

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, **Mrs. Sannidhi Rai**, USN: 4SU21CS089, **Mrs. 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.

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 analysing large datasets, the system can predict diseases like diabetes, heart disease, cancer, and more. The model uses algorithms such as decision trees, neural networks, 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.

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

Table of Contents

	Page No.
Abstract	i
Table of Contents	ii
List of Figures	iii
List of Tables	iv
Chapter 1 Introduction	1
1.1 Project Introduction	1
1.2 Problem Description	3
Chapter 2 Literature Review	3
2.1 Literature Survey	3
2.2 Comparative Analysis of Related Work	5
2.3 Summary	5
Chapter 3 Problem Formulation	6
3.1 Problem Statement	6
3.2 Objectives of the Present Study	6
3.3 Summary	7
Chapter 4 Requirements and Methodology	8
4.1 Hardware Requirements	8
4.2 Software Requirements	8
4.3 Methodology Used	8
Chapter 5 System Design	9
5.1 Architecture of the Proposed System	9
5.2 System Flowchart	10
Chapter 6 Implementation	11
6.1 Pseudocode	11
Chapter 7 System Testing, Results and Discussion	13
7.1 System Testing	13
7.2 Result Analysis	16
7.3 Summary	23

Chapter 8	Conclusion and Scope for Future Work	24
8.1	Conclusion	24
8.2	Scope for Future work	24
References		
Personal Profile		

List of Figures

	Page No.
Figure 5.1 Architecture of the Proposed System	10
Figure 5.2 Flowchart of the Proposed System	11
Figure 7.1 Graph analysis of the first set	18
Figure 7.2 Graph analysis of the second set	18
Figure 7.3 Graph analysis of the third set	19
Figure 7.4 Graph analysis of the fourth set	19
Figure 7.5 Sign Up page	20
Figure 7.6 Login page	20
Figure 7.7 Diabetes Prediction Page	21
Figure 7.8 Heart Disease Prediction Page	21
Figure 7.9 Parkinson's Prediction Page	22
Figure 7.10 Breast Cancer Prediction Page	22
Figure 7.11 Liver Disease Prediction Page	23

List of Tables

	Page No.
Table 2.1 Comparative Analysis	5
Table 4.1 Hardware Requirements	8
Table 4.2 Software Requirements	8
Table 7.1 Unit Test Cases	13
Table 7.2 Analysis of diseases using SVM	17
Table 7.3 Analysis of diseases using Logistic Regression Model	17
Table 6.1 Histogram of the bands of the fused images	99
Table 6.2 Entropy of the bands of the fused images	99

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.

Literature Review

2.1 Literature Survey

In the paper titled “**Multiple Disease Prediction Using Machine Learning Algorithms**” [1], the author **Chauhan et al. (2021)** have investigated using various ML algorithms, including SVM and Decision Trees, for multiple disease prediction, focusing on symptoms as input. It examines the performance of these algorithms on four diseases, including heart disease and diabetes. The authors emphasize the potential of predictive analytics in healthcare to assist practitioners in making timely decisions regarding patients' health. The work aims to address the challenge of early recognition and diagnosis of harmful diseases, given the shortage of medical infrastructure and a low ratio of doctors to the population. The paper unifies multiple diseases under a single user interface for predictions and highlights the significance of early detection in saving lives. The study is 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**” [2], the author **Kolli et al. (2021)** have proposed symptom-based disease prediction using various ML algorithms like Random Forest, Decision Trees, and LightGBM. While it focuses on 41 diseases, you could adapt the methodology to your specific diseases of interest. The system's predictions are reported to be highly accurate, and it is designed to assist medical professionals in making more informed decisions and providing better-targeted therapies. The work is a valuable contribution to the field of healthcare, offering a holistic and integrated approach to disease risk, early detection, and personalized interventions. The paper is available in the International Journal of Innovative Technology.

In the paper titled “**A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives**” [3], the author **Kamboj et al. (2020)** have proposed a framework for early disease prediction using an ensemble model combining Logistic Regression, SVM, and K-Nearest Neighbors. It showcases the effectiveness of this approach for multiple diseases, potentially including your chosen ones. The paper provides insights into the application of machine learning in healthcare for the early prediction of multiple

diseases, emphasizing the importance of accurate predictions and timely interventions to improve patient outcomes.

In the paper titled **“Multiple Disease Prediction Using Hybrid Deep Learning Architecture”** [4], the author **Al-Mallah et al. (2016)** have explored using a hybrid deep learning architecture for multiple disease prediction, encompassing diseases like diabetes and heart disease. Studying their approach might provide insights for applying deep learning techniques to your project. The authors utilize a comprehensive dataset of medical records and symptoms of various diseases, which are then analyzed using deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The proposed system involves three phases: data normalization, weighted normalized feature extraction, and prediction. The system's predictions are reported to be highly accurate, and it can assist medical professionals in making more informed decisions and providing better-targeted therapies. The work is a valuable contribution to the field of healthcare, offering insights into the application of deep learning for disease prediction and highlighting the potential of hybrid deep learning architectures to improve model performance.

In the paper titled **“Predictive Modeling for Multiple Diseases Using Machine Learning with Feature Engineering”** [5], the author **Krishnaiah et al. (2015)** : This paper delves into feature engineering techniques for improving multiple disease prediction using KNearest Neighbors and Fuzzy K-NN approaches. This could be helpful for optimizing your feature selection and data preparation. The work is a valuable contribution to the field of healthcare, offering insights into the application of machine learning for disease prediction and highlighting the potential of feature engineering techniques to improve model performance. The paper is available in the International Research Journal of Modernization in Engineering Technology and Science. The paper emphasizes the importance of feature selection, model optimization, and comparative analyses for the development of accurate and reliable disease prediction models

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.	Chauhan et al	SVM and Decision Trees	Accuracy
2.	Kolli et al	Random Forest, Decision Trees, and LightGBM	Accuracy
3.	Kamboj et al	Ensemble model combining Logistic Regression, SVM, and K-Nearest Neighbors.	Accuracy
4.	Al-Mallah et al	Hybrid deep learning architecture, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks	Accuracy
5.	Krishnaiah et al	KNearest 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, KNN, and Naive Bayes algorithms and implement the prediction model using the best algorithm in the project.

