

MULTIPLE DISEASE PREDICTION

A PROJECT REPORT

Submitted to

Visvesvaraya Technological University

BELAGAVI - 590 018

by

Punith L **4SU21CS071**

Sannidhi Rai **4SU21CS089**

Susha Jain **4SU21CS107**

Pradnyan Hegde **4SU22CS405**

Under the guidance of

Mr. Pradeep G S

Assistant Professor

in partial fulfillment of the requirements for the award of the degree of

Bachelor of Engineering



Department of Computer Science & Engineering

SDM INSTITUTE OF TECHNOLOGY

UJIRE - 574 240

2024-2025

SDM INSTITUTE OF TECHNOLOGY

(Affiliated to Visvesvaraya Technological University, Belagavi)

UJIRE – 574 240

Department of Computer Science and Engineering

CERTIFICATE

Certified that the Project Work titled ‘**Multiple Disease Prediction**’ is carried out by **Mr. Punith L**, USN: **4SU21CS071**, **Ms. Sannidhi Rai**, USN: **4SU21CS089**, **Ms. Susha Jain**, USN: **4SU21CS107**, **Mr. Pradnyan Hegde**, USN: **4SU22CS405**, bonafide students of SDM Institute of Technology, Ujire, in partial fulfillment for the award of the degree of **Bachelor of Engineering** in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2024-2025. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said Degree.

Mr. Pradeep G S
Asst. Professor and Guide

Dr. Thyagaraju G S
Professor and Head

Dr. Ashok Kumar T
Principal

Signature with date and seal:

External Viva

Name of the Examiners:

Signature with Date

1.

2.

Acknowledgement

It is our pleasure to express our heartfelt thanks to Mr. Pradeep G S, Assistant Professor, Department of Computer Science and Engineering, for his supervision and guidance which enabled us to understand and develop this project.

We are indebted to Dr. Thyagaraju G S, Head of the Department and Dr. Ashok Kumar T, Principal, for their advice and suggestions at various stages of the work. We also extend our heartfelt gratitude to Mr. Pradeep Rao, Project Coordinator for his assistance and the management of SDM Institute of Technology, Ujire, for providing us with a good learning environment, library and laboratory facilities.

Lastly, we take this opportunity to offer our regards to all of those who have supported us directly or indirectly in the successful completion of this project work.

Punith L
Sannidhi Rai
Susha Jain
Pradnyan Hegde

Abstract

The multiple disease prediction system aims to improve early diagnosis by utilizing machine learning techniques to predict various diseases based on patient data. It incorporates patient history, symptoms, and medical records to identify potential health risks. By analyzing large datasets, the system can predict diseases like diabetes, heart disease, Parkinson's, breast cancer and liver disease. The model uses algorithms such as random forest, logistic regression and support vector machines for classification. Data preprocessing techniques, such as normalization and missing value handling, are crucial for accurate predictions. The system provides a cost-effective, time-efficient approach to healthcare, allowing for early interventions and better patient outcomes. The integration of real-time data enhances prediction accuracy. Furthermore, continuous updates to the model help adapt to evolving medical trends. The system supports healthcare professionals by offering decision support, reducing human error. This multi-disease prediction tool has the potential to revolutionize preventive healthcare practices, offering personalized treatment options.

Table of Contents

	Page No.
Abstract	i
Table of Contents	ii
List of Figures	iv
List of Tables	v
Chapter 1 Introduction	1
1.1 Project Introduction	1
1.2 Problem Description	2
Chapter 2 Literature Review	3
2.1 Literature Survey	3
2.2 Comparative Analysis of Related Work	6
2.3 Summary	6
Chapter 3 Problem Formulation	7
3.1 Problem Statement	7
3.2 Objectives of the Present Study	8
3.3 Summary	8
Chapter 4 Requirements and Methodology	9
4.1 Hardware Requirements	9
4.2 Software Requirements	9
4.3 Methodology Used	10
Chapter 5 System Design	11
5.1 Architecture of the Proposed System	11
5.2 System Flowchart	12
Chapter 6 Implementation	13
6.1 Pseudocode	13
Chapter 7 System Testing, Results and Discussion	20
7.1 System Testing	20
7.2 Result Analysis	24
7.3 Summary	32

Chapter 8	Conclusion and Scope for Future Work	33
8.1	Conclusion	33
8.2	Scope for Future work	34

References

Personal Profile

List of Figures

	Page No.
Figure 5.1 Architecture of the Proposed System	11
Figure 5.2 Flowchart of the Proposed System	12
Figure 7.1 Graph analysis of the first set	25
Figure 7.2 Graph analysis of the second set	26
Figure 7.3 Graph analysis of the third set	26
Figure 7.4 Graph analysis of the fourth set	27
Figure 7.5 Graph analysis of the fifth set	27
Figure 7.6 Home page	28
Figure 7.7 About Us page	28
Figure 7.8 Sign Up page	28
Figure 7.9 Login page	29
Figure 7.10 Diabetes Prediction Page	29
Figure 7.11 Heart Disease Prediction Page	30
Figure 7.12 Parkinson's Prediction Page	30
Figure 7.13 Breast Cancer Prediction Page	31
Figure 7.14 Liver Disease Prediction Page	31

List of Tables

	Page No.
Table 2.1 Comparative Analysis	6
Table 4.1 Hardware Requirements	9
Table 4.2 Software Requirements	9
Table 7.1 Unit Test Cases	20
Table 7.2 Analysis of Diabetes using three algorithms	24
Table 7.3 Analysis of heart disease using three algorithms	24
Table 7.4 Analysis of the Parkinson's using three algorithms	24
Table 7.5 Analysis of the Breast Cancer using three algorithms	25
Table 7.6 Analysis of the Liver Disease using three algorithms	25

Introduction

1.1 Project Introduction

Multiple disease prediction is an emerging field that leverages machine learning and data analytics to identify the likelihood of various diseases in individuals. By analyzing a combination of clinical data, lifestyle factors, and patient history, predictive models can simultaneously forecast the risk of diseases such as heart disease, diabetes, Parkinson's disease, and breast cancer. These advanced systems examine patterns in health data, enabling early detection and timely interventions, which are crucial for improving patient outcomes. For example, predicting heart disease risk based on factors like cholesterol levels, or identifying early signs of Parkinson's through movement data, allows for more targeted and efficient treatments.

The integration of predictive models into healthcare systems has the potential to revolutionize the way we approach disease management. For instance, diabetes can be detected early through the analysis of blood sugar levels, and breast cancer can be predicted through mammogram images and genetic markers. Early identification of these diseases, especially those that progress silently, can greatly reduce the risk of complications and improve survival rates. With advancements in technology, these models can process vast amounts of data, making healthcare more proactive rather than reactive. This approach not only benefits individual patients but also supports healthcare providers in offering more personalized and effective treatment plans.

1.2 Problem Description

The problem of predicting multiple diseases, such as heart disease, diabetes, Parkinson's disease, and breast cancer, arises from the challenges of early diagnosis and the complexity of recognizing multiple conditions at once. Many of these diseases develop gradually and may not show clear symptoms until they are in advanced stages. As a result, patients often face serious health complications that could have been prevented with earlier detection. Traditional diagnostic methods rely heavily on individual symptoms and tests, which can sometimes be inaccurate or missed. Additionally, healthcare systems often struggle with managing the large volume of patient data and making sense of it to predict various diseases simultaneously. This creates a need for more efficient and accurate tools to assess a patient's risk for multiple diseases at once. Without such tools, patients may receive delayed treatments, or in some cases, incorrect diagnoses. A multi-disease prediction model could help by analyzing a patient's medical history, lab results, genetic information, and lifestyle factors to provide more comprehensive and timely insights. However, designing these predictive models comes with challenges, such as data privacy concerns, the complexity of integrating multiple health datasets, and ensuring that predictions are reliable and understandable to healthcare providers.

Chapter 2

Literature Review

2.1 Literature Survey

In the paper titled "Multimodal Machine Learning Approaches for Disease Prediction: A Survey" [1] (2022) by Sharma et al. explores the growing use of multimodal machine learning (ML) in healthcare, focusing on its ability to integrate diverse data sources such as clinical records, imaging, genetic profiles, and wearable sensor data to enhance disease prediction. The study reviews various ML algorithms, including deep learning architectures like CNNs, RNNs, and transformers, highlighting their applications in detecting and managing diseases such as cancer, cardiovascular conditions, and neurological disorders. While showcasing the advantages of multimodal ML in improving diagnostic accuracy and patient outcomes, the authors address challenges like data heterogeneity, integration complexity, and privacy concerns. They emphasize the need for standardized datasets, ethical considerations, and explainable AI to build reliable and generalizable systems. The paper concludes by advocating for advancements in real-time predictive models using IoT and wearable technologies to further revolutionize healthcare.

In the paper titled "Multiple Disease Prediction Using Machine Learning Algorithms" [2] by Chauhan et al. (2021) explores the application of various machine learning techniques, such as Support Vector Machines (SVM) and Decision Trees, for predicting multiple diseases using symptoms as input data. Focusing on conditions like heart disease and diabetes, the study evaluates the performance of these algorithms. It underscores the potential of predictive analytics to support healthcare professionals in making timely decisions, addressing challenges such as the scarcity of medical infrastructure and the low doctor-to-patient ratio. The authors highlight the importance of early detection in mitigating the impact of serious diseases and saving lives. Furthermore, the research integrates predictions for multiple diseases into a unified user interface, emphasizing its practical utility. This work was conducted by Indukuri Mohit, K. Santhosh Kumar, Avula Uday Kumar Reddy, and Badhagouni Suresh Kumar from Vardhaman College of Engineering, Hyderabad, India.

