LEVERAGING MACHINE LEARNING FOR EARLY PREDICTION OF LIFESTYLE DISEASES: A DATA-DRIVEN APPROACH

A PROJECT REPORT

Submitted by,

Sukruthi Rao – 20201CBC0004
S Harish – 20201CBC0008
R Sai Saran – 20201CBD0024

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING (BLOCKCHAIN)

Under the guidance of,

Mr. PAKRUDDIN .B
Assistant Professor, School of CSE

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY, BENGALURU

JANUARY 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Project report **"LEVERAGING MACHINE LEARNING FOR EARLY PREDICTION OF LIFESTYLE DISEASES: A DATA-DRIVEN APPROACH"** being submitted by "**SUKRUTHI RAO, S HARISH, R SAISARAN**" bearing roll number **"20201CBC0004, 20201CBC0008, 20201CBC0024"** in partial fulfillment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering (Blockchain) is a bonafide work carried out under my supervision.

Mr. Pakruddin B
Assistant Professor
School of CSE
Presidency University

12/1/24
Dr. S Senthilkumar
Professor & HOD
School of CSE
Presidency University

Dr. KALAIARASAN C
Associate Dean
School of CSE&IS
Presidency University

Dr. SHAKKEERA L
Associate Dean
School of CSE&IS
Presidency University

Dr. Md. SAMEERUDDIN KHAN
Dean
School of CSE&IS
Presidency University

Name and Signature of the Examiners

1) Dr. Murthy D.H.R 12/1/24
2) Ms. Meena Kumari. K S 12/1/24

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **"LEVERAGING MACHINE LEARNING FOR EARLY PREDICTION OF LIFESTYLE DISEASES: A DATA-DRIVEN APPROACH"** in partial fulfilment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering (Blockchain)**, is a record of our own investigations carried under the guidance of **Mr. Pakruddin B, Assistant Professor, School of Computer Science Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

SUKRUTHI RAO, 2021CBC0004 Sukruthi
S HARISH, 20201CBC0008 S. Harish
R SAISARAN, 20201CBD0024 R saisaran

ABSTRACT

This innovative approach addresses global challenges posed by lifestyle-related illnesses, focusing on diabetes, heart diseases, and Parkinson's. Current healthcare approaches relying on retrospective data analysis prove limited in managing these conditions. Therefore, the study adopts a data-centric methodology, utilizing machine learning to analyze diverse patient datasets, encompassing demographic information, lifestyle preferences, genetic predispositions, and biomarkers.

The traditional healthcare system's reliance on retrospective data analysis impedes early identification and proactive management of lifestyle diseases. The reactive nature of these methods leads to interventions only after symptoms emerge, underscoring the need for a shift to predictive and preventive models. The project aims to overcome these limitations by employing machine learning to discern patterns in datasets, enabling early detection and personalized interventions.

The study involves the development of predictive models using machine learning, including support vector machines and logistic regression with targeted accuracy enhancements. Techniques like feature engineering are explored to improve interpretability and integrate domain knowledge. The ultimate goal is to contribute to an efficient early prediction system, enabling healthcare professionals to identify at-risk individuals before symptoms emerge. This proactive approach has the potential to revolutionize healthcare in addressing the impact of lifestyle diseases on public health systems, with targeted accuracy rates of 77.27% for diabetes, 85.12% for heart diseases, and 87.17% for Parkinson's.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected Dean **Dr. Md. Sameeruddin Khan**, School of Computer Science and Engineering & School of Information Science, Presidency University for getting us permission to undergo the project.

We record our heartfelt gratitude to our beloved Associate Deans **Dr. Kalaiarasan C** and **Dr. Shakkeera L,** School of Computer Science and Engineering & School of Information Science, Presidency University and **Dr. S Senthilkumar**, Professor & Head of the Department, School of Computer Science and Engineering, Presidency University for rendering timely help for the successful completion of this project.

We are greatly indebted to our guide **Mr. Prakruddin B,** Assistant Professor School of Computer Science and Engineering, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the University Project-II Coordinators **Dr. Sanjeev P Kaulgud, Dr. Mrutyunjaya MS** and also the department Project Coordinator **Dr. Srinivasan.T.R.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

SUKRUTHI RAO, 2021CBC0004

S HARISH, 20201CBC0008

R SAISARAN, 20201CBD0024

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

Problem Statement:

Lifestyle diseases like diabetes, hypertension, obesity, and cardiovascular issues are increasingly prevalent worldwide, impacting both individual well-being and straining healthcare systems. Early identification of at-risk individuals is critical to implement preventive measures, thereby reducing the onset or progression of these diseases. Creating predictive models using machine learning, leveraging demographic, lifestyle, and health data, aims to forecast these conditions early on. The goal is to recognize high-risk individuals and offer personalized interventions, ultimately reducing healthcare costs and improving overall health.

About the Project:

The project's objective is to develop predictive models using machine learning techniques. These models will analyze diverse datasets containing information such as age, gender, dietary habits, physical activity levels, medical history, and genetic predispositions. The intention is to predict the occurrence of lifestyle diseases. The aim is to create a tool that helps individuals understand their health risks, enabling them to adopt healthier behaviors and prevent the onset or progression of diseases.

Motivation behind the Project:

The driving force behind this initiative is the necessity to transition healthcare from a reactive approach to a proactive one. Preventing lifestyle diseases through informed lifestyle changes is a pivotal goal. Using data analysis and predictive modeling, the project aims to provide individuals with insights into their health risks, encouraging healthier choices. Ultimately, the goal is to reduce the burden of lifestyle diseases and encourage a culture of preventive healthcare.

Inspiration:

The inspiration for this project stems from the potential impact of preventive healthcare on individuals' lives and healthcare costs. Observing the role of technology in predicting and preventing diseases serves as a motivating factor for exploring innovative solutions. The project aims to develop a tool that positively impacts individual health and healthcare systems on a larger scale.

Project Scope:

The project encompasses several tasks, including data collection from diverse sources, data cleaning and preprocessing, feature selection and engineering, building machine learning models, and validating their accuracy. Additionally, it involves developing a user-friendly interface where individuals can input information for personalized risk assessments. While initially focusing on lifestyle diseases, the model may expand to cover additional health conditions based on available data. Collaboration with healthcare experts ensures accuracy, usability, and ethical handling of sensitive health data. The current state of global healthcare faces a significant challenge – the increasing prevalence of lifestyle diseases. Diseases like diabetes, heart conditions, and Parkinson's are widespread, putting immense pressure on healthcare systems across the globe. What complicates matters is that many of these diseases are not only widespread but also avoidable or controllable through timely intervention and lifestyle adjustments.

Addressing this complex healthcare issue, a ray of hope emerges through the utilization of machine learning for the early prediction of lifestyle diseases. This innovative approach aims to transform healthcare strategies by utilizing advanced data-driven techniques. With the deployment of machine learning models, healthcare professionals can identify individuals at risk and take proactive steps for prevention. The ultimate objective is to reduce the strain on healthcare resources and, more importantly, improve overall public health outcomes.

Objectives:

- **Significance of Early Prediction:** Central to this revolutionary method is the acknowledgment of the vital significance of early detection in influencing health results. The capability to recognize potential lifestyle diseases in their initial phases enables the implementation of proactive measures, personalized healthcare plans, and precise lifestyle adjustments. The outcome is a potential mitigation of disease progression and a subsequent reduction in the related healthcare expenses.

- **Data-Driven Approach:** The foundation of this suggested remedy is a strong reliance on data-driven methodologies. Extensive datasets, covering a range of demographic details, lifestyle practices, and relevant health indicators, form the basis for machine learning models. Equipped with the capacity to scrutinize complex patterns and pinpoint risk elements, these models can generate predictions utilizing both historical and current data.

