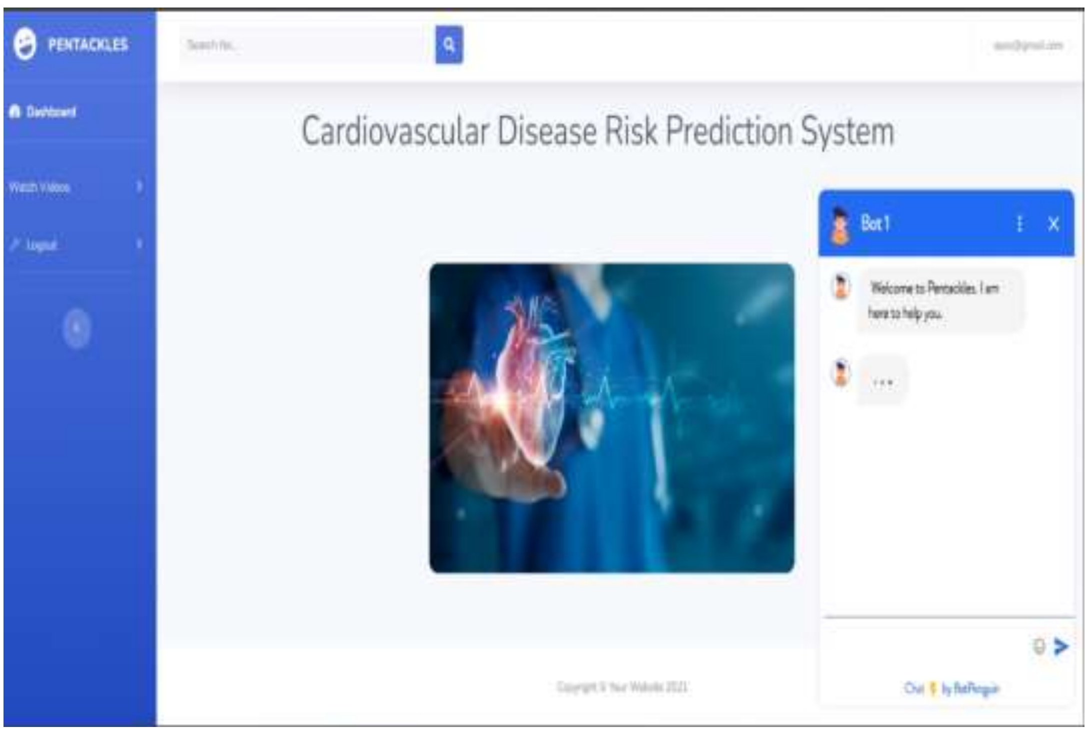


Chatbot

Dashboard

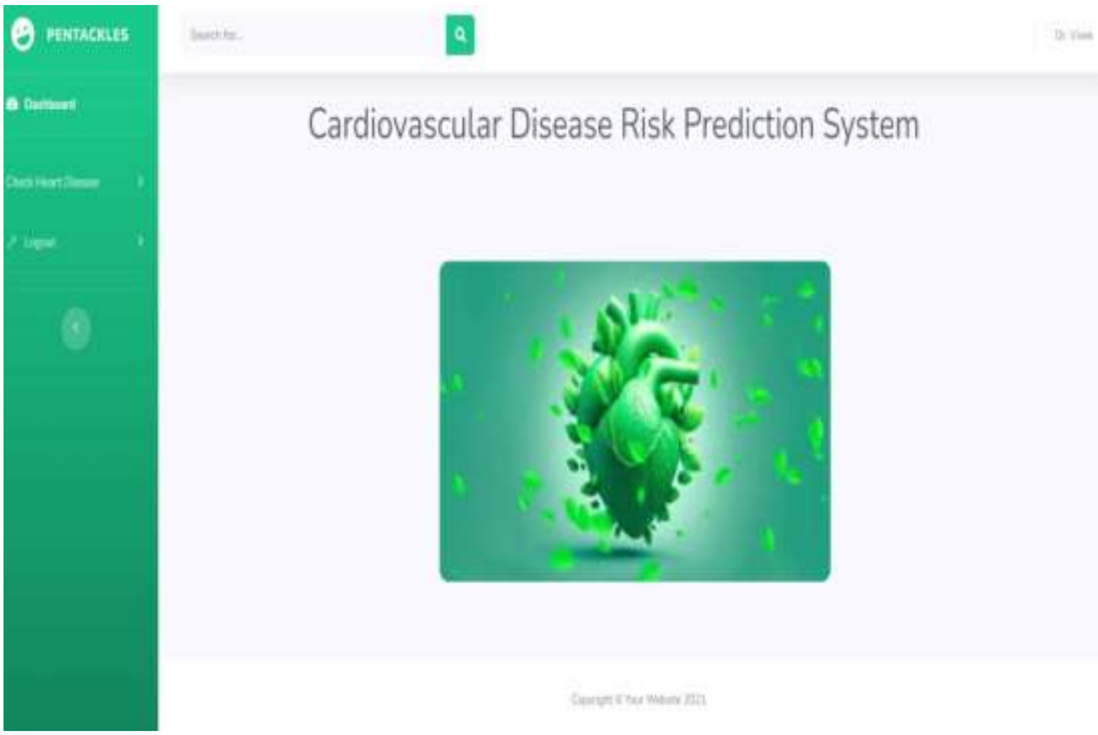
The system's dashboard presents a user-friendly interface designed for seamless interaction and easy navigation. At the right of the dashboard is a responsive chatbot, ready to assist users with queries or provide additional information. Users can engage in natural language conversations to gather insights about their heart health. On the main dashboard the application provided a chatbot for the help of user so that he/she can book in appointment in hospital for their checkup.