Problem Formulation

3.1 Problem Statement

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

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.

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	
3.	Framework	Streamlit
4.	Google Collab	

4.3 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, lifestyle factors (smoking, diet, exercise), clinical test results (blood pressure, cholesterol), imaging data (X-rays, MRIs), and genetic data.
- 2) Data pre-processing:** Then the next step is data pre-processing. It is the process of converting raw data into a 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 purpose. 70% of the data is used to train the model and 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.

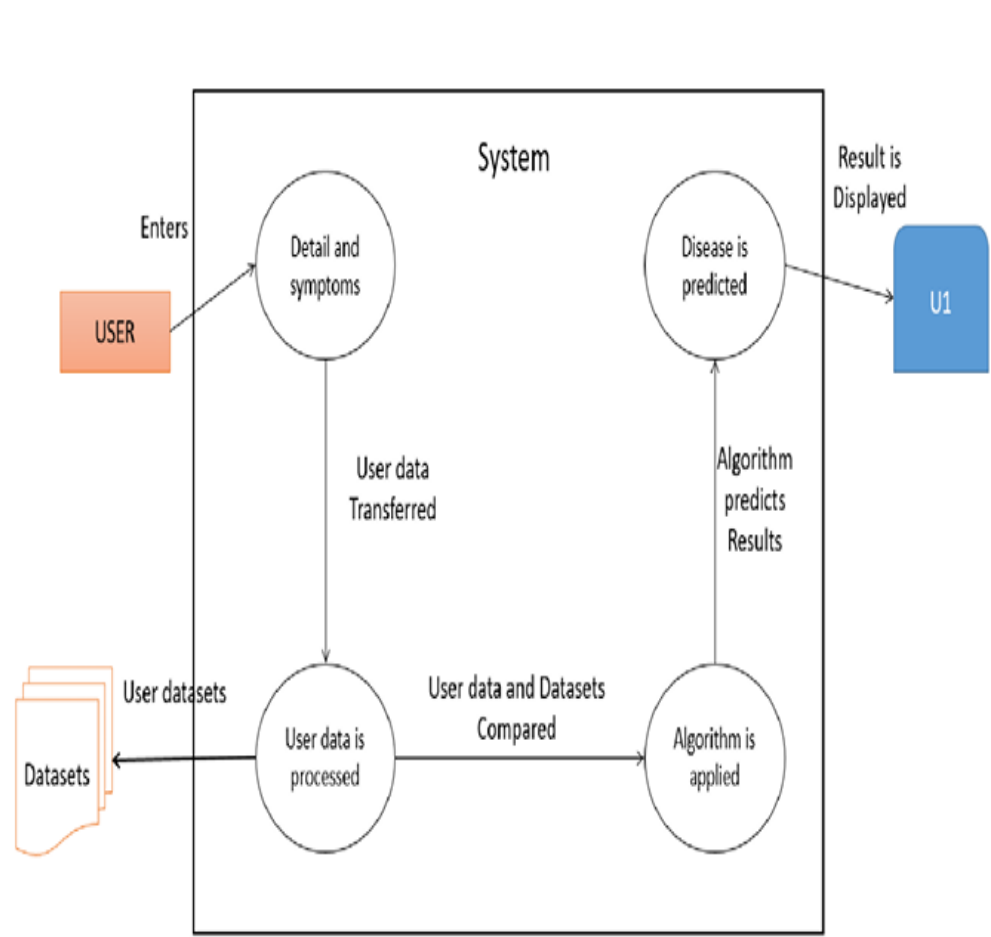


Figure 5.1: Architecture of the Proposed System

The 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.