In the paper titled "Symptoms Based Multiple Disease Prediction Model Using Machine Learning Approach" [3] (2021), Kolli et al. investigate the application of machine learning

(ML) techniques to predict multiple diseases based on patient symptoms. The paper focuses on addressing the challenges of timely disease detection and diagnosis, particularly in resource-constrained settings with limited healthcare infrastructure. The authors evaluate the effectiveness of various ML algorithms, including decision trees, support vector machines (SVM), and k-nearest neighbors (KNN), for symptom-based disease prediction. The research emphasizes the role of feature selection and preprocessing techniques in enhancing prediction accuracy and reducing computational complexity. The study also highlights the integration of user-friendly interfaces to provide accessible healthcare solutions, enabling users to input symptoms and receive predictions. Additionally, the authors explore the scalability of such models for predicting a range of diseases, including diabetes, heart disease, and respiratory conditions. While the findings demonstrate the potential of ML in improving diagnostic precision, the paper also identifies limitations such as reliance on the quality of input data and challenges in handling overlapping symptoms across diseases. Kolli et al. underscore the need for future work on improving data integration, model interpretability, and real-time deployment to enhance the practical utility of ML-driven disease prediction systems.

In the paper titled "A Comparative Study of Machine Learning Algorithms for Disease Prediction" [4] (2021), Gupta et al. analyse and compare the performance of various machine learning (ML) algorithms in predicting diseases based on clinical and diagnostic data. The study focuses on commonly used algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forest, Logistic Regression, and k-Nearest Neighbors (KNN), assessing their strengths and limitations in terms of accuracy, computational efficiency, and scalability. The authors emphasize the importance of selecting appropriate algorithms based on the nature of the dataset and disease characteristics. For instance, Random Forest demonstrated superior performance for complex datasets with high dimensionality, while Logistic Regression excelled in binary classification tasks. The study also highlights the role of preprocessing techniques, including feature scaling and dimensionality reduction, in improving model performance. Additionally, the authors explore the challenges of overfitting, imbalanced datasets, and the need for explainability in medical applications. Gupta et al. conclude by recommending ensemble learning approaches and hybrid models to leverage the strengths of individual algorithms. The paper underscores the necessity of optimizing ML techniques for real-world healthcare scenarios, advocating for further research on integrating diverse data types and improving the interpretability of predictive models to enhance their acceptance and reliability in clinical practice.

In the paper titled “Predictive Analytics for Disease Prevention and Diagnosis: A Machine Learning Approach ” [5] (2021), Bansal, et al highlights the transformative role of machine learning (ML) in healthcare by enabling early detection, accurate diagnosis, and personalized treatment of diseases. The authors review various ML techniques, including decision trees, support vector machines, neural networks, and ensemble methods, applied to predict chronic diseases like diabetes, cardiovascular conditions, and cancer, as well as neurodegenerative and infectious diseases. The survey emphasizes challenges such as data quality, privacy concerns, and the need for interpretability in ML models for clinical adoption. They also discuss tools and frameworks like Scikit-learn and TensorFlow used in healthcare analytics and underscore the importance of addressing data imbalance, ethical issues, and interdisciplinary collaboration to advance the field. This comprehensive review provides insights into current research trends and highlights opportunities for developing robust, user-friendly predictive systems in healthcare.

2.2 Comparative Analysis of the Related Work

The table 2.1 discusses the comparative analysis of the current systems in light of the suggested proposal.

Table 2.1: Comparative Analysis

Sl. No	Author(s)	Algorithms/Techniques	Performance Measures
1.	Sharma et al	Deep learning architectures like CNNs, RNNs, and transformers	Accuracy
2.	Chauhan et al	SVM and Decision Trees	Accuracy
3.	Kolli et al	Random Forest, Decision Trees, and LightGBM	Accuracy
4.	Gupta et al	Support Vector Machines (SVM), Decision Trees, Random Forest, Logistic Regression, and k-Nearest Neighbors (KNN)	Accuracy
5.	Bansal, et al	K-Nearest Neighbors and Fuzzy K-NN approaches	Accuracy

2.3 Summary

These were the research papers that we studied to gain a better understanding of the problem. Machine learning classification algorithms are more accurate compared to the traditional techniques when it comes to detecting breast cancer. Hence, we chose to analyze Logistic Regression, SVM and Random Forest algorithms and implement the prediction model using the best algorithm in the project.

Problem Formulation

3.1 Problem Statement

Most machine learning models are limited to analysing a single disease, such as liver disease, cancer, or lung disease, which restricts their scope. Additionally, users often must navigate multiple websites or platforms to predict different diseases, resulting in a fragmented and inefficient experience. There is a notable absence of a unified system capable of predicting multiple diseases simultaneously, highlighting the need for an integrated solution.

In 2024, the global healthcare community faced the growing challenge of accurately predicting multiple diseases such as heart disease, diabetes, Parkinson's, and breast cancer, with millions of cases going undiagnosed in their early stages. For instance, heart disease remains one of the leading causes of death worldwide, while diabetes affects over 460 million people globally. In many cases, these diseases are detected only after significant damage has already been done, leading to higher mortality rates. Traditional diagnostic methods, including blood tests and imaging, often fail to detect diseases early enough for effective intervention. However, advances in Data Science and Machine Learning offer a promising solution. By analyzing vast datasets, including medical histories, lab results, and genetic information, machine learning algorithms can predict the likelihood of multiple diseases, allowing for earlier diagnosis and personalized treatment plans. The goal is to create a robust model that accurately predicts multiple diseases, improving healthcare outcomes and reducing the overall burden on healthcare systems.

3.2 Objectives of the Present Study

The objectives of the proposed project are as follows:

1. To train the data set using the ML Classification algorithms, namely Logistic Regression and Support Vector Machine Algorithm.
2. To calculate and compare the accuracy of each model.
3. To find out the best model for early detection of Multiple Diseases.
4. To create a web interactive page for doctors for the detection of Disease.
5. To perform real-time analysis

3.3 Summary

Machine learning models typically focus on single-disease analysis, requiring users to rely on multiple platforms for different predictions. This fragmented approach lacks a unified system capable of predicting multiple diseases simultaneously, emphasizing the need for an integrated solution. The best solution for early detection of Multiple diseases is using machine learning techniques. The classification algorithms are more accurate when compared to traditional imaging techniques. Developing an early detection system can be useful for many doctors as well as patients by helping them be alert and take the required medications to prevent the spread of Diseases. In 2024, the healthcare sector faces significant challenges in accurately predicting and diagnosing diseases like heart disease, diabetes, Parkinson's, and breast cancer at an early stage, with many cases going undetected until they are advanced. This delay results in higher mortality rates and increased pressure on healthcare systems, as traditional diagnostic methods such as blood tests and imaging often fail to provide timely detection. Advances in Data Science and Machine Learning, however, present a promising solution by leveraging large datasets, including medical histories and lab results, to predict disease likelihood and enable early intervention. The proposed project aims to address these challenges by training datasets using classification algorithms like Logistic Regression, Random Forest and Support Vector Machine, evaluating and comparing their accuracies to identify the most effective model for early disease detection. Additionally, the project seeks to create a web-based interactive platform for doctors to aid in diagnosis and facilitate real-time disease analysis, ultimately improving healthcare outcomes and reducing the burden on healthcare systems.

Chapter 4

Requirements and Methodology

4.1 Hardware Requirements

The hardware requirements for the proposed project are depicted in Table 4.1.

Table 4.1: Hardware Requirements

Sl. No	Hardware/Equipment	Specification
1.	Graphics Card	Intel 621 Graphics card or 2GB
2.	RAM	4GB or above

4.2 Software Requirements

The software requirements for the proposed project are depicted in Table 4.2.

Table 4.2: Software Requirements

Sl. No	Software	Specification
1.	Anaconda	Anaconda 64 bit
2.	Spyder	Version 5.4.3
3.	Framework	Streamlit
4.	Google Collab	Cloud-based Python Notebook
5.	scikit-learn	Version 1.3.0

Methodology Used

The proposed multiple disease prediction system is implemented using the following steps:

- 1) **Data collection:** Gather data related to patients, such as demographic information (age, gender), medical history, clinical test results (blood pressure, cholesterol) and genetic data.
- 2) **Data pre-processing:** Then the next step is data pre-processing. It is the process of converting raw data into clean data. In this process missing values, noisy and inconsistent data in the dataset are handled.
- 3) **Train Test split:** Here data is split for training and testing purposes. 70% of the data is used to train the model and the remaining 30% of the data is used for testing. Then dataset will be trained using Support Vector Machine classifies and Logistic Regression.
- 4) **Modelling:** After the training and testing of the models, confusion matrix is plotted, and accuracy score is computed for each algorithm. Then based on the accuracy score best suited algorithm for the prediction of disease is identified.

System Design

5.1 Architecture of the Proposed System

Figure 5.1 shows the architecture of the proposed system.