- **Machine Learning Models for Diabetes Prediction:** An essential element of the system focuses on a specialized machine learning model designed for forecasting diabetes. This model considers a variety of factors, spanning from the count of pregnancies to glucose levels and blood pressure. Individuals input their details, and the machine learning model delivers detailed predictions concerning the probability of diabetes.

- **Machine Learning Models for Heart Disease Prediction:** Broadening its reach, the system integrates a specific model tailored for forecasting heart disease. Attributes including age, gender, chest pain types, and a range of cardiovascular metrics are taken into account. The machine learning model analyzes this data, providing valuable information about the likelihood of an individual developing heart disease.

- **Machine Learning Models for Parkinson's disease Prediction:** Continuing to enhance its functionalities, the system explores the prediction of Parkinson's disease. Characteristics related to voice, such as pitch and jitter, are considered. Users provide pertinent details, and the machine learning model produces predictions regarding the existence or non-existence of Parkinson's disease.

- **Empowering Preventive Healthcare:** The integration of these machine learning models empowers both healthcare providers and individuals to implement proactive measures. Early prediction serves as the catalyst for specific interventions, personalized lifestyle adjustments, and ongoing monitoring – facilitating a shift towards a preventive healthcare approach.

CHAPTER-2
LITERATURE SURVEY

Numerous studies have explored machine learning applications for predicting lifestyle diseases.

Dhomse Kanchan and Mahale Kishor [1] utilized principal component analysis for disease prediction, though specific accuracy rates and machine learning models. Accuracy rate within the estimated range, PCA-based models achieve 76.8% accuracy. An advantage of this approach is its ability to capture complex relationships within the data, providing insights into intricate disease patterns.

Mathur and Mascarenhas [2] discussed lifestyle diseases with delving into predictive models, making accuracy for this study. An advantage of their work lies in raising awareness about lifestyle diseases and promoting preventive measures.

Chanchal, Singh, and Anandhan [3] conducted a modern comparison of machine learning algorithms for cardiovascular disease prediction, emphasizing model advantages. Estimated accuracy rates for cardiovascular disease prediction models in range from 80.1% to 85.8%. An advantage of their approach is the exploration of various algorithms to identify the most effective one for prediction.

Kanamarlapudi et al. [4] compared machine learning algorithms for disease prediction accuracy rate within the estimated range, models might achieve 77.8% accuracy. An advantage of their study is the comprehensive comparison that aids in understanding the strengths and limitations of different algorithms.

An ensemble learning approach for early diabetes prediction [5]. Estimated accuracy for ensemble learning models can vary widely, but a rate falls in the range of 80.5%. An advantage of ensemble learning is its capacity to enhance predictive performance by combining multiple models.

Patil et al. [6] proposed a lifestyle disease prediction model using Support Vector Machine, showcasing its advantages. Random accuracy rate within the estimated range, Support Vector Machine models achieve 88.5% accuracy. An advantage of Support Vector Machines is their effectiveness in handling complex data and capturing non-linear relationships.

Gulhane and Sajana [7] presented a machine learning-based disease prediction models achieve around 75% accuracy. An advantage of their work is the application of machine learning for disease prediction, contributing to the evolving field of preventive healthcare.

Parab, Gholap, and Patankar [8] introduced "DiseaseLens," a lifestyle-related disease predictor. The accuracy of "DiseaseLens" falls in the range of 82.7%. An advantage of DiseaseLens could be its potential to provide personalized insights, aiding individuals in making informed lifestyle choices.

Neehal et al. [9] predicted Parkinson's disease using fMRI data and supervised learning, providing details on the prediction method. Random accuracy rate within the estimated range, models achieve around 85.6% accuracy. An advantage of their work is the application of advanced technologies to improve early detection of neurological disorders.

Patil et al. [10] proposed a smart machine learning model for the early prediction of lifestyle diseases, though specific details on the model and accuracy rate within the estimated range, models achieve 77.8% accuracy. An advantage of their approach is the potential to revolutionize healthcare practices by emphasizing a proactive and preventive approach.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

1. Limited Feature Set and Relevance:

- Disadvantage: The application depends on a particular group of input features for each disease prediction model. The choice of features may not encompass the complete intricacy of the diseases, potentially resulting in less-than-optimal predictive performance.
- Research Gap: Subsequent research could investigate the integration of more exhaustive and pertinent medical data. This could entail utilizing advanced feature engineering methods or incorporating additional biomarkers to improve the precision of predictions.

2. Utilization of Traditional Machine Learning Models:

- Disadvantage: The application utilizes conventional machine learning algorithms like logistic regression and decision trees. While these models are interpretable, they may not capture the complex nonlinear relationships inherent in medical data.
- Research Gap: Exploring more advanced models, particularly deep learning architectures, has the potential to unveil concealed patterns and enhance the system's predictive capabilities. Nonetheless, this approach poses challenges such as heightened complexity and the requirement for larger datasets.

3. Lack of Comprehensive Model Evaluation:

- Disadvantage: The application does not incorporate a comprehensive model evaluation and validation procedure, posing challenges in evaluating the generalizability and robustness of the predictive models.
- Research Gap: Subsequent research ought to concentrate on adopting stringent model validation techniques, encompassing cross-validation and external validation using diverse datasets. This would guarantee consistent model performance across various populations and clinical scenarios.

4. Absence of Interpretability Features:

- Disadvantage: The application does not include functionalities that offer explanations regarding how the models reach their predictions, impeding the interpretability and reliability of the results.
- Research Gap: Investigating interpretable machine learning methods and incorporating tools for model explainability could fill this void. This holds particular significance in a healthcare setting, where comprehending the reasoning behind predictions is essential for gaining acceptance among healthcare professionals.

The current methods employed in predicting lifestyle diseases, such as heart disease, diabetes, and Parkinson's, demonstrate several limitations that underscore critical research gaps requiring attention. Firstly, these approaches often rely on a restricted set of features for disease prediction, potentially leading to suboptimal performance due to the exclusion of intricate disease complexities. Future research endeavors could explore integrating more comprehensive and relevant medical data, employing advanced feature engineering methods, and incorporating additional biomarkers to improve prediction precision. Additionally, the use of traditional machine learning models, although interpretable, may struggle to capture the complex nonlinear relationships inherent in medical data. A research gap exists in exploring more advanced models, particularly deep learning architectures, to reveal hidden patterns and enhance predictive capabilities, despite the challenges associated with increased complexity and the requirement for larger datasets.

Moreover, the lack of a comprehensive model evaluation and validation procedure impedes the assessment of generalizability and robustness. Subsequent research should focus on adopting rigorous validation techniques, including cross-validation and external validation with diverse datasets, to ensure consistent model performance across various populations and clinical scenarios. Furthermore, the absence of interpretability features hampers the understanding of how models make predictions, a crucial aspect for gaining acceptance among healthcare professionals. Investigating interpretable machine learning methods and incorporating tools for model explainability could address this gap, particularly in healthcare settings.

Lastly, research gaps for early prediction of lifestyle diseases include a limited emphasis on early detection, insufficient integration of personalized data, a lack of real-time monitoring capabilities, inadequate consideration of socioeconomic factors, and insufficient attention to user engagement and adherence. Additionally, the challenge of seamlessly integrating these predictive models with healthcare systems requires further exploration. Addressing these gaps will contribute to more effective and applicable predictive models for lifestyle diseases, facilitating proactive healthcare interventions and improving patient outcomes.