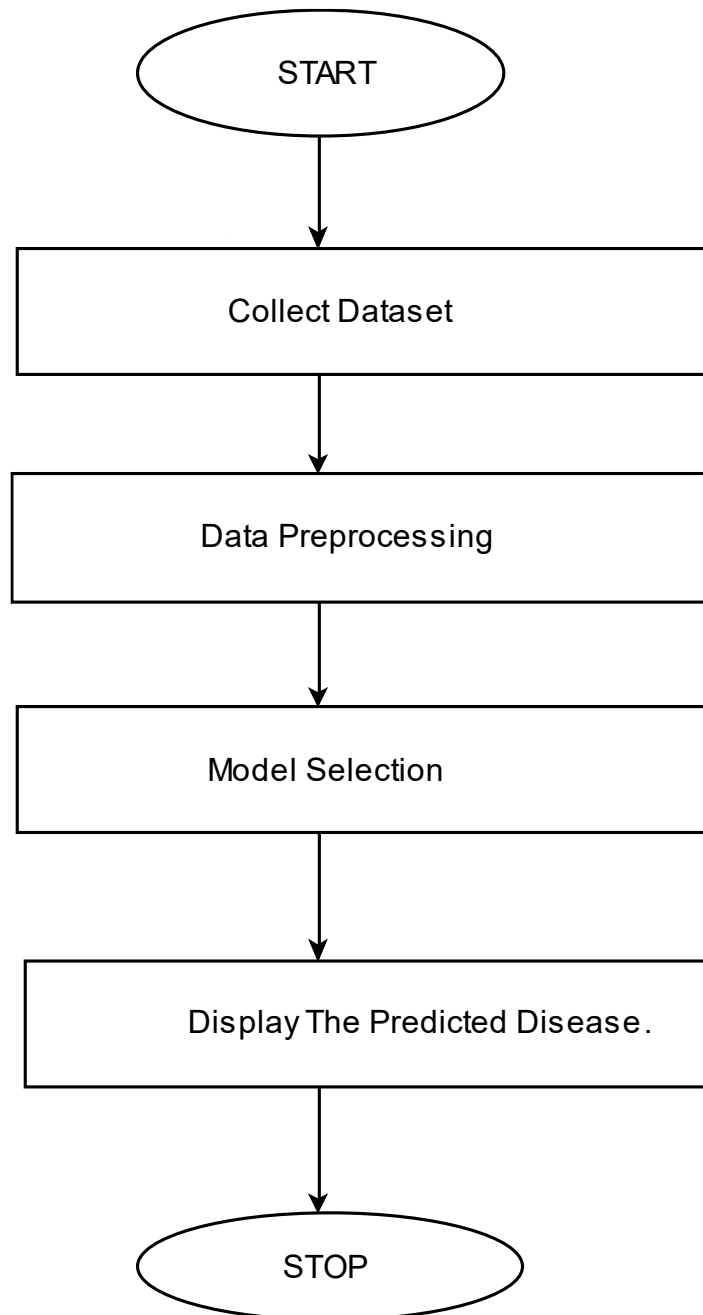


Figure 5.2: System Flowchart

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

Implementation

6.1 Pseudocode

```
# Step 1: Import Libraries
import required_libraries_for_GUI
import machine_learning_model_library
import numpy as np

# Step 2: Load Pre-trained Models
load_model = load('path_to_saved_model')

# Step 3: GUI Initialization
initialize_GUI()

# Step 4: User Input Section
if selected_option == "Specific Disease Prediction":
    # Create input fields for disease-specific parameters
    inputs = get_user_inputs()

# Step 5: Pre-process Inputs
processed_inputs = convert_inputs_to_float(inputs)

# Step 6: Predict Disease
prediction_result = model.predict(processed_inputs)

# Step 7: Interpret Results
if prediction_result indicates disease:
    display("The person has the disease.")
else:
    display("The person does not have the disease.")

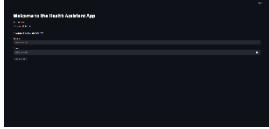
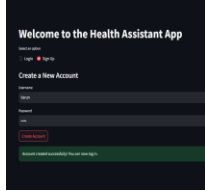
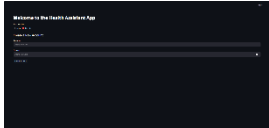
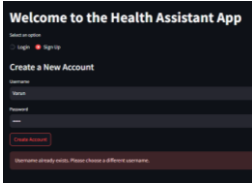
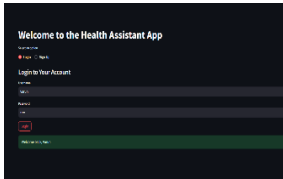
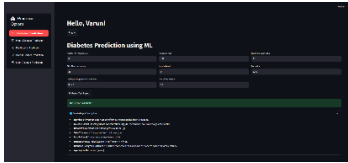
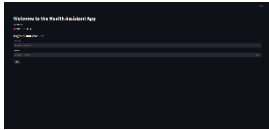
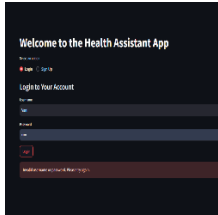
# Step 8: Optional - Display Detailed Explanations
display_expandable_information(disease_specific_details)
```


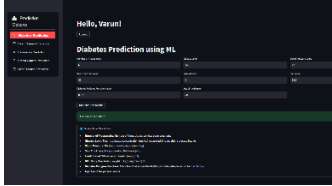

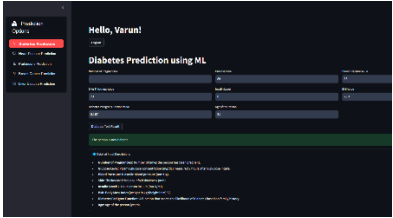

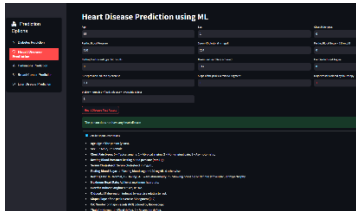

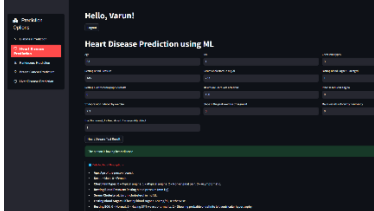
```
# Step 9: Repeat for Other Diseases
if selected_option == "Another Disease Prediction":
    # Repeat steps 4 to 8 for other diseases.
```


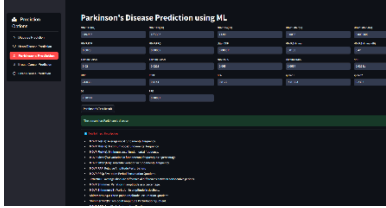



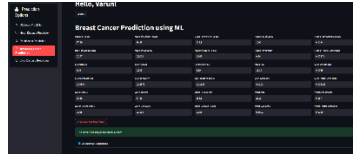
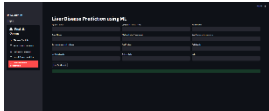
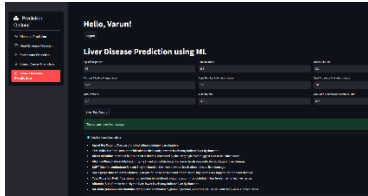
System Testing, Results and Discussion