This is the dashboard for user as patient where he/she can tackle through the website.



Dashboard for patient

This is the dashboard for user as doctor/lab researcher where he/she can tackle through the website.

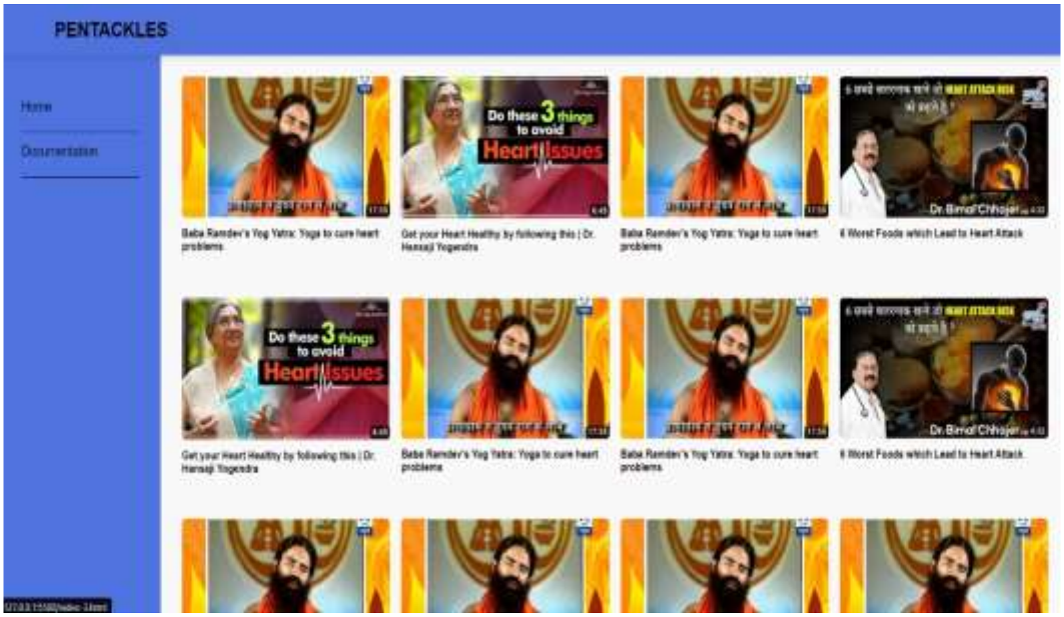


Dashboard for Doctor/Lab Researcher

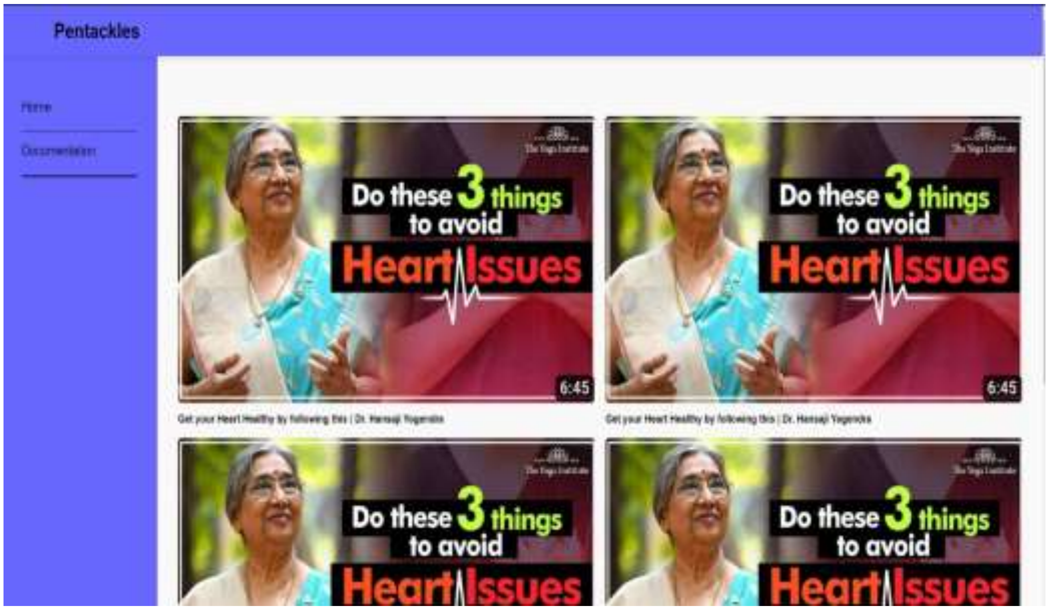
Watch Video Section

Heart Health Education The "Watch Video" section on the dashboard is a dedicated space where users can access informative and engaging videos related to heart health. This section is designed to provide valuable insights into heart exercises, dietary recommendations, and do's and don'ts for maintaining a healthy cardiovascular system. Here's a breakdown of the content:

Videos Section



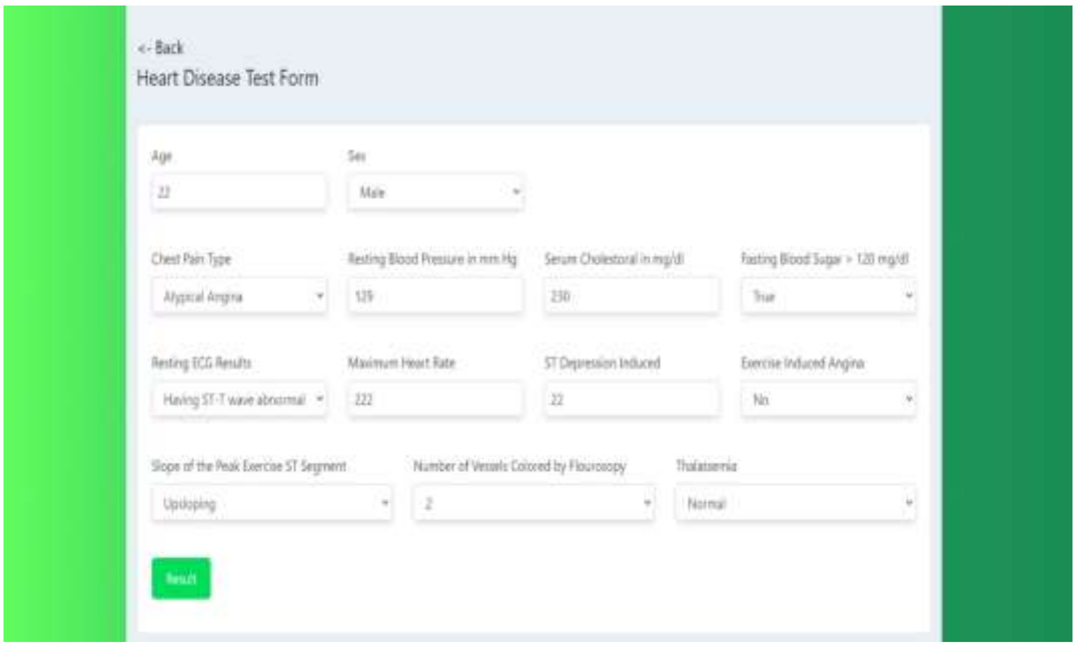
Videos Section



Videos Section

Heart Disease Prediction Section

Prediction of heart disease is done in this section of the application where user which can be lab operator or doctor uses the report of patient and enter the values of parameters asked in the heart disease test form and according to the values in the parameters the result is predicted whether the patient is likely to be a heart patient or not.



Form for heart disease prediction

CHAPTER 8

CONCLUSION AND FUTURE SCOPE

Conclusion

Cardiovascular diseases (CVDs) which also called heart diseases are the most life-threatening disease nowadays which develops tragical compilations such as heart attack. In contrast to which we aimed to build a Machine learning system to predict cardiovascular disease. We have studied the data sets and throughout the process of developing the cardiovascular disease prediction system, various machine learning algorithms were explored and estimated such as Random Forest, K nearest neighbor, logistic regression and support vector machine, leading to the selection of the most suitable models based on their accuracy, reliability, and specificity.

Future Scope

In future studies, we will collect more samples, data sets and try to add more parameters to find out the prediction fastly and accurately we also try to optimize models and build deep learning-based models to more accurately predict the development of CVD, as well as treatment and prevention recommendations.