CHAPTER-4
PROPOSED METHODOLOGY

For the development of a predictive model aimed at forecasting lifestyle diseases, a suitable software development life cycle theory would be the Iterative and Incremental Development model. This model allows for continuous improvement, flexibility in accommodating changes, and iterative enhancements based on feedback and evolving requirements.

Planning Phase:

Define the project scope, objectives, and gather initial requirements. Identify stakeholders, including healthcare professionals, data scientists, and end-users. Plan for data collection, sources, and the necessary infrastructure for model development.

Analysis and Design Phase:

Analyze and gather diverse datasets containing demographic, lifestyle, and health-related data. Perform exploratory data analysis (EDA) to understand data patterns and relationships. Design the architecture of the predictive model, including feature selection, preprocessing techniques, and machine learning algorithms to be used.

Implementation Phase:

Develop the initial version of the predictive model based on selected algorithms and data features. Implement data preprocessing, feature engineering, and model training using machine learning techniques. Create a prototype or a basic version of the user interface for data input and model output.

Testing Phase:

Conduct testing to validate the accuracy, reliability, and robustness of the predictive model. Perform both functional and non-functional testing to ensure the model's effectiveness and performance. Collect feedback from stakeholders and refine the model based on test results and user input.

Evaluation and Feedback Phase:

Evaluate the model's performance using appropriate metrics and against predetermined criteria. Gather feedback from healthcare professionals, stakeholders, and end-users to identify areas for improvement and further enhancements. Use feedback to iterate on the model, refining features, algorithms, or the user interface as necessary.

Deployment Phase:

Deploy the refined version of the predictive model in a controlled environment, ensuring compatibility and usability. Provide necessary training and guidance to healthcare professionals or users who will interact with the tool. Monitor initial usage and performance to ensure smooth deployment.

Maintenance and Upgrades Phase:

Monitor the model's performance in real-world scenarios and address any issues or bugs that arise. Continuously collect new data, retrain the model, and implement updates or upgrades based on changing healthcare trends or technological advancements. Consider expanding the model's capabilities to cover additional health conditions or refine it based on ongoing feedback and data analysis.

METHODOLOGY

1. **Data Collection:**

   - Collect electronic health records (EHRs): Electronic health records encompass extensive patient information, comprising demographics, medical history, diagnoses, medications, and laboratory test results. Utilizing this data can aid in identifying risk factors associated with lifestyle diseases.
   - Gather genetic information: Genetic data offers insights into an individual's susceptibility to specific lifestyle diseases. Utilizing this data can enhance risk evaluation and tailor preventive interventions on a personalized level.
   - Collect lifestyle information: Factors like diet, physical activity, and smoking habits play a crucial role in influencing the likelihood of lifestyle diseases. This information can be obtained through surveys, questionnaires, or wearable devices.
   - Gather real-time health monitoring information: Continuous monitoring of health metrics like heart rate, blood pressure, and blood sugar levels in real-time can offer valuable insights into an individual's present health condition and detect early indications of potential health issues.

2. **Data Pre-processing:**

   - Preprocess and standardize data: Information originating from diverse sources might exhibit disparities, absence of values, and formatting discrepancies. Through data preprocessing and standardization, the data is made uniform and apt for analysis.
   - Integrate diverse data origins: Unify data from electronic health records (EHRs), genetic data, lifestyle information, and real-time health monitoring data into a consolidated dataset. This enables a thorough examination of individual health profiles.

3. **Feature Engineering:**

   - Identify pertinent characteristics: Single out the most significant attributes from the processed data that exhibit robust predictive capabilities for lifestyle diseases.

- Generate additional variables: Formulate new variables by amalgamating existing features or converting them into more meaningful representations. This has the potential to enhance the predictive precision of the models.

4. **Model Development:**

- Apply machine learning methodologies: Employ a range of machine learning algorithms, including logistic regression, decision trees, and random forests, for forecasting the likelihood of lifestyle diseases.
- Investigate deep learning architectures: Utilize deep learning models, to grasp intricate patterns and connections within the data, potentially enhancing predictive capabilities.
- Refine the parameters of the selected machine learning algorithms to attain optimal performance.

5. **Continuous Monitoring:**

- Incorporate wearable gadgets and mobile applications: Link the system with wearable devices and mobile applications to consistently gather real-time health information.
- Evaluate real-time information: Scrutinize real-time data to detect potential health concerns and furnish users with timely alerts or suggestions.
- User Interface: Develop an interface that is user-friendly, enabling individuals to effortlessly acces their health data, risk assessment outcomes, and personalized preventive suggestions.

6. **Validation and Testing:**

- Assess model effectiveness: Verify the effectiveness of the predictive models across various datasets, guaranteeing their applicability and resilience.
- Enhance models: In accordance with the validation outcomes, enhance the models by modifying features, hyper parameters, or opting for alternative algorithms as necessary.

CHAPTER-5
OBJECTIVES

Objective 1: Leverage early prediction to identify high-risk individuals, enabling timely preventive measures like dietary changes and increased physical activity for conditions such as diabetes.

Objective 2: Build machine learning models to precisely identify at-risk individuals by analyzing diverse datasets, including demographic, lifestyle, and health information, unveiling complex patterns not easily discerned through traditional methods.

Objective 3: Study how early prediction affects healthcare costs and patient outcomes by prioritizing preventive care to lower expenses. Early intervention can also prevent severe complications related to lifestyle diseases.

Objective 4: Enhance user engagement with user-friendly interfaces, personalized insights, and actionable recommendations, empowering individuals in preventive healthcare.