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 	Preview Page	Should successfully create a new account	Account created successfully 	P
2	Create an account with existing username and password 	Preview Page	Should not create an account with same username and password	Error message generated 	P
3	Login with username and password 	Preview Page	Should successfully log in	Successfully logged into next page 	P
4	Login with non-existent username and password 	Preview page	Should not login	Generates an error message 	P

5	<p>Insert all the values to the displayed parameters in diabetes prediction Page (person is diabetic)</p> 	Preview page	Should display result as The person is diabetic	<p>The person is diabetic</p> 	P
6	<p>Insert all the values to the displayed parameters in diabetes prediction Page (person is not diabetic)</p> 	Preview page	Should display result as The person is not diabetic	<p>The person is not diabetic</p> 	P
7	<p>Insert all the values to the displayed parameters heart disease prediction Page (person does not have a heart disease)</p> 	Preview page	Should display result as The person does not have heart disease	<p>The person does not have heart disease</p> 	P
8	<p>Insert all the values to the displayed parameters in heart disease prediction Page (person has heart disease)</p> 	Preview page	Should display result as The person has heart disease	<p>The person has heart disease</p> 	P

9	<p>Insert all the values to the displayed parameters in Parkinson's prediction Page (person has Parkinson's)</p> 	Preview page	Should display result as The person has Parkinson's	<p>The person has Parkinson's</p> 	P
10	<p>Insert all the values to the displayed parameters in Parkinson's prediction Page (person does not have Parkinson's)</p> 	Preview page	Should display result as The person does not have Parkinson's	<p>The person does not have Parkinson's</p> 	P
11	<p>Insert all the values to the displayed parameters in Breast Cancer prediction Page (person has Breast Cancer)</p> 	Preview page	Should display result as The person has Breast Cancer	<p>The person has Breast Cancer</p> 	P
12	<p>Insert all the values to the displayed parameters in liver disease prediction Page (person has liver disease)</p> 	Preview page	Should display result as The person has liver disease	<p>The person has liver disease</p> 	P

7.2 Result Analysis

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

Table 7.2: Analysis of the diseases using SVM

Training Size	Testing Size	Accuracy (%) of diseases using Support Vector machine Classification		
		diabetes	Parkinson's	Breast cancer
80%	20%	78.33	87.18	96.92
70%	30%	78.03	88.97	96.98

Table 7.3: Analysis of diseases using Logistic Regression Model

Training Size	Testing Size	Accuracy (%) of diseases using Logistic Regression Model	
		Heart disease	Liver disease
80%	20%	85.12	71.42
70%	30%	84.43	71.43

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

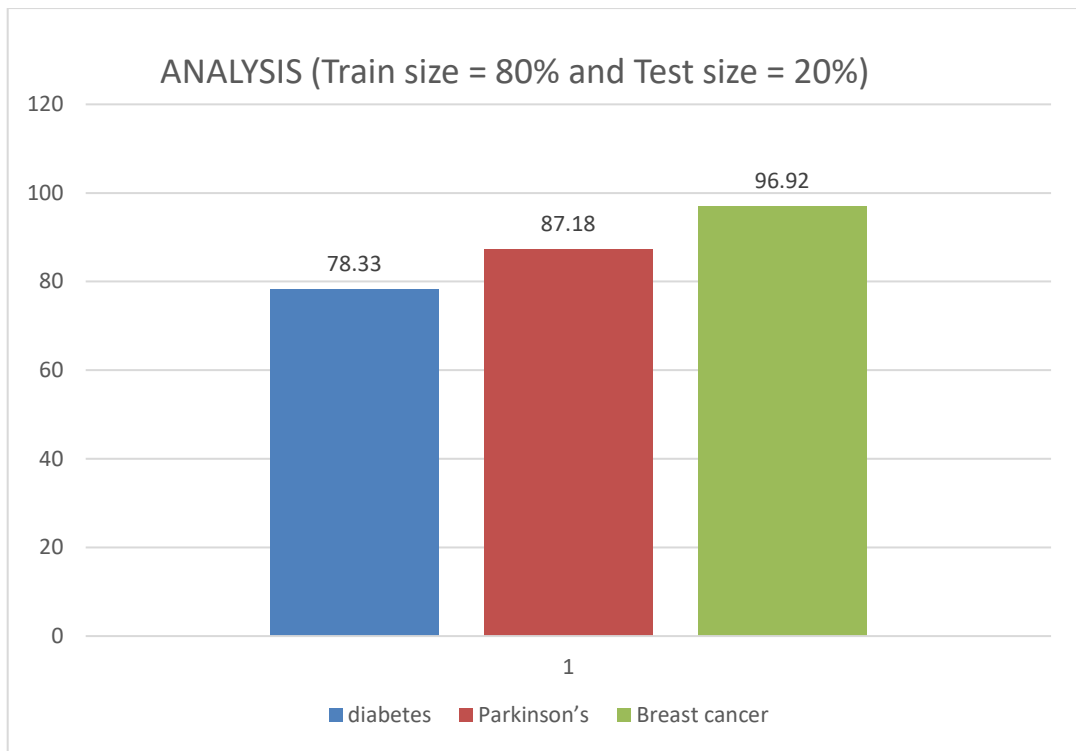


Figure 7.1: Graph analysis of the first set

Figure 7.2 shows the bar graph for the accuracy of the three diseases using SVM classification where the train set size was 70% and the test set size was 30%.