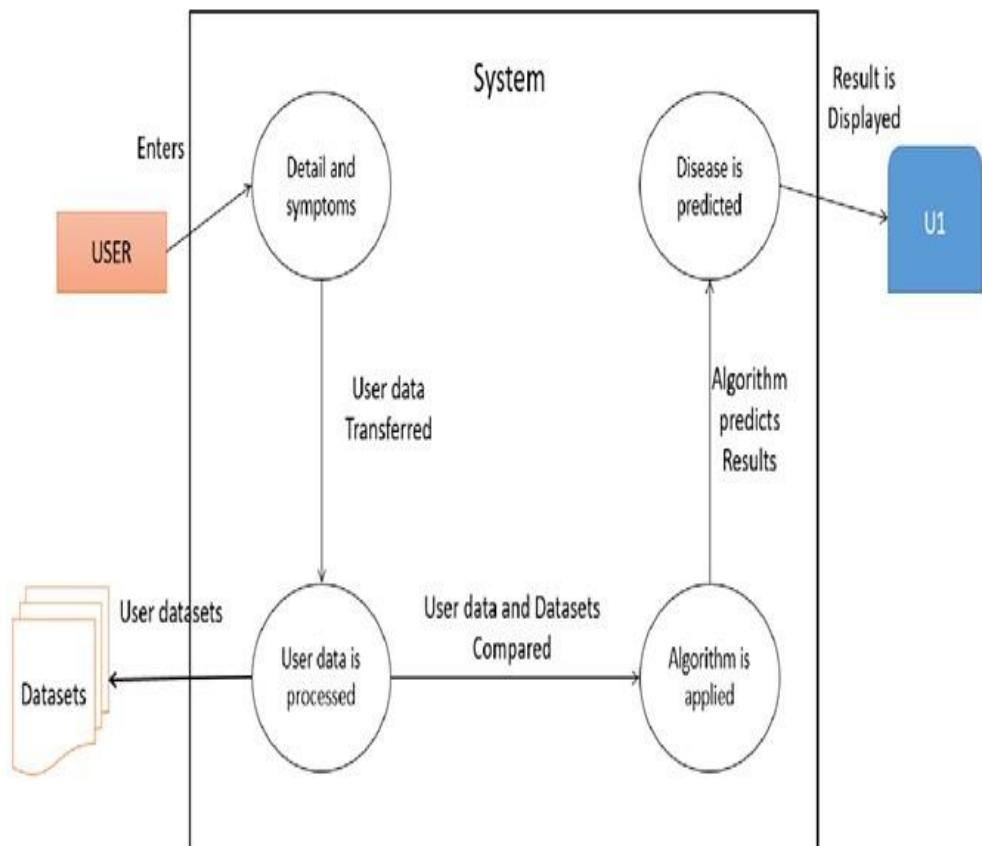


Figure 5.1: Architecture of the Proposed System

Figure 5.1 is an architecture depicting a system for predicting diseases based on user data and datasets. The process starts with the user entering details and symptoms into the system. The user data is then transferred and processed, where it is compared with existing datasets. An algorithm is applied to predict the disease, and the results are displayed to the user.

5.2 System Flowchart

A system flowchart is a way of depicting how data flows in a system and how decisions are made to control events. Figure 5.2 depicts the system flowchart.

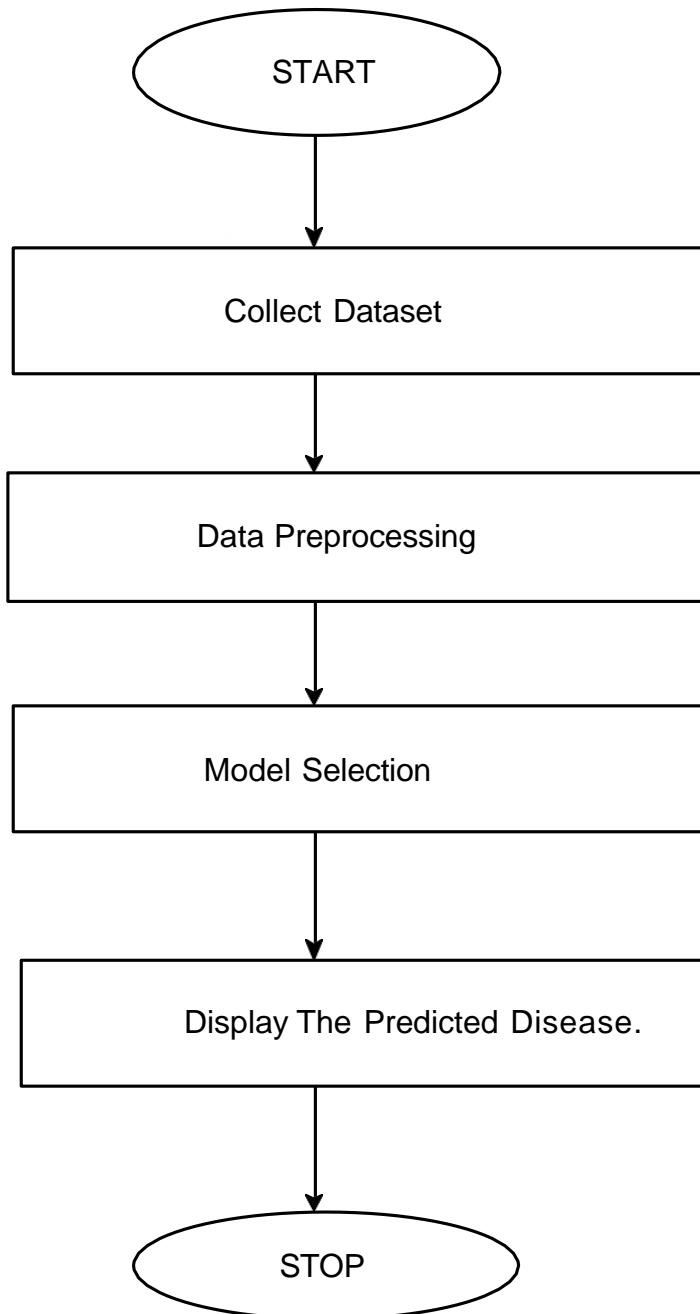


Figure 5.2: System Flowchart

The raw dataset must be loaded, cleaned, and preprocessed. The data frame is created with the selected features. The prediction model is created using SVM, Logistic Regression and Random Forest. This model classifies the data into positive and negative results.

Implementation

6.1 Pseudocode

Steps Overview

1. Environment Setup: Import necessary libraries like os, pickle, and streamlit.
2. Load Pre-trained Models: Use pickle to load machine learning models for diseases (diabetes, heart disease, Parkinson's, breast cancer and liver disease).
3. Configure Streamlit: Set up the Streamlit web application, including custom CSS for styling.
4. Implement Authentication: Handle user login, account creation, and logout functionality.
5. Create Navigation Menu: Use option_menu to allow users to navigate between home, login, and disease prediction pages.
6. Disease Prediction Pages:
 - Collect user inputs for each disease's required features.
 - Process inputs and use the corresponding ML model to make predictions.
 - Display the result (e.g., whether the user is likely to have a specific disease).
7. Result Presentation: Show predictions alongside helpful descriptions of input features.
8. Real-Time Interaction: Allow for seamless interaction and model predictions directly on the web app.

```
import os
import pickle
import streamlit as st
from streamlit_option_menu import option_menu

# Load models
```

```

working_dir = os.path.dirname(os.path.abspath(__file__))
diabetes_model = pickle.load(open(f'{working_dir}/model/diabetes_model.sav', 'rb'))
heart_disease_model = pickle.load(open(f'{working_dir}/model/heart_disease_model.sav',
'rb'))
parkinsons_model = pickle.load(open(f'{working_dir}/model/parkinsons_model.sav', 'rb'))
breast_cancer_model = pickle.load(open(f'{working_dir}/model/breast_cancer_model.sav',
'rb'))
liver_model = pickle.load(open(f'{working_dir}/model/liver_model.sav', 'rb'))

# Streamlit page configuration
st.set_page_config(page_title="Disease Prediction System", layout="wide",
page_icon="🧬")

# Custom CSS for styling
def apply_custom_css():
    st.markdown("""
<style>
    body { background-color: #f8f9fa; font-family: Arial, sans-serif; }
    [data-testid="stSidebar"] { background-color: #343a40; }
    .stButton > button { background-color: #007bff; color: white; border-radius: 5px; }
    .stButton > button:hover { background-color: #0056b3; }
</style>
""", unsafe_allow_html=True)

apply_custom_css()

# Navigation menu
selected = option_menu(
    menu_title="Navigation",
    options=["Home", "Diabetes Prediction", "Heart Disease Prediction", "Parkinson's
Prediction", "Breast Cancer Prediction", "Liver Disease Prediction"],
    icons=["house", "activity", "heart", "person", "gender-female", "shield-plus"],
    menu_icon="cast",
    default_index=0,
    orientation="horizontal",
)

```

```

)

# Pages
if selected == "Home":
    st.title("Welcome to the Disease Prediction System")
    st.write("This system helps predict various diseases using machine learning models.")
    st.image("home_image.png", use_column_width=True)

elif selected == "Diabetes Prediction":
    st.title("Diabetes Prediction")
    col1, col2, col3 = st.columns(3)

    with col1: Pregnancies = st.text_input('Number of Pregnancies')
    with col2: Glucose = st.text_input('Glucose Level')
    with col3: BloodPressure = st.text_input('Blood Pressure value')
    with col1: SkinThickness = st.text_input('Skin Thickness value')
    with col2: Insulin = st.text_input('Insulin Level')
    with col3: BMI = st.text_input('BMI value')
    with col1: DiabetesPedigreeFunction = st.text_input('Diabetes Pedigree Function value')
    with col2: Age = st.text_input('Age')

if st.button('Get Diabetes Prediction'):
    try:
        user_input = [
            float(Pregnancies), float(Glucose), float(BloodPressure),
            float(SkinThickness), float(Insulin), float(BMI),
            float(DiabetesPedigreeFunction), float(Age)
        ]
        result = diabetes_model.predict([user_input])[0]
        if result == 0:
            st.success("Good news! You are not diabetic.")
        else:
            st.error("Unfortunately, you are diabetic. Please consult a doctor.")
    except ValueError:
        st.error("Please enter valid numeric inputs.")