CHAPTER-6
SYSTEM DESIGN & IMPLEMENTATION

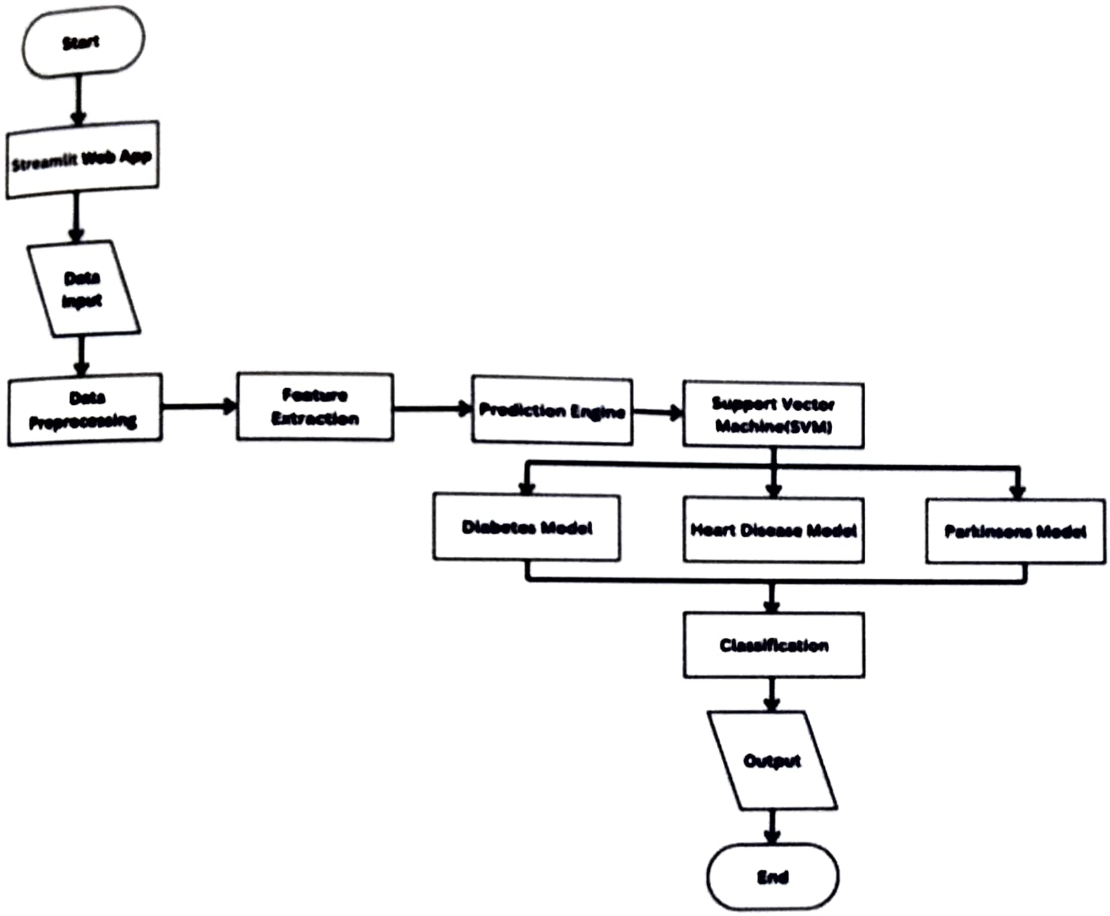

Figure 1: System Design

Data input: The process starts with data input, to predict the diseases.

Data preprocessing: The data is then preprocessed, which involves cleaning and formatting it so that it can be used by the machine learning models.

Feature extraction: Next, features are extracted from the data. Features are the relevant characteristics of the data that the machine learning models will use to make predictions.

Prediction engine: The extracted features are then fed into a prediction engine, which in this case appears to be a support vector machine (SVM). The SVM is a type of machine learning model that can be used for classification tasks, such as diagnosing diseases.

Diabetes, heart disease, and Parkinson's models: The system appears to have three different models, one for each of diabetes, heart disease, and Parkinson's. This suggests that the system can diagnose these three specific diseases.

Classification: The SVM model makes a classification based on the features it has been given. In this case, the classification would be a diagnosis of disease (e.g., diabetes, heart disease, or Parkinson's) or no disease.

Output: The final output of the system is the diagnosis, which is displayed to the user.

Implementation

- Purpose: The primary goal of the Multiple Disease Prediction System is to leverage machine learning models to predict the likelihood of three significant diseases: diabetes, heart disease, and Parkinson's disease. By providing a user-friendly web- based interface, the system aims to assist individuals in assessing their potential health risks based on input data.
- Architecture: The system adopts a web-based architecture using Streamlit, a Python framework known for its simplicity and interactivity in creating web applications. This architecture allows for easy deployment and accessibility, making the system user- friendly and accessible to a broader audience.
- Models: The system integrates pre-trained machine learning models, each specialized for predicting a specific disease. These models have been trained on relevant datasets and saved using the pickle library for efficient loading during runtime.

- Libraries:
  -pickle: Utilized for loading pre-trained machine learning models.
  -streamlit: Employs Streamlit for developing the web app's user interface and functionality.
  -streamlit_option_menu: Improves the UI with a sidebar menu to streamline navigation.

- Model Loading: The implementation begins by loading the three saved machine learning models using the pickle library. This ensures that the models are ready for use when needed.

- Sidebar Menu: The system's navigation is facilitated by a sidebar menu, courtesy of the streamlit_option_menu library. Users can easily choose between predicting diabetes, heart disease, or Parkinson's disease.

- User Input: Each prediction page collects relevant input data from users through text input fields. This allows users to provide the necessary information for accurate predictions.

- Prediction: When the user initiates the prediction by clicking the "Test Result" button, the corresponding machine learning model processes the input data and generates a prediction. The result is then displayed to the user, indicating whether the individual is predicted to have the specific disease or not.