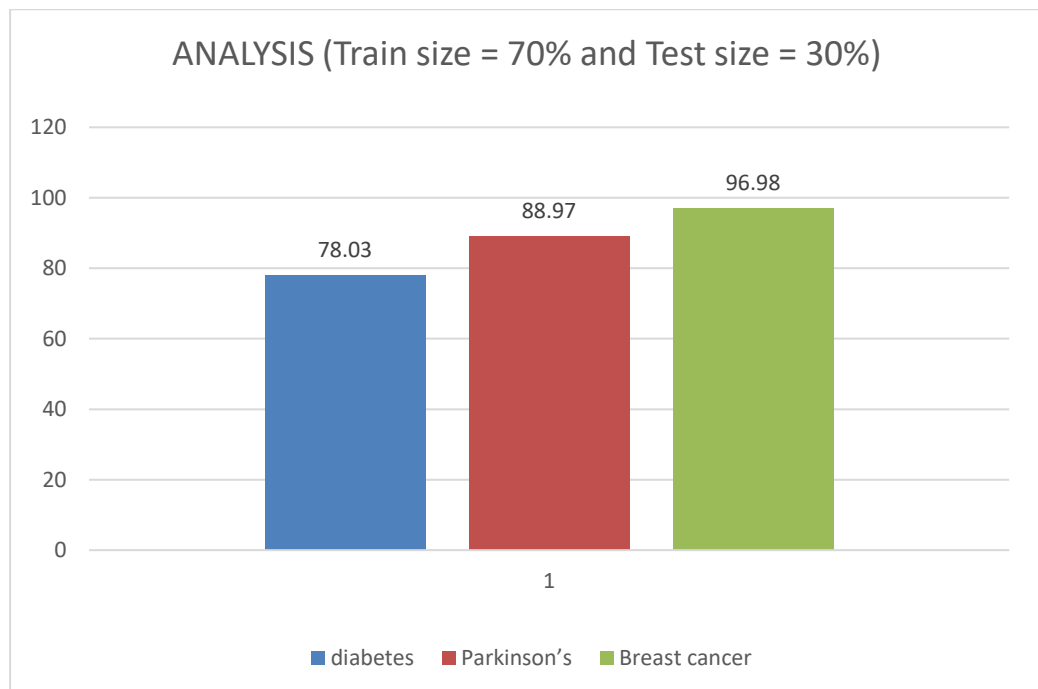


Figure 7.2: Graph analysis of the second set

Figure 7.3 shows the bar graph for the accuracy of the two diseases using Logistic Regression where the train set size was 80% and the test set size was 20%.

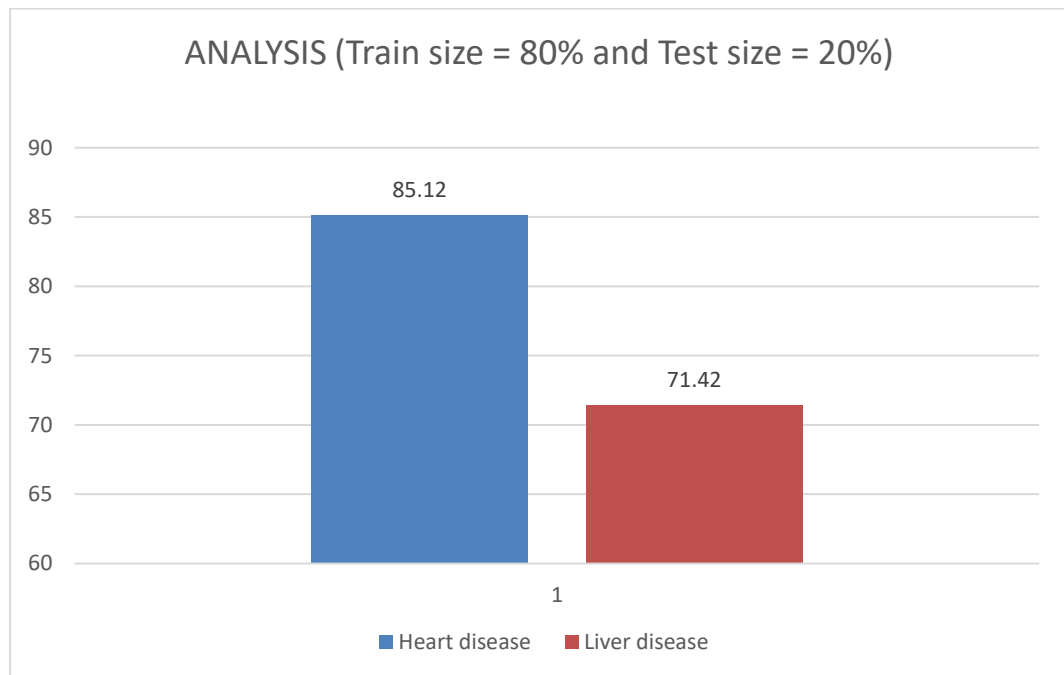


Figure 7.3: Graph analysis of the third set

Figure 7.4 shows the bar graph for the accuracy of the two diseases using Logistic Regression where the train set size was 70% and the test set size was 30%.