```

```

elif selected == "Heart Disease Prediction":
    st.title("Heart Disease Prediction")
    col1, col2, col3 = st.columns(3)

    with col1: Age = st.text_input('Age')
    with col2: Sex = st.selectbox('Sex', options=["Male", "Female"])
    with col3: CP = st.text_input('Chest Pain Type')
    with col1: Trestbps = st.text_input('Resting Blood Pressure')
    with col2: Chol = st.text_input('Serum Cholestorol')
    with col3: Fbs = st.selectbox('Fasting Blood Sugar > 120 mg/dl', options=["Yes", "No"])
    with col1: Restecg = st.text_input('Resting Electrocardiographic Results')
    with col2: Thalach = st.text_input('Maximum Heart Rate Achieved')
    with col3: Exang = st.selectbox('Exercise Induced Angina', options=["Yes", "No"])

if st.button('Get Heart Disease Prediction'):
    try:
        user_input = [
            float(Age), 1 if Sex == "Male" else 0, float(CP), float(Trestbps),
            float(Chol), 1 if Fbs == "Yes" else 0, float(Restecg), float(Thalach),
            1 if Exang == "Yes" else 0
        ]
        result = heart_disease_model.predict([user_input])[0]
        if result == 1:
            st.error("High risk of heart disease. Please consult a doctor.")
        else:
            st.success("No signs of heart disease. Stay healthy!")
    except ValueError:
        st.error("Please enter valid numeric inputs.")

elif selected == "Parkinson's Prediction":
    st.title("Parkinson's Prediction")
    col1, col2, col3, col4 = st.columns(4)

    with col1: Fo = st.text_input('MDVP:Fo(Hz)')

```

```

with col2: Fhi = st.text_input('MDVP:Fhi(Hz)')
with col3: Flo = st.text_input('MDVP:Flo(Hz)')
with col4: Jitter_percent = st.text_input('MDVP:Jitter(%)')
with col1: Jitter_Abs = st.text_input('MDVP:Jitter(Abs)')
with col2: RAP = st.text_input('MDVP:RAP')
with col3: PPQ = st.text_input('MDVP:PPQ')
with col4: DDP = st.text_input('Jitter:DDP')

if st.button('Get Parkinson's Prediction'):
    try:
        user_input = [
            float(Fo), float(Fhi), float(Flo), float(Jitter_percent), float(Jitter_Abs),
            float(RAP), float(PPQ), float(DDP)
        ]
        result = parkinsons_model.predict([user_input])[0]
        if result == 1:
            st.error("You may have Parkinson's disease. Please consult a specialist.")
        else:
            st.success("No signs of Parkinson's disease. Stay healthy!")
    except ValueError:
        st.error("Please enter valid numeric inputs.")

elif selected == "Breast Cancer Prediction":
    st.title("Breast Cancer Prediction")
    col1, col2, col3 = st.columns(3)

    with col1: Radius = st.text_input('Mean Radius')
    with col2: Texture = st.text_input('Mean Texture')
    with col3: Perimeter = st.text_input('Mean Perimeter')
    with col1: Area = st.text_input('Mean Area')
    with col2: Smoothness = st.text_input('Mean Smoothness')

    if st.button('Get Breast Cancer Prediction'):
        try:
            user_input = [

```

```

    float(Radius), float(Texture), float(Perimeter), float(Area), float(Smoothness)
]

result = breast_cancer_model.predict([user_input])[0]
if result == 1:
    st.error("High risk of breast cancer. Please consult an oncologist.")
else:
    st.success("No signs of breast cancer. Stay healthy!")

except ValueError:
    st.error("Please enter valid numeric inputs.")

elif selected == "Liver Disease Prediction":
    st.title("Liver Disease Prediction")
    col1, col2, col3 = st.columns(3)

    with col1: Age = st.text_input('Age')
    with col2: Total_Bilirubin = st.text_input('Total Bilirubin')
    with col3: Direct_Bilirubin = st.text_input('Direct Bilirubin')
    with col1: Alkphos = st.text_input('Alkaline Phosphatase')
    with col2: Sgpt = st.text_input('Alamine Aminotransferase (SGPT)')
    with col3: Sgot = st.text_input('Aspartate Aminotransferase (SGOT)')
    with col1: Total_Proteins = st.text_input('Total Proteins')
    with col2: Albumin = st.text_input('Albumin')
    with col3: AG_Ratio = st.text_input('Albumin/Globulin Ratio')

if st.button('Get Liver Disease Prediction'):
    try:
        user_input = [
            float(Age), float(Total_Bilirubin), float(Direct_Bilirubin), float(Alkphos),
            float(Sgpt), float(Sgot), float(Total_Proteins), float(Albumin), float(AG_Ratio)
        ]
        result = liver_model.predict([user_input])[0]
        if result == 1:
            st.error("High risk of liver disease. Please consult a hepatologist.")
        else:
            st.success("No signs of liver disease. Stay healthy!")
    
```

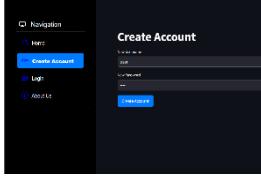
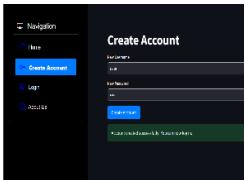
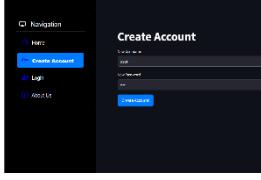
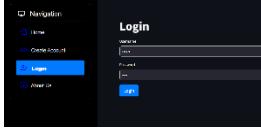
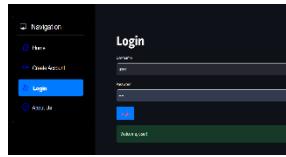
```
except ValueError:  
    st.error("Please enter valid numeric inputs.")
```

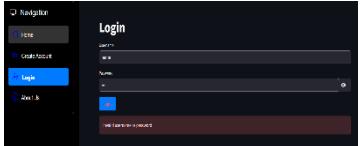
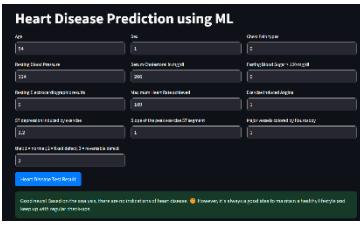
This Python code creates a Disease Prediction System using Streamlit for the user interface, Spyder as the development environment, and scikit-learn (sklearn) for machine learning model operations. The application loads pre-trained models for predicting diabetes, heart disease, Parkinson's disease, breast cancer, and liver disease using pickle. Users can input relevant health data on the respective disease prediction pages, and the app uses the loaded models to generate predictions. The system provides feedback based on the predictions, advising users to consult a doctor if needed. The app includes a custom-designed interface with navigation menus and styling, ensuring an interactive experience for users. It also includes error handling for invalid inputs. This solution leverages Streamlit for easy deployment and sklearn for the machine learning models, developed and tested within the Spyder IDE.

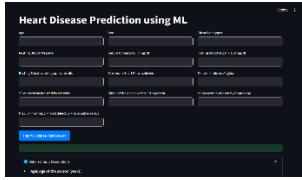
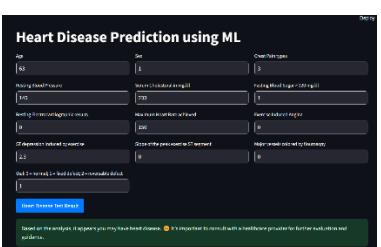
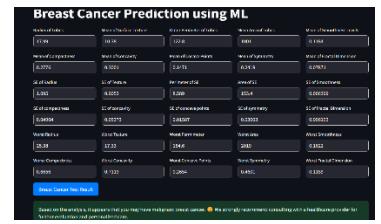
System Testing, Results and Discussion

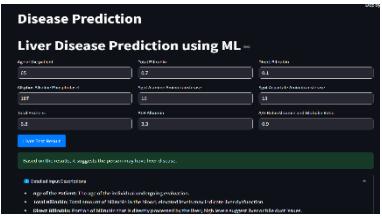
7.1 System Testing

Table 7.1: Unit Test Cases

Test Case Number	Input	Stage	Expected Behavior	Observed Behavior	Status P=Pass F=Fail
1	Create a new account 	Registration page	Creates a new account	Account created successfully 	P
2	Create an account with existing username and password 	Registration Page	Should not create an account with same username and password	Error message generated 	P
3	Login with username and password 	Login Page	Should successfully log in	Successfully logged into next page 	P