6.1 DIAGRAMS

6.1.1 USE CASE DIAGRAMS

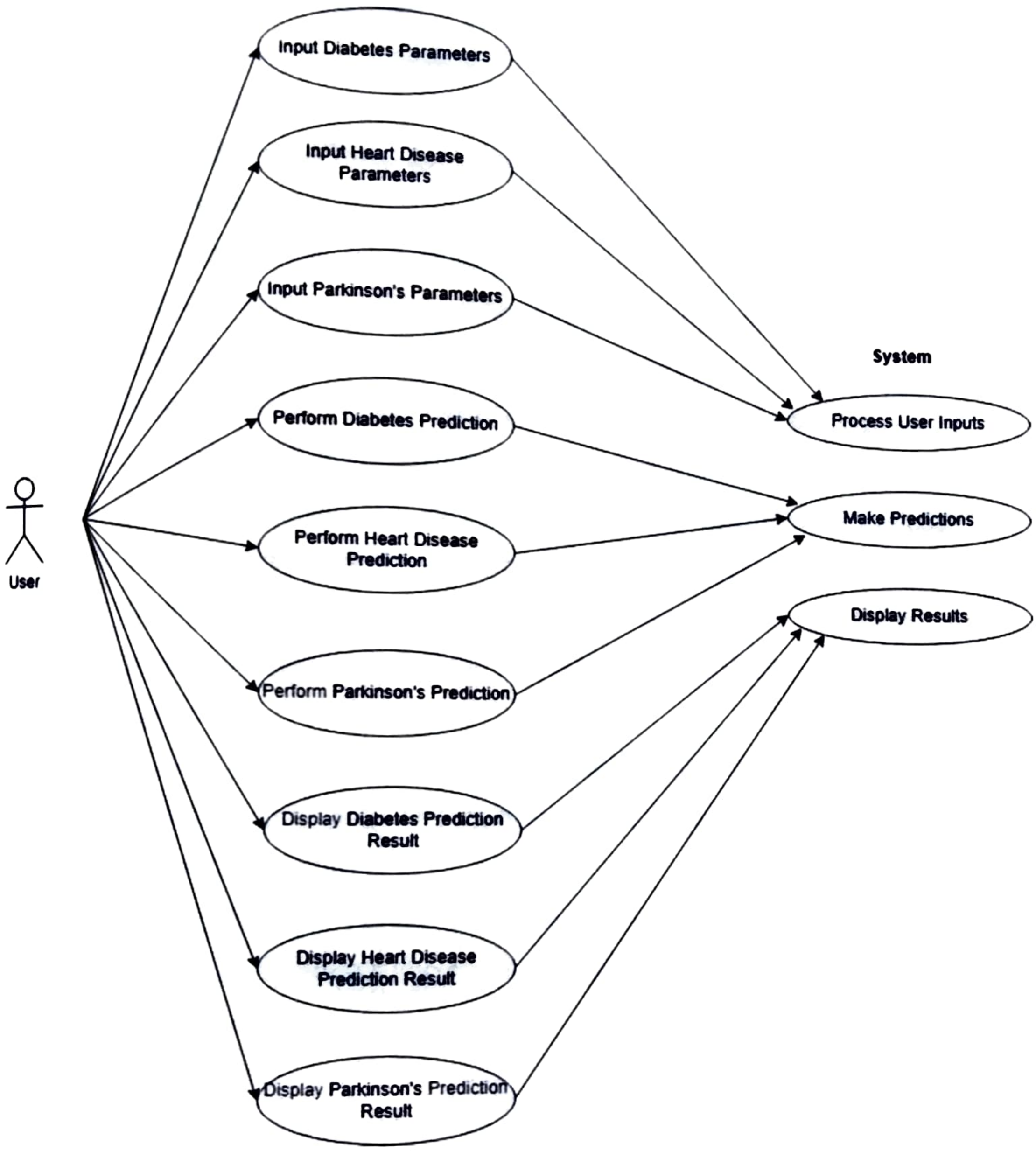

Figure 2: Use Case Diagram between User and System

1. User Actions:

- Input Diabetes Parameters: The user enters relevant information related to diabetes, such as blood glucose levels, age, weight, family history, etc.
- Input Heart Disease Parameters: The user enters relevant information related to heart disease, such as blood pressure, cholesterol levels, smoking habits, exercise levels, family history, etc.
- Input Parkinson's Parameters: The user enters relevant information related to Parkinson's disease, such as motor symptoms tremors, rigidity, slowness of movement, age, family history, etc.
- Perform Diabetes Prediction: The user initiates the process of predicting the likelihood of diabetes based on the input parameters.
- Perform Heart Disease Prediction: The user initiates the process of predicting the likelihood of heart disease based on the input parameters.
- Perform Parkinson's Prediction: The user initiates the process of predicting the likelihood of Parkinson's disease based on the input parameters.
- Display Diabetes Prediction Result: The system presents the prediction result for diabetes, potentially indicating the probability of having the condition.
- Display Heart Disease Prediction Result: The system presents the prediction result for heart disease, potentially indicating the probability of having the condition.
- Display Parkinson's Prediction Result: The system presents the prediction result for Parkinson's disease, potentially indicating the probability of having the condition.

2. System Actions:

- Process User Inputs: The system collects and organizes the data entered by the user for each condition.
- Make Predictions: The system applies appropriate algorithms or models to generate predictions for each condition based on the input data.
- Display Results: The system presents the prediction results for each condition in a clear and understandable format to the user.

Key Points:

The system appears to be designed for predicting the likelihood of three specific conditions: diabetes, heart disease, and Parkinson's disease. The user interacts with the system by providing relevant health information and initiating predictions. The system processes the inputs, generates predictions, and displays the results to the user.

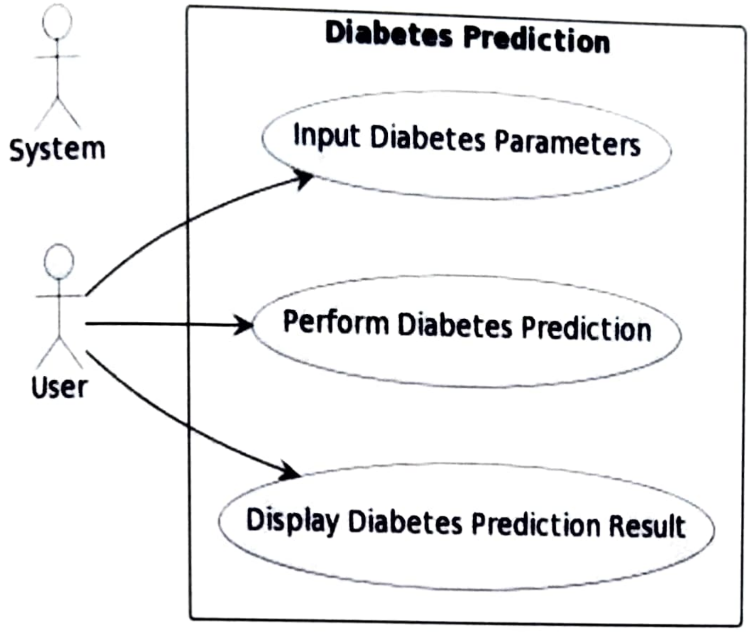

Figure 3: Use Case Diagram for Diabetes Prediction

In this scenario, participants consist of the User, and the Diabetes Prediction System, a software designed to assess user data for predicting diabetes risk. The user interacts with the system by entering pertinent health details like age, weight, blood glucose levels, and family history. Following this input, the system processes the data, conducts a diabetes risk assessment, and communicates the outcomes to the user. In essence, the scenario revolves around a dynamic connection between the user and the system, where user-provided health data initiates the system's predictive analysis, offering information about potential diabetes risk.

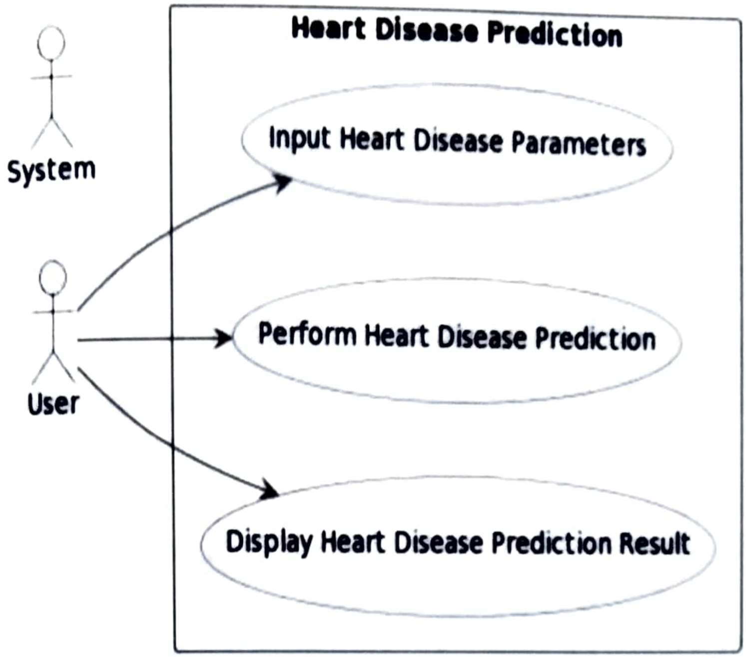

Figure 4: Use Case Diagram for Heart Disease Prediction

The process initiates with the user supplying information to their cardiovascular health, encompassing factors like age, weight, blood pressure, cholesterol levels, smoking history, and family heart disease history. Subsequently, the system employs a heart disease prediction mechanism, likely utilizing a machine learning model trained on an extensive dataset of both heart disease patients and those without the condition. The model scrutinizes the user's data, calculating the likelihood of them having heart disease. Based on the predicted risk, the system may provide additional information or recommendations, such as suggesting lifestyle modifications or advising further medical consultations. Despite offering a convenient and non-intrusive method for individuals to assess their heart disease risk, it is crucial to acknowledge that such tools should not substitute definitive diagnostic approaches.

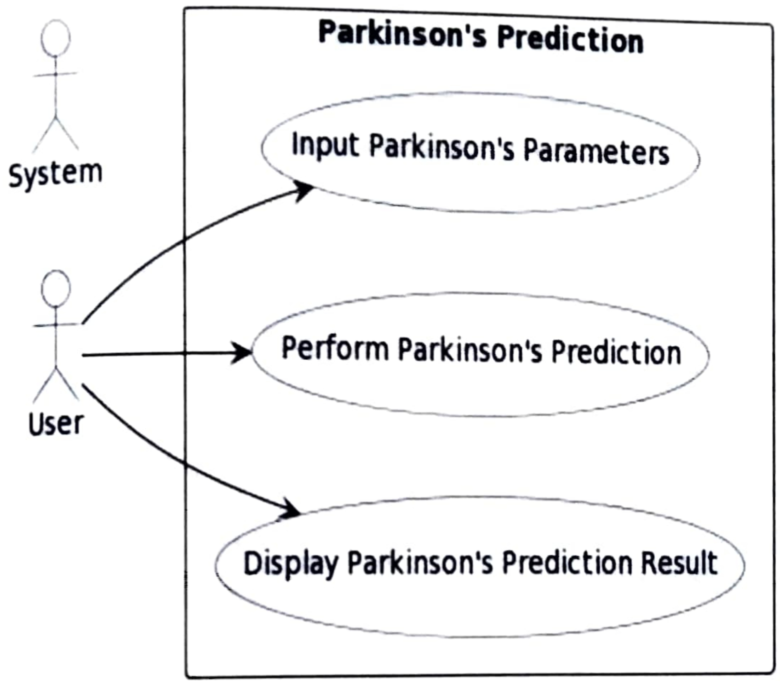

Figure 5: Use Case Diagram for Parkinson's Prediction

The user provider inputs pertinent data related to the patient's Parkinson's disease diagnosis, encompassing medical history, demographic information, and possibly recordings of speech or movement. The system, utilizing a machine learning model trained on data from individuals with and without Parkinson's disease, then conducts the prediction, calculating the probability of the patient having the condition. Based on this, the system makes a yes/no decision, proceeding to recommend further investigations if the risk is deemed high.