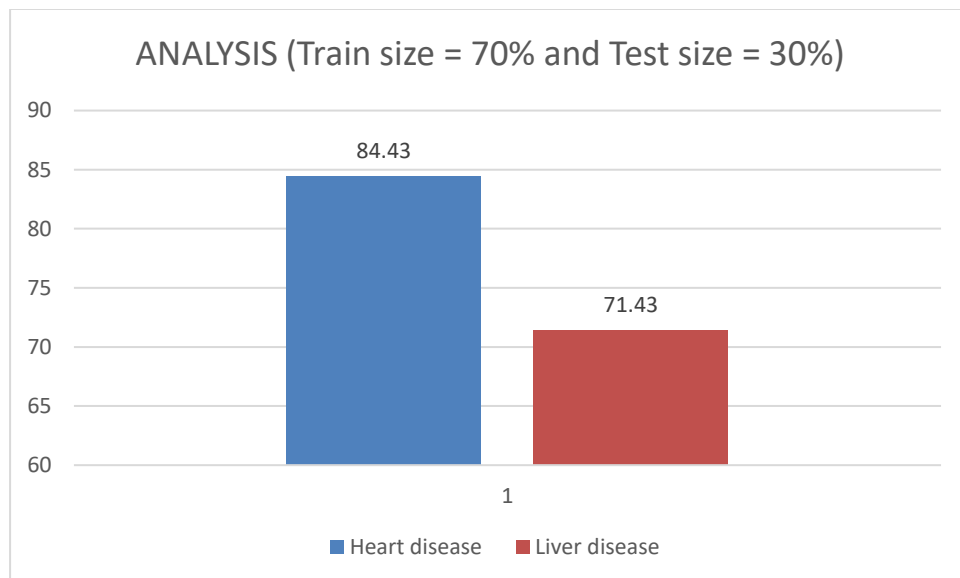
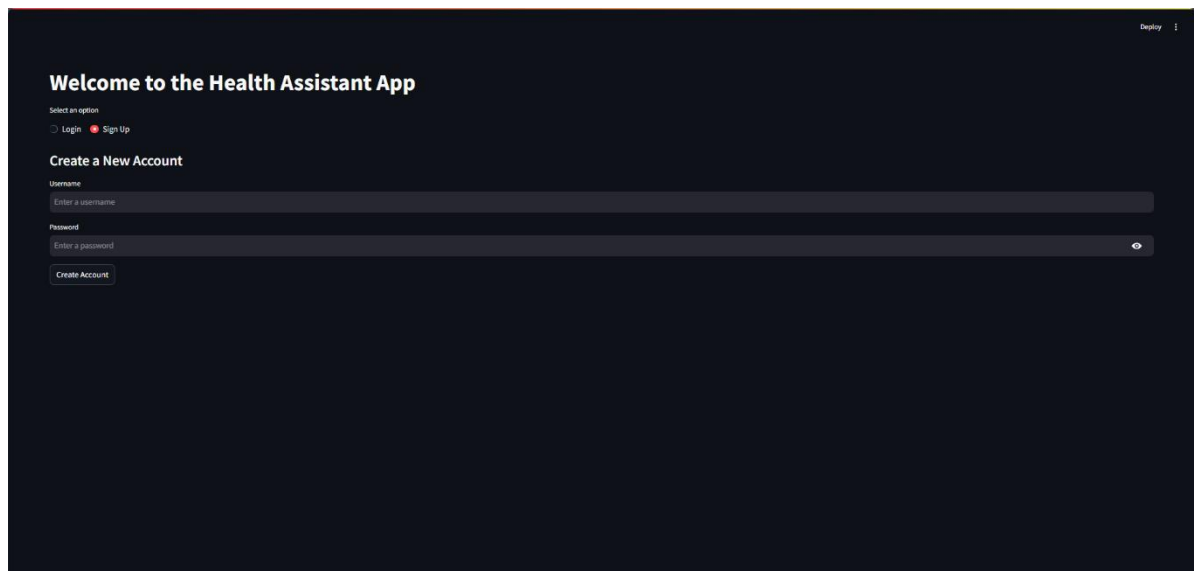


Figure 7.4: Graph analysis of the fourth set

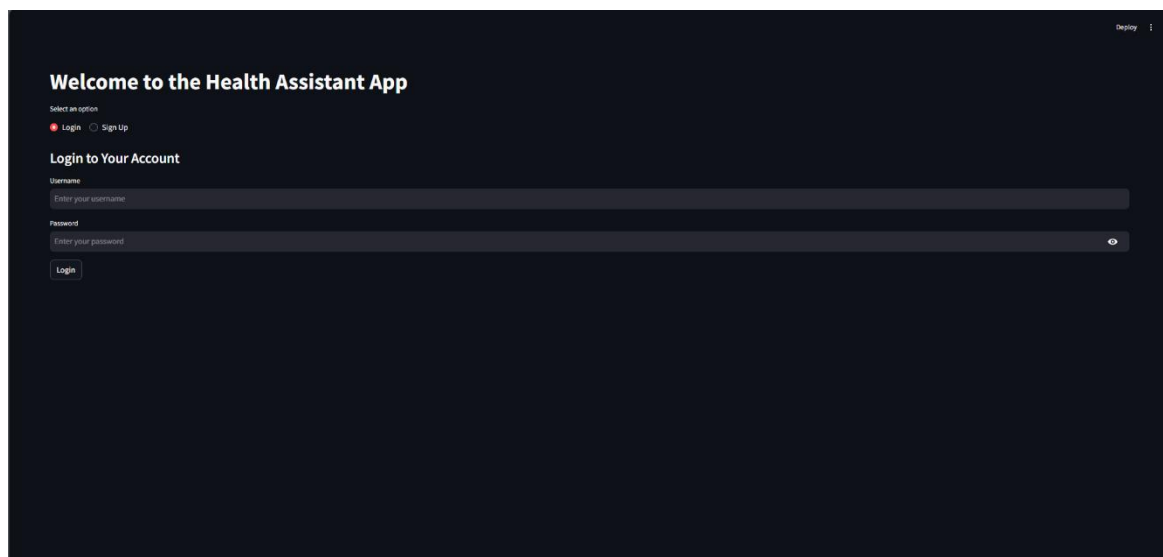
Figure 7.5 is the sign-up page to the users to create an account.