4	Login with non-existed username and password	Login page	Should not login	Generates an error message	P
					
5	Insert all the values to the displayed parameters in diabetes prediction Page (person is diabetic)	Prediction page	Should display result as "The person is diabetic"	The person is diabetic  Diabetes Prediction using ML	P
6	Insert all the values to the displayed parameters in diabetes prediction Page (person is not diabetic)	Prediction page	Should display result as "The person is not diabetic"	The person is not diabetic  Diabetes Prediction using ML	P
7	Insert all the values to the displayed parameters heart disease prediction Page (person does not have a heart disease)	Prediction page	Should display result as "The person does not have heart disease"	The person does not have heart disease  Heart Disease Prediction using ML	P

8	Insert all the values to the displayed parameters in heart disease prediction Page (person has heart disease)	Prediction page	Should display result as “The person has heart disease”	The person has heart disease	P
					
9	Insert all the values to the displayed parameters in Parkinson's prediction Page (person has Parkinson's)	Prediction page	Should display result as “The person has Parkinson's”	The person has Parkinson's	P
					
10	Insert all the values to the displayed parameters in Parkinson's prediction Page (person does not have Parkinson's)	Prediction page	Should display result as “The person does not have Parkinson's”	The person does not have Parkinson's	P
					
11	Insert all the values to the displayed parameters in Breast Cancer prediction Page (person has Malignant Breast Cancer)	Prediction page	Should display result as “The person has Malignant Breast Cancer”	The person has Malignant Breast Cancer	P
					

12	<p>Insert all the values to the displayed parameters in liver disease prediction Page (person has liver disease)</p> 	<p>Predict ion page</p>	<p>Should display result as “The person has liver disease”</p>	<p>The person has liver disease</p> 	<p>P</p>
----	---	---------------------------------	--	---	----------

7.2 Result Analysis

The main aim of the project was to predict multiple diseases using machine learning algorithms. Table 7.2 to 7.6 shows the analysis that was performed on the five diseases with different training and testing sizes.

Table 7.2: Analysis of Diabetes using three algorithms

Training Size	Testing Size	Accuracy (%)		
		SVM	LR	RF
80%	20%	77.27	75.32	75.32

Table 7.3: Analysis of heart disease using three algorithms

Training Size	Testing Size	Accuracy (%)		
		SVM	LR	RF
80%	20%	81.96	81.96	77.05

Table 7.4: Analysis of the Parkinson's using three algorithms

Training Size	Testing Size	Accuracy (%)		
		SVM	LR	RF
80%	20%	87.179	87.18	84.61

Table 7.5: Analysis of the Breast Cancer using three algorithms

Training Size	Testing Size	Accuracy (%)		
		SVM	LR	RF
80%	20%	94.74	92.98	94.73

Table 7.6: Analysis of the Liver disease using three algorithms

Training Size	Testing Size	Accuracy (%)		
		SVM	LR	RF
80%	20%	NA	71.87	99.65

Figure 7.1 shows the bar graph for the accuracy of diabetes using three algorithms where the train set size was 80% and the test set size was 20%.

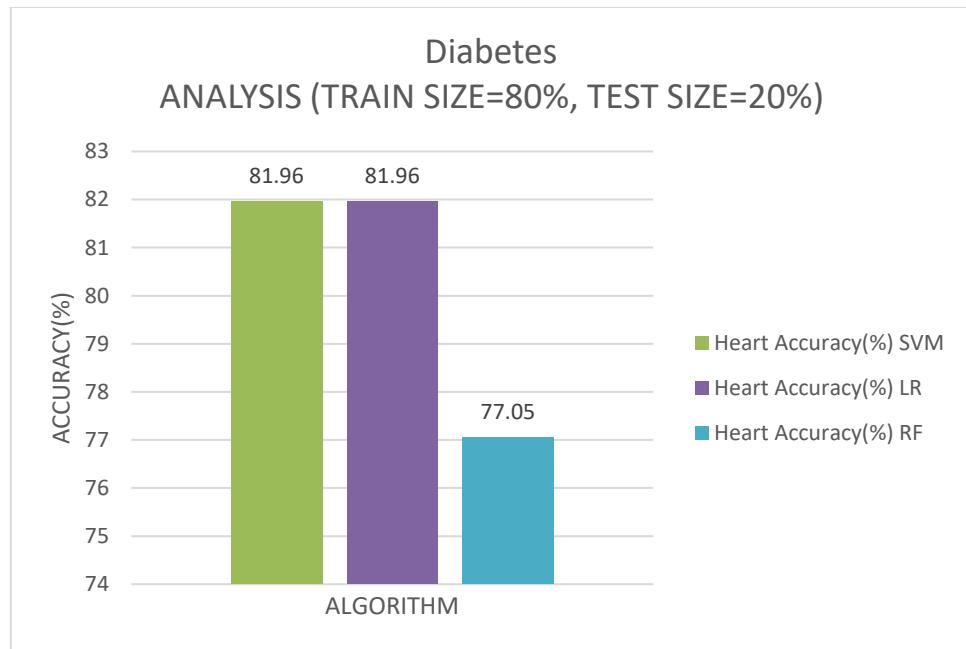


Figure 7.1: Graph analysis of the first set

Figure 7.2 shows the bar graph for the accuracy of heart disease using three algorithms where the train set size was 80% and the test set size was 20%.

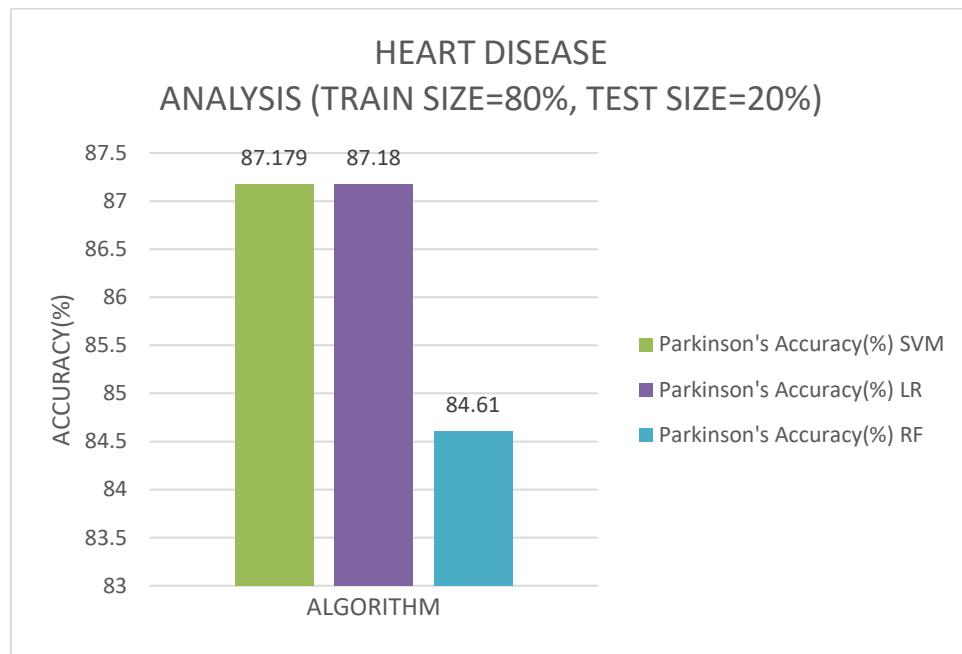


Figure 7.2: Graph analysis of the second set

Figure 7.3 shows the bar graph for the accuracy of the Parkinson's using three algorithms where the train set size was 80% and the test set size was 20%.

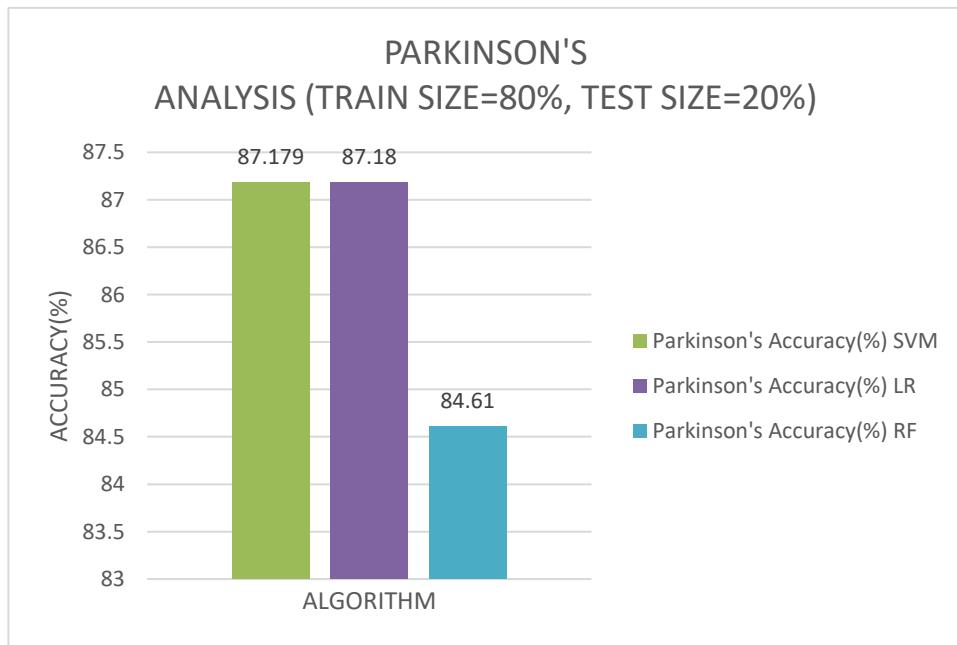


Figure 7.3: Graph analysis of the third set

Figure 7.4 shows the bar graph for the accuracy of breast cancer using three algorithms where the train set size was 80% and the test set size was 20%.