REFERENCE

[1] A. S. D. M. ,. S. Mandakini Priyadarshani Behera, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with," Procedia Computer Science, 2022.

[2] M. SeYongJang, "Machine Learning Model Prediction of Mortality in Patients With and Without Heart Failure," JACC: Advances, vol. 2, 2023.

[3] A. L. P. B. P. V. R. K. Alireza Ghasemieh, "A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease Patients," Decision Analytics Journal, 2023.

[4] 1. ,. Y. R. a. ,. H. Y. a. ,. Y. D. b. ,. Y. L. a. ,. Y. Y. a. ,. A. M. a. ,. T. Y. d. ,. Y. W. c. ,. F. X. c. ,. Q. H. e. Y. Z. a. Yue Huang a, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," Computers in Biology and Medicine, 2022.

[5] C.-k. T. b. C.-J. W. b. S.-H. C. b. Y.-S. H. b. L.-J. K. b. Chun-Ling Lin a, "Title-Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome," International Measurement Confederation , 2023.

[6] V. S. A.-D. b. ,. N. M. c. ,. A. A. H. d. ,. I. M. a. ,. S. A. e. ,. K. F. A.-t. f. ,. K. A. g. ,. F. M. B. g. ,. M. A. M. Md. Mamun Ali a, "Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure Patients," Healthcare Analytics, 2023.

[7] L. V. R. G. Stojanov, ": Machine Learning-Based Predictive Models for Heart Disease Diagnosis," King Saud University, 2023.

[8] S. M. P. K. P. S. M. Subasish Mohapatra, "Heart Diseases Prediction based on Stacking Classifiers Model," International Conference on Machine Learning and Data Engineering, 2023.

[9] A. C. A. R. W. E. E. Qi Li, "Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data," International Journal of Medical Informatics, 2022.

[10] R. R. N. Y. N. B. S. K. D. Khandaker Mohammad Mohi Uddin, "Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset," Intelligence-Based Medicine, 2023.

Github Link-

https://github.com/KIET-Github/CS-2024-C/tree/main/PCS24-8-Aakriti

RESEARCH PAPER ACCEPTANCE PROOF

Submission Summary

Conference Name	INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS AND COMPUTING TECHNOLOGIES
Track Name	Track-3 (Computing Technologies)
Paper ID	124
Paper Title	Towards Proactive Cardiovascular Health: Machine Learning Models for Early Detection of Heart Disease
Abstract	Cardiovascular diseases (CVDs), also known as heart disease, has replaced other diseases like malaria, strokes, cancer, trachea, bronchus which was once considered the most significant diseases worldwide, as the number one killer globally, notably this is true for India as well as other countries. This means that such illnesses should be identified early enough for effective treatment to be undertaken hence there should be such as reliable and appropriate system. Using machine learning on various medical datasets, these approaches have automated large scale and complex data analysis. For the advancement in growth of healthcare our primary goal of this study work is to determine which patient, depending on several medical parameters, is more likely to develop a cardiac condition. To determine whether the patient is likely to receive a heart disease diagnosis or not, we developed a heart disease prediction algorithm. use the patient's medical history. We employed many machine learning algorithms, including KNN and logistic regression, support vector machine, random forest algorithm are used to categorize and predict heart disease patients. Quite beneficial method was employed to control the model's application in order to increase the forecast accuracy of a heart attack in any person. The suggested model's strength was rather pleasing; it could identify signs of heart illness in a specific person using support vector machines, which demonstrated high accuracy when compared to earlier classifiers like KNN and logistic regression, etc. Thus, by utilizing the provided model to determine the likelihood that the classifier can correctly and precisely diagnose cardiac illness, a considerable amount of pressure has been released. The Given heart disease prediction system lowers costs and improves medical treatment. This research provides us with important information that may be used to forecast the people who will have heart disease.
Created	3/9/2024, 12:45:26 PM
Last Modified	3/9/2024, 12:45:26 PM
Authors	**Aakriti Gupta** (KIET Group of Institutions) <aakritig31101@gmail.com> Aakansha Tyagi (KIET Group of Institutions) <tyagiaakansha99270@gmail.com> Ayush Jaiswal (KIET Group of Institutions) <ayushjai1267@gmail.com> Ayush Tyagi (KIET Group of Institutions) <ayushtyagi4810poco@gmail.com> Vivek Sharma (KIET Group of Institutions) <vivek.sharma@kiet.edu>
Submission Files	Towards Proactive Cardiovascular Health Machine Learning Models for Early Detection of Heart Disease.docx (751.3 Kb, 3/9/2024, 12:43:14 PM)