The image shows a dark-themed web interface for a 'Health Assistant App'. At the top right, there is a 'Deploy' link. The main heading is 'Welcome to the Health Assistant App'. Below it, there is a section 'Select an option' with two radio buttons: 'Login' and 'Sign Up'. The 'Sign Up' option is selected. Underneath, there is a section 'Create a New Account'. It contains two input fields: 'Username' with the placeholder text 'Enter a username' and 'Password' with the placeholder text 'Enter a password'. To the right of the password field is an eye icon for toggling visibility. At the bottom of this section is a 'Create Account' button.

Figure 7.5: Sign Up Page

Figure 7.6 is the Login page to the users who wants to access the Page



The image shows a dark-themed web interface for a 'Health Assistant App'. At the top right, there is a 'Deploy' link. The main heading is 'Welcome to the Health Assistant App'. Below it, there is a section 'Select an option' with two radio buttons: 'Login' and 'Sign Up'. The 'Login' option is selected. Underneath, there is a section 'Login to Your Account'. It contains two input fields: 'Username' with the placeholder text 'Enter your username' and 'Password' with the placeholder text 'Enter your password'. To the right of the password field is an eye icon for toggling visibility. At the bottom of this section is a 'Login' button.

Figure 7.6: Login Page

Figure 7.7 is the page used to predict diabetes. Here the user will enter the values of the parameters. Eight Features are present. These features are the ones that are responsible for the result.

Diabetes Prediction using ML

Number of Pregnancies:

Glucose Level:

Blood Pressure value:

Skin Thickness value:

Insulin Level:

BMI value:

Diabetes Pedigree Function value:

Age of the Person:

Diabetes Test Result:

Detailed Input Descriptions:

- Number of Pregnancies: Number of times the person has been pregnant.
- Glucose Level: Plasma glucose concentration (mg/dL) measured 2 hours after a glucose intake.
- Blood Pressure: Diastolic blood pressure (mm Hg).
- Skin Thickness: Triceps skinfold thickness (mm).
- Insulin Level: 2-Hour serum insulin (mu U/ml).
- BMI: Body Mass Index (weight in kg/height in m²).
- Diabetes Pedigree Function: A function that scores the likelihood of diabetes based on family history.
- Age: Age of the person (years).

Figure 7.7: Diabetes Prediction Page

Figure 7.8 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 the ones that are responsible for the result.

Heart Disease Prediction using ML

Age:

Sex:

Chest Pain types:

Resting Blood Pressure:

Serum Cholesterol in mg/dl:

Fasting Blood Sugar > 120 mg/dl:

Resting Electrocardiographic results:

Maximum Heart Rate achieved:

Exercise Induced Angina:

ST depression induced by exercise:

Slope of the peak exercise ST segment:

Major vessels colored by fluoroscopy:

that: 0 = normal, 1 = fixed defect, 2 = reversible defect

Heart Disease Test Result:

Detailed Input Descriptions:

- Age: Age of the person (years).
- Sex: 1 = Male, 0 = Female.
- Chest Pain types: 0 = Typical angina, 1 = Atypical angina, 2 = Non-anginal pain, 3 = Asymptomatic.
- Resting Blood Pressure: Resting blood pressure (mm Hg).
- Serum Cholesterol: Serum cholesterol in mg/dL.
- Fasting Blood Sugar: 1 if fasting blood sugar > 120 mg/dL, 0 otherwise.
- Resting ECG: 0 = Normal, 1 = Having ST-T wave abnormality, 2 = Showing probable or definite left ventricular hypertrophy.
- Maximum Heart Rate: Achieved maximum heart rate.
- Exercise Induced Angina: 1 = Yes, 0 = No.
- Oldpeak: ST depression induced by exercise relative to rest.
- Slope: Slope of the peak exercise ST segment (0-2).
- CA: Number of major vessels (0-3) colored by fluoroscopy.
- Thall: 0 = Normal, 1 = Fixed defect, 2 = Reversible defect.

Figure 7.8: Heart Disease Prediction Page

Figure 7.9 is the page used to predict Parkinson's disease. Here the user will enter the values of the parameters. 22 features are present. These features are the ones that are responsible for the result.

Parkinson's Disease Prediction using ML

MDVP:F0(Hz) MDVP:F1(Hz) MDVP:F2(Hz) MDVP:Jitter(Hz) MDVP:Jitter(Abs)

MDVP:RAP MDVP:FPPQ Jitter:DDP MDVP:Shimmer MDVP:Shimmer(dB)

Shimmer:APQ3 Shimmer:APQ5 MDVP:APQ Shimmer:DDA NNR

NNR RPDE DFA Spread1 spread2

D2 PPE

Parkinson's Test Result

Detailed Input Descriptions

- MDVP:F0(Hz): Average vocal fundamental frequency.
- MDVP:F1(Hz): Maximum vocal fundamental frequency.
- MDVP:F2(Hz): Minimum vocal fundamental frequency.
- MDVP:Jitter(Hz): Variation in fundamental frequency as a percentage.
- MDVP:Jitter(Abs): Absolute variation in fundamental frequency.
- MDVP:RAP: Relative Amplitude Perturbation.
- MDVP:FPPQ: Five-point Period Perturbation Quotient.
- Jitter:DDP: Average absolute difference of differences between consecutive periods.
- MDVP:Shimmer: Variation in amplitude as a percentage.
- MDVP:Shimmer(dB): Variation in amplitude in decibels.
- Shimmer:APQ3: Three-point Amplitude Perturbation Quotient.
- Shimmer:APQ5: Five-point Amplitude Perturbation Quotient.
- MDVP:APQ: Amplitude Perturbation Quotient.

Figure 7.9: Parkinson's Prediction Page

Figure 7.10 is the page used to predict Breast Cancer. Here the user will enter the values of the parameters. 25 Features are present. These features are the ones that are responsible for the result.