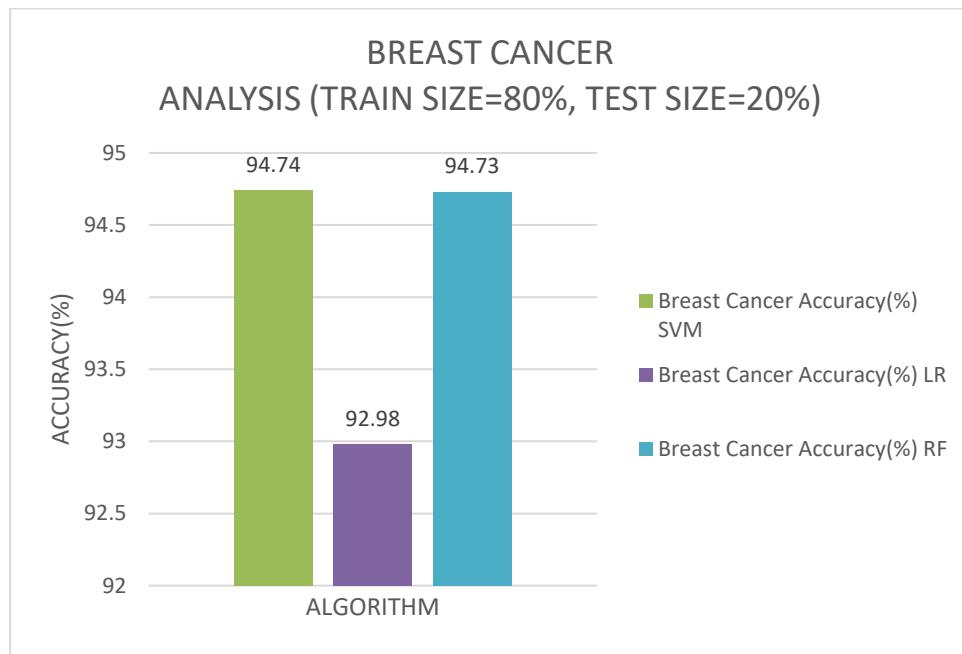


Figure 7.4: Graph analysis of the fourth set

Figure 7.5 shows the bar graph for the accuracy of the liver disease using three algorithms where the train set size was 80% and the test set size was 20%.

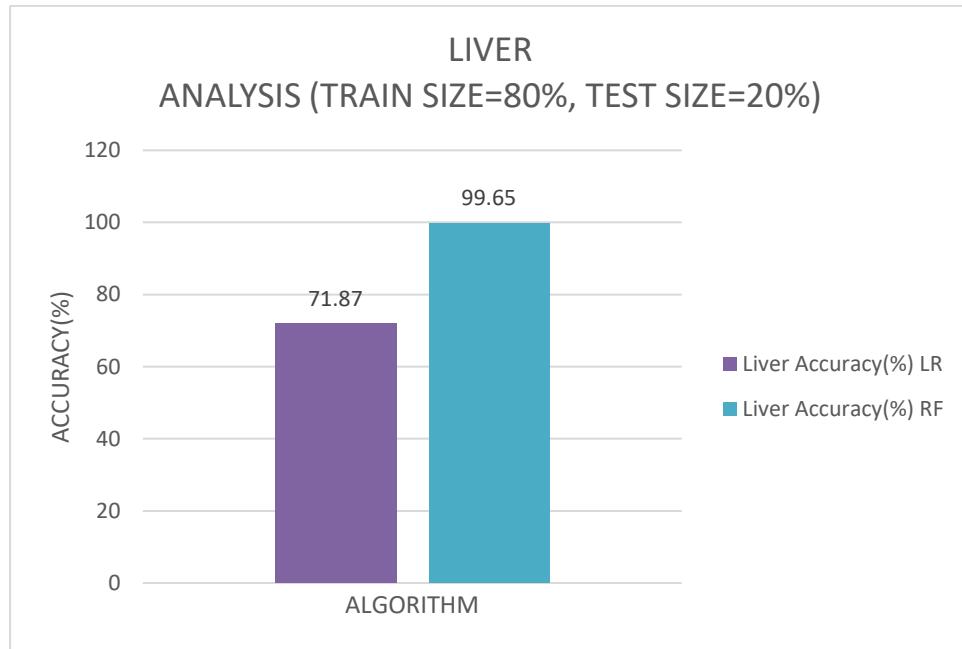


Figure 7.5: Graph analysis of the fifth set

Figure 7.6 is the home page of our website.

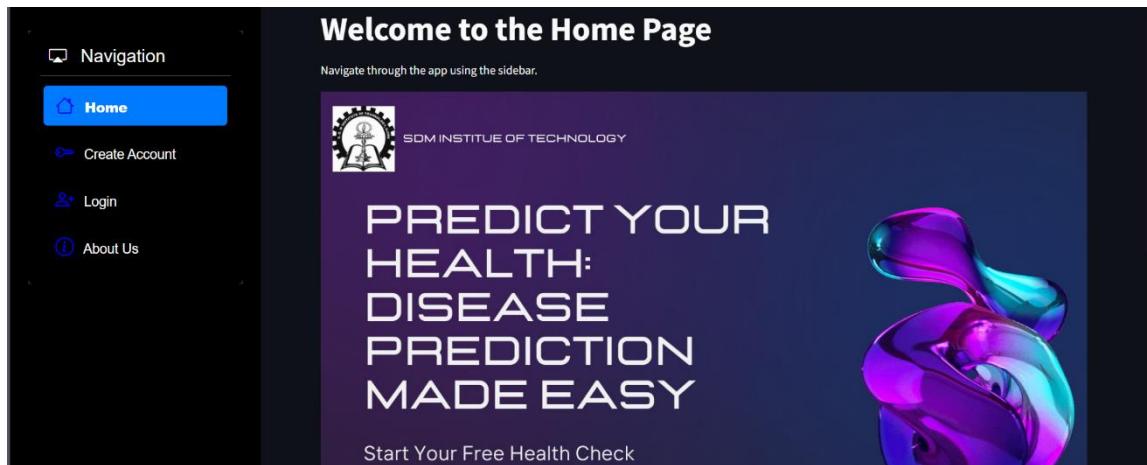


Figure 7.6: Home Page

Figure 7.7 is the about us page of our website.

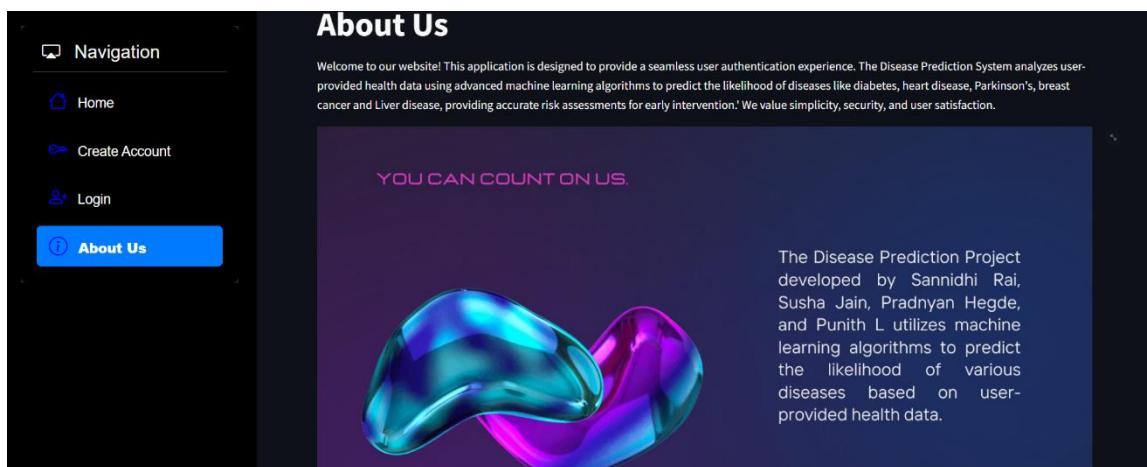


Figure 7.7: About Us Page

Figure 7.8 is the sign-up page to the users to create an account.