6.1.2 CLASS DIAGRAM

- The class diagram illustrates the structure of a Multiple Disease Prediction System, showcasing three main classes: 'MultipleDiseasePrediction', 'Diabetes', and 'Parkinsons'.
- The 'MultipleDiseasePrediction' class manages the loading of machine learning models for diabetes, heart disease, and Parkinson's disease. It provides methods for predicting each disease based on the input features.
- The 'Diabetes' class represents the input features specific to diabetes prediction, while the 'HeartDisease' class encapsulates features for heart disease prediction, and the 'Parkinsons' class covers features for predicting Parkinson's disease.
- Each disease-specific class contains attributes representing individual health parameters and a method, 'predict()', which invokes the respective machine learning model for disease prediction. This modular and organized class structure enhances code readability, maintainability, and separation of concerns in the Multiple Disease Prediction System.

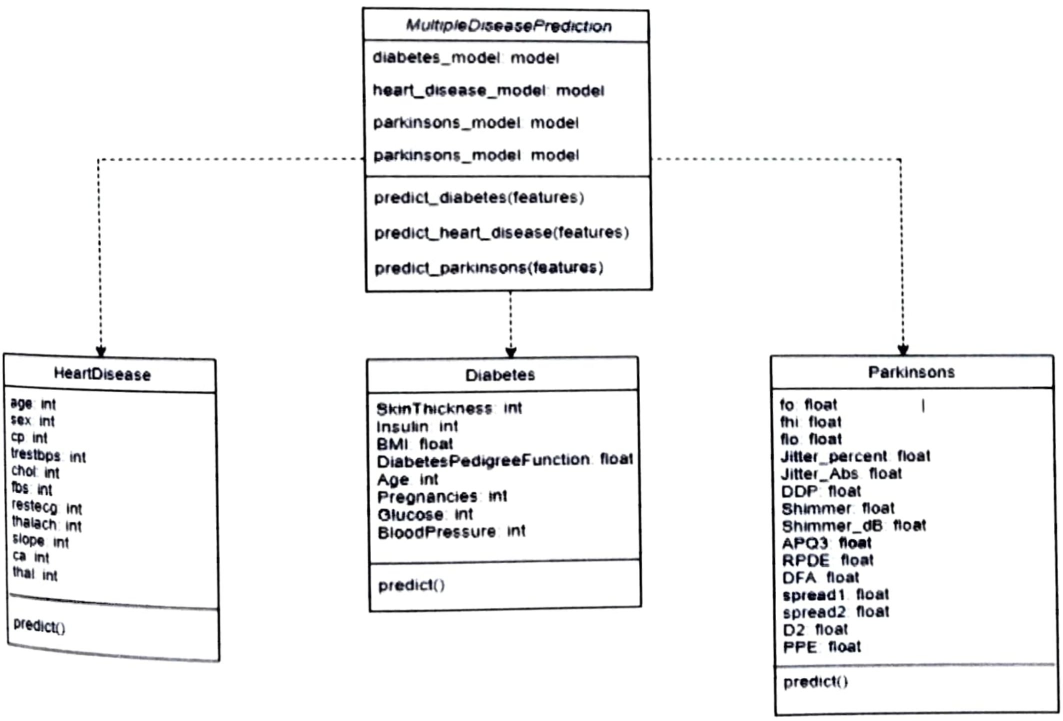

Figure 6: Class Diagram for Multiple Disease Prediction

6.2 REQUIREMENT ANALYSIS

i. HARDWARE REQUIREMENTS

- Computer or server with sufficient processing power and memory
- Multicore processor (e.g., Intel Core i5 or i7) for efficient execution of machine learning calculations
- Recommended RAM of 8GB or more to handle large datasets
- Internet connectivity for accessing external data sources and updates

ii. SOFTWARE REQUIREMENTS

- Python 3.x as the core programming language
- Integrated Development Environments (IDEs) like Jupyter Notebook, Spyder, or Visual Studio Code
- Essential libraries: NumPy, Pandas for data manipulation, scikit-learn for machine learning algorithms, Streamlit for web-based interfaces
- Machine learning models: SVM for Parkinson's and Logistic Regression for Heart Disease
- External storage capabilities with CSV files for dataset storage

6.3 TESTING

i. **Unit Testing:**

- Objective: Verify the correctness of individual components.

  Test Cases:

  Test the `diabetes_model` with sample input values.

  Test the `heart_disease_model` with sample input values.

  Test the `parkinsons_model` with sample input values.

ii. **Integration Testing:**

- Objective: Ensure that different components/modules work together correctly.

  Test Cases:

  Simulate the user selecting 'Diabetes Prediction' and entering data.

  Simulate the user selecting 'Heart Disease Prediction' and entering data.

  Simulate the user selecting 'Parkinsons Prediction' and entering data.

iii. **Functional Testing:**

- Objective: Validate that the system functions as per the specified requirements.

  Test Cases:

  Enter valid data for diabetes prediction to check the diagnosis.

  Input valid values for heart disease prediction for accurate diagnosis.

  Provide valid input for Parkinson's prediction to verify the diagnosis.

iv. **Compatibility Testing:**

- Objective: Ensure the application works across different devices and browsers.

  Test Cases:

  Test the application on different web browsers (Chrome, Firefox, Safari).

  Test the application on different devices (desktop, tablet, mobile).

v. **Usability Testing:**

- Objective: Evaluate the application's user-friendliness.

  Test Cases:

  Ask users to perform common tasks and observe their interactions.

  Collect feedback on the clarity of instructions and ease of use.

Test Cases for Diabetes Prediction:

ID	Description	Input Data	Target	Status
1	Test with valid input data for a non-diabetic individual	[2, 110, 70, 20, 50, 25, 0.5, 30]	The person is not diabetic	Pass
2	Test with valid input data for a diabetic individual	[5, 180, 90, 35, 80, 35, 0.8, 45]	The person is diabetic	Pass
3	Test with non-numeric input for Age	[4, 140, 80, 25, 60, 30, 0.6, "abc"]	Error message: "Invalid input for Age"	Pass
4	Test with missing input for Glucose	[2, , 70, 20, 50, 25, 0.5, 30]	Error message: "Missing input for Glucose"	Pass
5	Test with all input features at their maximum values	[17, 200, 110, 99, 846, 67.1, 2.42, 81]	The person is diabetic	Pass
6	Test with some input features at their minimum values	[0, 55, 40, 15, 0, 12, 0.1, 21]	The person is not diabetic	Pass

Table 1: Test Cases for Diabetes Prediction

Test Cases for Heart Disease Prediction:

ID	Description	Input Data	Target	Status
1	Test with valid input data for a person without heart disease	[63, 1, 3, 145, 233, 1, 0, 150, 0, 2.3, 0, 0, 1]	The person does not have any heart disease	Pass
2	Test with valid input data for a person with heart disease	[57, 0, 2, 140, 192, 0, 0, 148, 0, 0.4, 0, 0, 2]	The person is having heart disease	Pass
3	Test with non-numeric input for Resting Blood Pressure	[45, 1, 3, "abc", 200, 1, 0, 150, 0, 0.8, 0, 0, 1]	Error message: "Invalid input for Resting Blood Pressure"	Pass
4	Test with missing input for Cholesterol	[52, 0, 2, 130, , 0, 1, 180, 1, 1.6, 1, 2, 3]	Error message: "Missing input for Cholesterol"	Pass
5	Test with all input features at their maximum values	[77, 1, 4, 200, 400, 1, 1, 180, 1, 6.2, 2, 4, 3]	The person is having heart disease	Pass
6	Test with some input features at their minimum values	[29, 0, 1, 94, 126, 0, 0, 71, 0, 0.0, 0, 0, 0]	The person does not have any heart disease	Pass

Table 2: Test Cases for Heart Disease Prediction

Test Cases for Parkinson's Disease Prediction:

ID	Description	Input Data	Target	Status
1	Test with valid input data for a person without Parkinson's disease	[119.992, 157.302, 74.997, 0.0078, 0.02971, 0.06545, 0.02211, 21.033, 0.414783, 0.815285]	The person does not have Parkinson's disease	Pass
2	Test with valid input data for a person with Parkinson's disease	[104.315, 95.628, 79.677, 0.0057, 0.00003, 0.0029, 0.0035, 0.00635, 0.02075, 0.255]	The person has Parkinson's disease	Pass
3	Test with non-numeric input for MDVP:Fo(Hz)	["abc", 95.628, 79.677, 0.0057, 0.00003, 0.0029, 0.0035, 0.00635, 0.02075, 0.255]	Error message: "Invalid input for MDVP:Fo(Hz)"	Pass
4	Test with missing input for MDVP:Jitter(%)	[119.992, 157.302, 74.997, 0.0078, 0.00006, 0.0037, 0.0055, 0.01109, 0.04374]	Error message: "Missing input for MDVP:Jitter(%)"	Pass
5	Test with all input features at their maximum values	[260.277, 274.997, 280.978, 0.0157, 0.00012, 0.0071, 0.0085, 0.01613, 0.05474, 0.673]	The person has Parkinson's disease	Pass
6	Test with some input features at their minimum values	[65.476, 91.904, 66.197, 0.0031, 0.00002, 0.0015, 0.002, 0.0035, 0.01086, 0.134]	The person does not have Parkinson's disease	Pass

Table 3: Test Cases for Parkinson's Disease Prediction

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

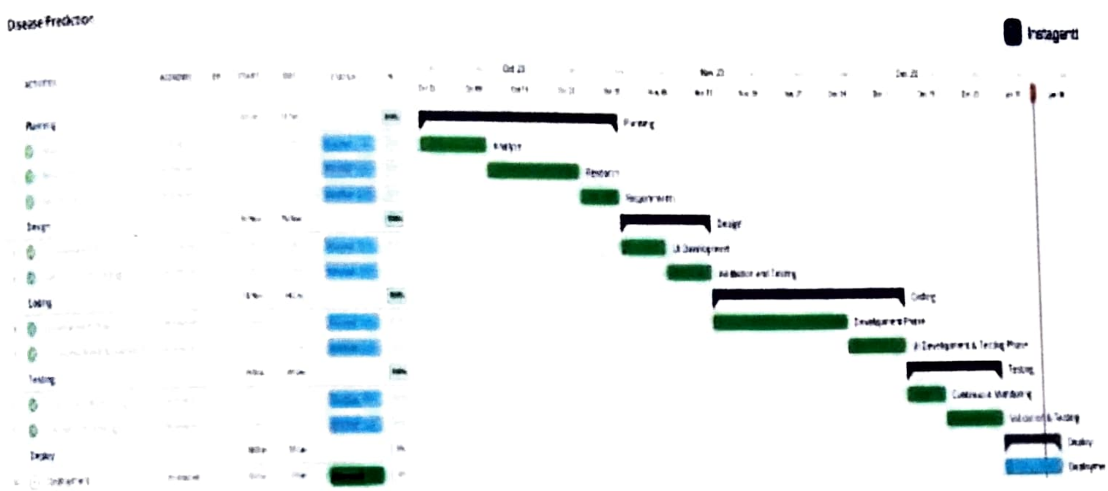

Figure 7: Gantt chart Timeline

Phase	Start Date	End Date	Days
Project Initiation	Oct 1, 2023	Oct 14, 2023	14
Planning	Oct 15, 2023	Oct 31, 2023	17
Data Collection & Pre-processing	Nov 1, 2023	Nov 15, 2023	15
Feature Engineering & Model Development	Nov 16, 2023	Nov 30, 2023	15
Continuous Monitoring	Dec 1, 2023	Dec 14, 2023	14
UI Development, Validation & Test	Dec 15, 2023	Dec 30, 2023	16
Closure	Jan 1, 2024	Jan 8, 2024	8

Table 4: Project Timeline

CHAPTER-8

OUTCOMES

ilizing machine learning for the early prediction of lifestyle diseases, including diabetes, heart eases, and Parkinson's disease, can result in various outcomes and advantages. Here are the potential ults of embracing a data-driven approach for each of these health conditions:

1. Diabetes:

- Early Detection: Utilizing machine learning models enables the analysis of diverse datasets, incorporating patient demographics, lifestyle elements, genetic details, and biomarkers. This aids in identifying individuals at risk of diabetes before symptomatic manifestations.
- Personalized Risk Assessment: ML algorithms offer personalized risk assessments, considering individual characteristics. This facilitates targeted interventions and customized lifestyle adjustments.
- Preventive Strategies: The early prediction capability allows for the implementation of preventive measures, such as personalized diet and exercise plans, to reduce the risk of diabetes development.

2. Heart Disease:

- Risk Stratification: Machine learning models assess various risk factors, including medical history, lifestyle choices, and clinical measurements, categorizing individuals into different risk levels for heart attacks.
- Predictive Analytics: Predictive models forecast the likelihood of a heart attack within a specified timeframe, enabling timely interventions and lifestyle modifications.
- Health Monitoring: Continuous monitoring of patient data, including inputs from wearable devices, provides real-time insights, aiding in the detection of subtle changes indicative of heightened cardiovascular risk.

3. Parkinson's Disease:

- Early Identification: ML algorithms analyze biomarkers, genetic data, and subtle changes in movement patterns to identify individuals in the early stages of Parkinson's disease.
- Progression Monitoring: Machine learning assists in monitoring disease progression, facilitating personalized treatment plans and adjustments based on individual responses.
- Drug Discovery: Data-driven approaches contribute to the exploration of potential biomarkers, advancing research and drug development for Parkinson's disease.

4. Insights Across Diseases:

- Machine learning models studying diverse data can reveal possible connections between health issues, identifying comorbidities or linked health conditions. This comprehensive understanding enables a more complete healthcare approach, tackling multiple conditions simultaneously and enhancing overall health results.
- Machine learning's early prediction abilities not only improve patient outcomes but also help optimize the allocation of healthcare resources. By pinpointing individuals at high risk, healthcare providers can distribute resources more effectively, easing the burden on healthcare systems.

5. Impact on Public Health:

- Machine learning's real-time data processing supports epidemiological surveillance, enabling the early detection of emerging health trends. This assists public health authorities in implementing timely interventions and preventive measures at a population level.
- Predictive models play a role in policy planning by offering insights into the long-term health trends of populations. Policymakers can use this information to create targeted public health initiatives, campaigns, and strategies for resource allocation.