Breast Cancer Prediction using ML

Radius of Lobes Mean of Surface Texture Outer Perimeter of Lobes Mean Area of Lobes Mean of Smoothness Levels

Mean of Compactness Mean of Concavity Mean of Concave Points Mean of Symmetry Mean of Fractal Dimension

SE of Radius SE of Texture Perimeter of SE Area of SE SE of Smoothness

SE of compactness SE of concavity SE of concave points SE of symmetry SE of Fractal Dimension

Worst Radius Worst Texture Worst Perimeter Worst Area Worst Smoothness

Worst Compactness Worst Concavity Worst Concave Points Worst Symmetry Worst Fractal Dimension

Breast Cancer Test Result

Detailed Input Descriptions

- Radius Mean: Mean of distances from the center to points on the perimeter.
- Texture Mean: Mean of gray-level intensity variations.
- Perimeter Mean: Mean of tumor perimeter lengths.
- Area Mean: Mean of the tumor area.
- Smoothness Mean: Mean of local variation in radius lengths.
- Compactness Mean: Mean of $(\text{perimeter}^2 / \text{area} - 1.0)$.
- Concavity Mean: Mean of severity of concave portions of the contour.
- Concave Points Mean: Mean of the number of concave portions of the contour.
- Symmetry Mean: Mean symmetry of the tumor shape.
- Fractal Dimension Mean: Means of coarsening approximation (1 = smooth dimension).

Figure 7.10: Breast Cancer Prediction Page

Figure 7.11 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 the ones that are responsible for the result.

Prediction Options

Diabetes Prediction

Heart Disease Prediction

Parkinsons Prediction

Breast Cancer Prediction

Liver Disease Prediction

Hello, Varun!

Logout

Liver Disease Prediction using ML

Age of the patient

65

Total Bilirubin

0.7

Direct Bilirubin

0.1

Alkaline Phosphatase

187

SGPT Alanine Aminotransferase

36

SGPT Aspartate Aminotransferase

18

Total Proteins

6.8

ALB Albumin

3.3

A/G Ratio Albumin and Globulin Ratio

0.9

Liver Test Result

The person have liver disease

Detailed Input Descriptions

- **Age of the Patient:** The age of the individual undergoing evaluation.
- **Total Bilirubin:** Total amount of bilirubin in the blood; elevated levels may indicate liver dysfunction.
- **Direct Bilirubin:** Portion of bilirubin that is directly processed by the liver; high levels suggest liver or bile duct issues.
- **Alkaline Phosphatase (Alphes):** Enzyme linked to the bile ducts; increased levels may indicate blockage or liver disease.
- **SGPT (Alanine Aminotransferase):** Enzyme found in the liver; elevated levels often indicate liver damage.
- **SGPT (Aspartate Aminotransferase):** Enzyme found in the liver and other tissues; high levels can suggest liver or heart damage.
- **Total Proteins (Prots):** Total amount of proteins in the blood, including albumin and globulin; low levels may reflect liver issues.
- **Albumin (Alb):** Proteins made by the liver; lower levels may indicate liver dysfunction.
- **A/G Ratio (Albumin and Globulin Ratio):** Ratio of albumin to globulin proteins; abnormal values can indicate liver or kidney disease.

Figure 7.11: Liver Disease Prediction Page

7.3 Summary

The application was developed using the Flask framework. The programming languages that were used were Python, HTML, and CSS. The figures in the previous section showed the snapshots of various pages of the application. Since Logistic Regression was found to be the most accurate among the four algorithms, the prediction model was created using it.

23

Conclusion and Scope for Future Work

8.1 Conclusion

In conclusion, the Multiple Disease Prediction System project represents more than a technological endeavor; it is a glimpse into the future of healthcare. The accurate predictive models, patient-specific risk assessments, emphasis on interpretability, contributions to preventive healthcare practices, and cross disciplinary collaboration collectively position the project at the forefront of healthcare innovation. As the system transitions from the research and development phase to potential real-world applications, its impact on healthcare practices and patient outcomes is poised to be substantial. The project underscores the transformative potential of technology when aligned with the principles of healthcare ethics, patient centered care, and collaborative innovation. In shaping the future of preventive medicine, the Multiple Disease Prediction System emerges not just as a technological tool but as a beacon guiding the way towards a healthier and more resilient society. As this project concludes, it serves as a call to action for continued innovation in healthcare technology. The Multiple Disease Prediction System is not the culmination of a journey but a milestone in an ongoing exploration of possibilities. The challenges faced, the ethical considerations addressed, and the successes achieved form the foundation for future endeavors. The call to action extends to researchers, policymakers, healthcare professionals, and technologists to collectively contribute to the evolution of healthcare practices and the integration of technology for the betterment of global health. In reflecting on the technological journey undertaken in this project, it is essential to acknowledge the collaborative efforts, the dedication of the project team, and the resilience in overcoming challenges. The iterative process of development, testing, and refinement has not only led to the creation of a powerful tool but has also contributed to the collective knowledge base in health technology. The lessons learned in this project can inform future endeavors, laying the groundwork for further innovation in the intersection of technology and healthcare.

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] C. Chauhan, et al., "Multiple Disease Prediction Using Machine Learning Algorithms," 2021.
- [2] S. Kolli, et al., "Symptoms Based Multiple Disease Prediction Model Using Machine Learning Approach," 2021.
- [3] A. Kamboj, et al., "A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives," 2020.
- [4] H. Al-Mallah, et al., "Multiple Disease Prediction Using Hybrid Deep Learning Architecture," 2016.
- [5] P. Krishnaiah, et al., "Predictive Modeling for Multiple Diseases Using Machine Learning with Feature Engineering," 2015.

Personal Profile



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

Mr. Praddep 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



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