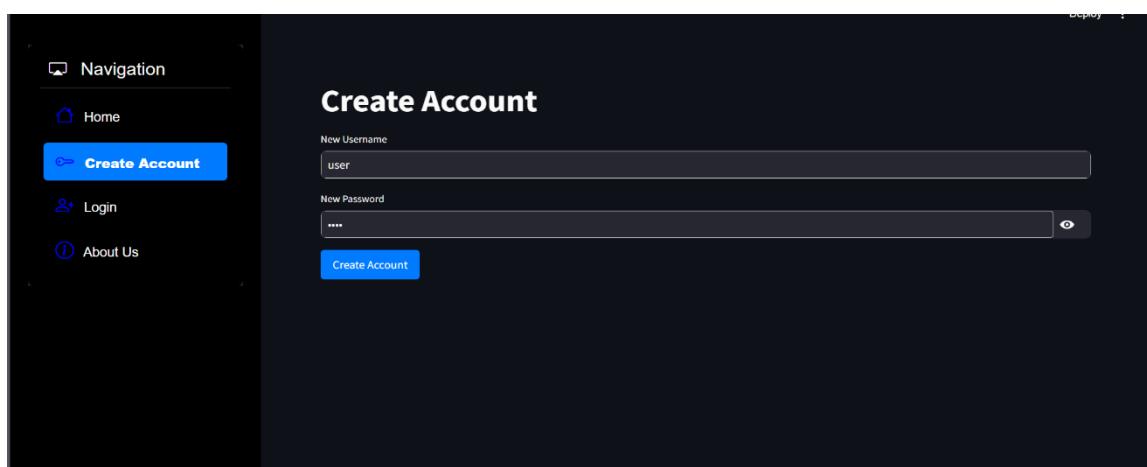


Figure 7.8: Sign Up Page

Figure 7.9 is the Login page for the users who wants to access the Page

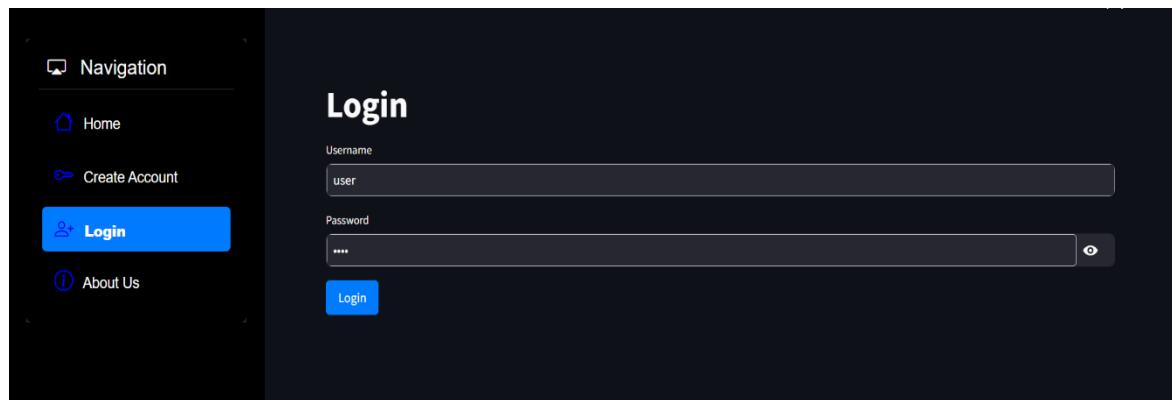


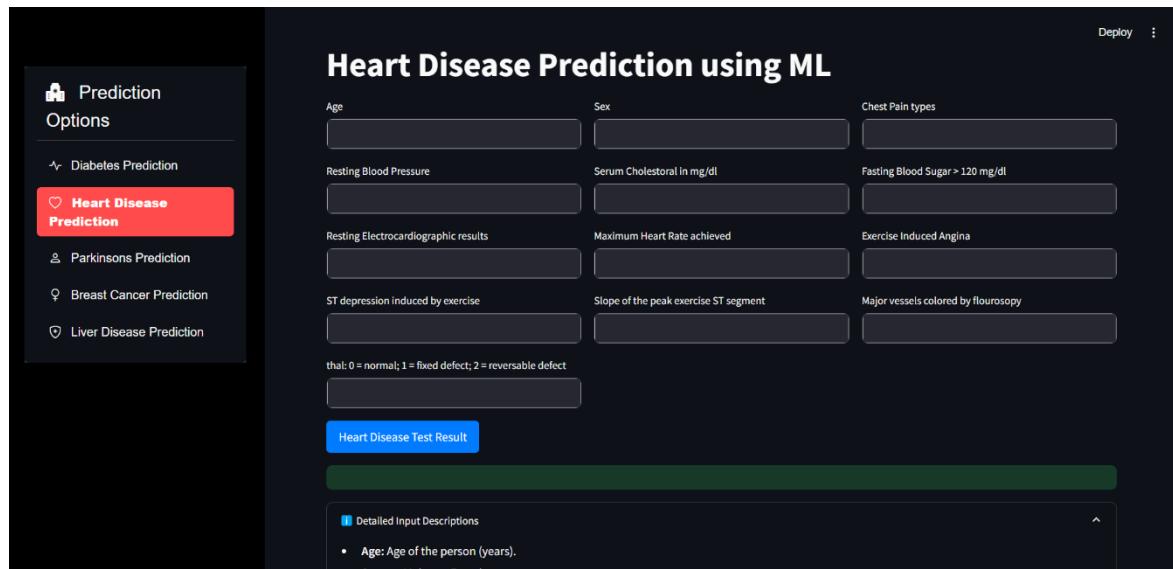
Figure 7.9: Login Page

Figure 7.10 is the page used to predict diabetes. Here the user will enter the values of the parameters. Eight Features are present. These features are responsible for the result.

Detailed Input Descriptions

Figure 7.10: Diabetes Prediction Page

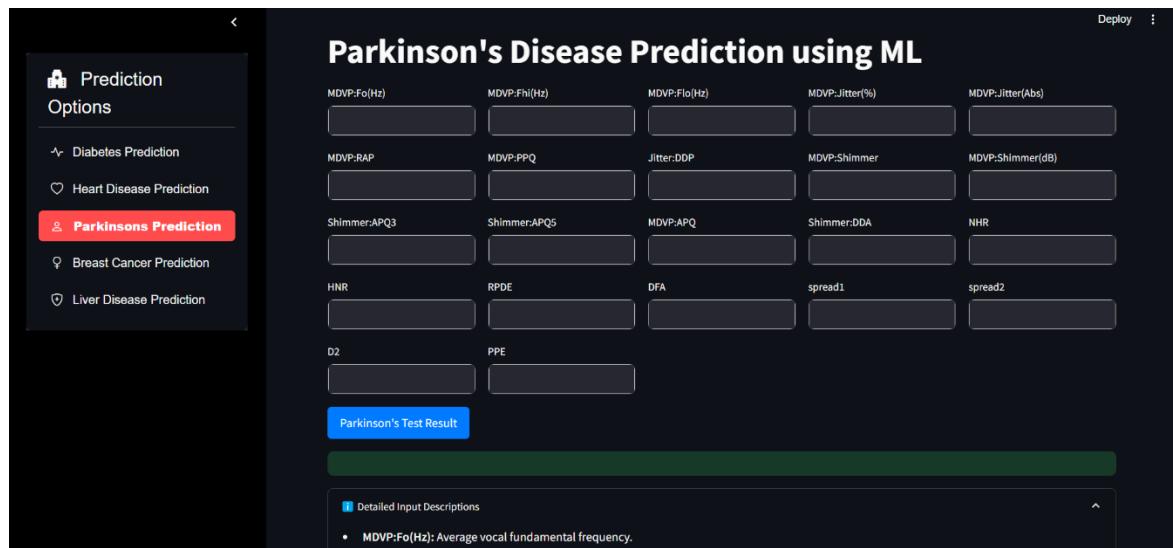
Figure 7.11 is the page used to predict heart disease. Here the user will enter the values of the parameters. Thirteen Features are present. These features are responsible for the result.



The screenshot shows a web-based machine learning application for heart disease prediction. On the left, a sidebar titled "Prediction Options" lists "Diabetes Prediction", "Heart Disease Prediction" (which is highlighted in red), "Parkinsons Prediction", "Breast Cancer Prediction", and "Liver Disease Prediction". The main area is titled "Heart Disease Prediction using ML". It contains 13 input fields arranged in three rows. Row 1: Age, Sex, Chest Pain types. Row 2: Resting Blood Pressure, Serum Cholesterol in mg/dl, Fasting Blood Sugar > 120 mg/dl. Row 3: Resting Electrocardiographic results, Maximum Heart Rate achieved, Exercise Induced Angina. Below these are two more rows of input fields: "ST depression induced by exercise", "Slope of the peak exercise ST segment", and "Major vessels colored by fluoroscopy". A note below the first row states: "that: 0 = normal; 1 = fixed defect; 2 = reversible defect". At the bottom is a blue "Heart Disease Test Result" button. A "Detailed Input Descriptions" section at the bottom provides definitions for the input fields, such as "Age: Age of the person (years)".

Figure 7.11: Heart Disease Prediction Page

Figure 7.12 is the page used to predict Parkinson's disease. Here the user will enter the values of the parameters. Twenty-two features are present. These features are responsible for the result.



The screenshot shows a web-based machine learning application for Parkinson's disease prediction. On the left, a sidebar titled "Prediction Options" lists "Diabetes Prediction", "Heart Disease Prediction", "Parkinsons Prediction" (which is highlighted in red), "Breast Cancer Prediction", and "Liver Disease Prediction". The main area is titled "Parkinson's Disease Prediction using ML". It contains 22 input fields arranged in five rows. Row 1: MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), MDVP:Jitter(%), MDVP:Jitter(Abs). Row 2: MDVP:RAP, MDVP:PPQ, Jitter:DDP, MDVP:Shimmer, MDVP:Shimmer(dB). Row 3: Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA, NHR. Row 4: HNR, RPDE, DFA, spread1, spread2. Row 5: D2, PPE. Below these is a blue "Parkinson's Test Result" button. A "Detailed Input Descriptions" section at the bottom provides definitions for the input fields, such as "MDVP:Fo(Hz): Average vocal fundamental frequency".

Figure 7.12: Parkinson's Prediction Page

Figure 7.13 is the page used to predict Breast Cancer. Here the user will enter the values of the parameters. Twenty-five features are present. These features are responsible for the result.

The screenshot shows a user interface for 'Breast Cancer Prediction using ML'. On the left, a sidebar titled 'Prediction Options' lists several prediction models: Diabetes Prediction, Heart Disease Prediction, Parkinsons Prediction, Breast Cancer Prediction (which is highlighted with a red background), and Liver Disease Prediction. Below this is a 'Breast Cancer Test Result' button. The main area is titled 'Breast Cancer Prediction using ML' and contains a grid of input fields for 25 different features. The features are grouped into five rows of five fields each. The first row includes: Radius of Lobes, Mean of Surface Texture, Outer Perimeter of Lobes, Mean Area of Lobes, and Mean of Smoothness Levels. The second row includes: Mean of Compactness, Mean of Concavity, Mean of Cocave Points, Mean of Symmetry, and Mean of Fractal Dimension. The third row includes: SE of Radius, SE of Texture, Perimeter of SE, Area of SE, and SE of Smoothness. The fourth row includes: SE of compactness, SE of concavity, SE of concave points, SE of symmetry, and SE of Fractal Dimension. The fifth row includes: Worst Radius, Worst Texture, Worst Perimeter, Worst Area, and Worst Smoothness. The sixth row includes: Worse Compactness, Worst Concavity, Worst Concave Points, Worst Symmetry, and Worst Fractal Dimension. At the bottom of the main area is a large green progress bar.

Figure 7.13: Breast Cancer Prediction Page

Figure 7.14 is the page used to predict Liver disease. Here the user will enter the values of the parameters. Nine features are present. These features are responsible for the result.

The screenshot shows a user interface for 'Liver Disease Prediction using ML'. On the left, a sidebar titled 'Prediction Options' lists several prediction models: Diabetes Prediction, Heart Disease Prediction, Parkinsons Prediction, Breast Cancer Prediction, and Liver Disease Prediction (which is highlighted with a red background). Below this is a 'Liver Test Result' button. The main area is titled 'Liver Disease Prediction using ML' and contains a grid of input fields for nine features. The features are grouped into three rows of three fields each. The first row includes: Age of the patient, Total Bilirubin, and Direct Bilirubin. The second row includes: Alkphos Alkaline Phosphotase, SGpt Alanine Aminotransferase, and SGot Aspartate Aminotransferase. The third row includes: Total Proteins, ALB Albumin, and A/G Ratio Albumin and Globulin Ratio. At the bottom of the main area is a detailed input descriptions section with a title 'Detailed Input Descriptions' and a list of nine items explaining the meaning of each feature. The list includes: Age of the Patient, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase (Alkphos), SGPT (Alanine Aminotransferase), SGOT (Aspartate Aminotransferase), and Total Proteins (Prot).

Figure 7.14: Liver Disease Prediction Page

7.3 Summary

The project leverages Streamlit to build an interactive and user-friendly web interface, enabling easy input of health data by users. Spyder is used as the primary integrated development environment (IDE) for coding and testing the application. Anaconda helps manage the required libraries and dependencies, ensuring a smooth setup of the project's environment. scikit-learn is utilized for implementing machine learning models that process the user input and generate predictions for various health conditions.

The website allows users to enter specific health metrics, such as blood pressure, glucose levels, and more, depending on the disease being predicted. After the data is submitted, the machine learning models analyse it to predict whether the user may be at risk for a particular disease. Based on the prediction, the system provides feedback, such as advice to consult a doctor or reassurance about the user's health.

With these technologies, the system provides a scalable and efficient tool for disease prediction. The use of Streamlit allows for quick deployment of the web app, while scikit-learn ensures accurate and reliable disease predictions. The entire project aims to make disease prediction accessible and user-friendly, leveraging the power of machine learning and a responsive web interface.

Conclusion and Scope for Future Work

8.1 Conclusion

In conclusion, the Multiple Disease Prediction System project is a significant step toward revolutionizing healthcare. By leveraging accurate predictive models, tailoring risk assessments to individual patients, and prioritizing transparency and collaboration across disciplines, this initiative exemplifies innovation in healthcare technology. As the project moves from research and development to practical applications, it holds the potential to reshape healthcare practices and improve patient outcomes. This system demonstrates how technology, when guided by ethical principles, patient-focused care, and interdisciplinary teamwork, can transform preventive medicine. Far from being a mere technological advancement, the system symbolizes a vision for a healthier and more resilient future. As the project reaches its completion, it invites ongoing efforts to innovate within healthcare technology. Rather than marking the end of a journey, it serves as a critical milestone in a broader exploration of opportunities. The lessons learned, ethical challenges navigated, and successes achieved lay a robust foundation for future developments. This project calls on researchers, healthcare professionals, policymakers, and technologists to collaboratively drive progress in healthcare systems and harness technology to enhance global health. Reflecting on this endeavour highlights the dedication, teamwork, and persistence of those involved. The process of designing, testing, and refining the system has not only resulted in a powerful solution but has also contributed to the growing body of knowledge in health technology. These insights will undoubtedly inspire and guide future advancements at the intersection of technology and medicine.

8.2 Scope for Future Work

In the future, the multiple disease prediction system can be expanded to include a wider range of diseases, allowing the API to identify and predict additional health conditions beyond diabetes, heart disease, and Parkinson's. This will involve continually updating the dataset and refining the model to improve its predictive accuracy. By incorporating advanced machine learning techniques and more comprehensive datasets, the goal is to enhance the system's ability to detect diseases earlier, thereby reducing the mortality rate. Additionally, the user interface can be made more intuitive and accessible, ensuring that both healthcare professionals and patients can easily interact with the system. To further improve the user experience, a chatbot feature can be integrated, allowing users to ask general health-related queries and receive instant, informative responses. This will make the system not only a powerful predictive tool but also a helpful resource for users seeking guidance on health-related matters.

References

- [1] V. Sharma, et al., "Multimodal Machine Learning Approaches for Disease Prediction: A Survey," 2022.
- [2] C. Chauhan, et al., "Multiple Disease Prediction Using Machine Learning Algorithms," 2021.
- [3] S. Kolli, et al., "Symptoms Based Multiple Disease Prediction Model Using Machine Learning Approach," 2021.
- [4] A. Gupta, et al., "A Comparative Study of Machine Learning Algorithms for Disease Prediction," 2021.
- [5] Bansal, et al., "Predictive Analytics for Disease Prevention and Diagnosis: A Machine Learning Approach," 2021.
- [6] D. Patel, et al., "Disease Prediction using Ensemble Learning Techniques in Healthcare," 2021.
- [7] S. Kumar, et al., "Feature Selection and Classification for Disease Prediction: A Machine Learning Approach," 2020.
- [8] R. Singh, et al., "Machine Learning Models for Disease Diagnosis and Prognosis in Healthcare," 2020.
- [9] A. Kamboj, et al., "A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives," 2020.
- [10] M. Rodriguez, et al., "Ensemble Machine Learning Models for Healthcare Disease Prediction," 2020.
- [11] P. Rajendran, et al., "Artificial Intelligence and Machine Learning Techniques in Healthcare Disease Prediction," 2020.
- [12] L. Brown, et al., "Utilizing Machine Learning to Improve Disease Diagnosis and Treatment," 2019.
- [13] J. Smith, et al., "Application of Deep Learning in Healthcare: A Survey of Techniques and Trends," 2019.
- [14] S. Zhang, et al., "Disease Prediction and Risk Assessment using Machine Learning Algorithms in Healthcare," 2018.
- [15] H. Al-Mallah, et al., "Multiple Disease Prediction Using Hybrid Deep Learning Architecture," 2016.
- [16] P. Krishnaiah, et al., "Predictive Modeling for Multiple Diseases Using Machine Learning with Feature Engineering," 2015.

Personal Profile



Mr. Pradeep G S has completed his B.E in the year 2010 and M.Tech in the year 2013 from VTU Belagavi. His areas of interest are Machine Learning and Network Security. He has published 3 international research papers.

Email ID: pradeepgs@sdmit.in

Mr. Pradeep G S,
Asst. Prof.
Project Guide



Name: Punith L
USN: 4SU21CS071
Address: Hirandalli, Bangalore-560049
E-mail ID: punipunith1924@gmail.com
Contact Phone No: 8147830890



Name: Sannidhi Rai
USN: 4SU21CS089
Address: Samrudha House, Santhekatte, Belthangady
E-mail ID: sannidhirai012@gmail.com
Contact Phone No: 6366163345



Name: Susha Jain
USN: 4SU21CS107
Address: Sumedha House, Ujire, Belthangady
E-mail ID: sushajain2003@gmail.com
Contact Phone No: 9686644196



Name: Pradnyan Hegde
USN: 4SU22CS405
Address: Undyapu Nagara, Ujire, Belthangady
E-mail ID: hegdepradnyan@gmail.com
Contact Phone No: 7338628423