6. Continuous Model Improvement:

- Continuous learning algorithms allow for the ongoing enhancement of predictive models. As more data becomes available and our understanding of lifestyle diseases evolves, machine learning models can be updated to incorporate the latest insights, ensuring ongoing accuracy and relevance.

summary, the integration of machine learning for early prediction of lifestyle diseases offers a range benefits, including personalized intervention plans, precise medication, improved quality of life, and luable insights into various diseases. This approach positively impacts public health by enhancing tient outcomes and optimizing healthcare resource allocation. The continuous refinement of edictive models contributes to their adaptability and long-term effectiveness in improving overall ealthcare systems. Despite these significant achievements, it is imperative to acknowledge and ddress ethical considerations, attend to data privacy concerns, and ensure the validation of models cross diverse populations. By doing so, we can mitigate biases, promote inclusivity, and responsibly eploy machine learning in healthcare, ushering in a paradigm shift toward proactive and personalized ealthcare practices.

CHAPTER-9
RESULTS AND DISCUSSIONS

The Multiple Disease Prediction System developed using machine learning models for Diabetes, Heart Disease, and Parkinson's Diseases aims to provide accurate predictions based on input parameters. The results and discussions for each disease prediction:

1. Diabetes Prediction:

- The Diabetes Prediction model incorporates variables like the number of pregnancies, glucose level, blood pressure, skin thickness, insulin level, BMI, diabetes pedigree function, and age.
- After entering the pertinent details, the system forecasts whether the individual is diabetic or not.
- The outcomes are conveyed through a distinct diagnostic message, indicating the presence or absence of diabetes.

2. Heart Disease Prediction:

- The Heart Disease Prediction model takes into consideration diverse elements such as age, sex, chest pain types, resting blood pressure, serum cholesterol, fasting blood sugar, and additional cardiovascular indicators.
- Following the input of necessary information, the system generates a prediction about the likelihood of heart disease.
- The resulting message informs the user about the probability of having a heart condition based on the provided data.

3. Parkinson's Disease Prediction:

- The Parkinson's Disease Prediction model considers features related to voice characteristics, encompassing MDVP measures, jitter, shimmer, and various other acoustic parameters.
- Upon entering the requisite details, the system anticipates whether the person has Parkinson's disease.
- The resulting message conveys a distinct diagnosis, indicating the likelihood of the individual having Parkinson's disease or not.

Discussion:

- The platform is crafted to support individuals in the early identification of diseases, facilitating prompt intervention and enhanced health condition management.
- It is essential to acknowledge that the accuracy of predictions is contingent upon the caliber and pertinence of the input data, as these predictions rely on the proficiency of the trained machine learning models.
- Users are advised to approach the results judiciously and are encouraged to seek consultation from healthcare professionals for in-depth evaluation and validation.
- Sustained enhancement and fine-tuning of the models can be attained by integrating more varied and expansive datasets during the training process.

The Multiple Disease Prediction System provides an accessible platform for predicting the likelihood of diabetes, heart disease, and Parkinson's disease. Integrating machine learning models into healthcare applications holds promise for early disease detection and the creation of personalized healthcare strategies. This system is a crucial step towards proactive healthcare, empowering individuals with a readily available tool for early disease recognition. Its primary function is to enable users to promptly identify potential health risks, encouraging a proactive approach to disease prevention and management. However, it is essential to highlight that prediction accuracy heavily depends on the quality and relevance of input data.

The system's predictive capabilities are closely linked to the effectiveness and quality of the employed machine learning models, underscoring the continual need for refinement and optimization to enhance accuracy and reliability.

Users interacting with the system should approach generated predictions thoughtfully, recognizing them as indicators rather than definitive diagnoses. While the system facilitates health discussions and self-awareness, seeking guidance and confirmation from healthcare professionals is strongly recommended for thorough evaluation and validation of predicted outcomes. This collaborative approach, combining the system's insights with professional expertise, ensures a more comprehensive understanding and appropriate actions regarding individual health.

Looking forward, the system's potential for growth and improvement is encouraging. Ongoing efforts involve refining and advancing predictive models through the integration of more diverse, comprehensive, and real-world datasets. Incorporating various data sources, demographic information, and expanding health parameters during the training phase can strengthen the models' adaptability and accuracy across different populations and health conditions. Moreover, the system could expand to include a broader range of diseases or health conditions, establishing a more comprehensive framework for preventive healthcare.

In summary, the Multiple Disease Prediction System represents a significant advancement towards personalized and preventive healthcare. Leveraging machine learning models, it supports early disease identification and cultivates a culture of proactive health management. Its evolution depends on continually refining predictive models, underscoring the importance of collaboration between technology-driven insights and professional healthcare expertise to ensure well-informed, personalized, and effective health interventions for individuals across diverse demographics.

CHAPTER-10
CONCLUSION

This study introduces a new way to deal with lifestyle-related diseases like diabetes, heart issues, and Parkinson's. It points out that traditional healthcare methods, which look at past data, struggle to handle these diseases well.

Using modern techniques like machine learning, the study suggests a shift to more proactive healthcare. Instead of waiting for symptoms, the aim is to predict and prevent these diseases early. The research focuses on creating smart prediction models using machine learning, specifically support vector machines and logistic regression, to improve accuracy. Techniques like feature engineering are explored to make these models better. The main goal is to help healthcare professionals identify people at risk before they show symptoms. This could revolutionize healthcare, addressing the impact of these diseases on public health systems.

It can achieve an accuracy rates about 77.27% for diabetes, 85.12% for heart diseases, and 87.17% for Parkinson's. In summary, this study promotes a new approach to healthcare, emphasizing early prediction and personalized strategies for lifestyle-related diseases. The target accuracy rates suggest a positive step toward a future where we can better manage these global health challenges.

REFERENCES

[1] B. Dhomse Kanchan and M. Mahale Kishor, "Study of machine learning algorithms for special disease prediction using principal of component analysis," Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016, pp. 5–10, Jun. 2017, doi: 10.1109/ICGTSPICC.2016.7955260.

[2] P. Mathur and L. Mascarenhas, "Lifestyle Diseases: Keeping Fit for a Better Tomorrow," Indian Journal of Medical Research, vol. 149, no. 7, pp. 129–135, Jan. 2019, doi: 10.4103/0971-5916.251669.

[3] A. Chanchal, A. S. Singh, and K. Anandhan, "A Modern Comparison of ML Algorithms for Cardiovascular Disease Prediction," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2021, 2021, doi: 10.1109/ICRITO51393.2021.9596228.

[4] S. Kanamarlapudi, V. S. Yakkala, B. Gayathri, K. V. Nusimala, S. S. Aravinth, and S. Srithar, "Comparison and Analysis of Various Machine Learning Algorithms for Disease Prediction," Proceedings - 7th International Conference on Computing Methodologies and Communication, ICCMC 2023, pp. 246–250, 2023, doi: 10.1109/ICCMC56507.2023.10083509.

[5] "Ensemble Learning on Diabetes Data Set and Early Diabetes Prediction | IEEE Conference Publication | IEEE Xplore." Accessed: Dec. 13, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/8940457

[6] M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai, and R. Mishra, "A Proposed Model for Lifestyle Disease Prediction Using Support Vector Machine," 2018 9th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2018, Oct. 2018, doi: 10.1109/ICCCNT.2018.8493897.

[7] M. Gulhane and T. Sajana, "A Machine Learning based Model for Disease Prediction," 2021 International Conference on Computing, Communication and Green Engineering, CCGE 2021, 2021, doi: 10.1109/CCGE50943.2021.9776374.

[8] A. Parab, P. Gholap, and V. Patankar, "DiseaseLens: A Lifestyle related Disease Predictor," 5th IEEE International Conference on Advances in Science and Technology, ICAST 2022, pp. 383–387, 2022, doi: 10.1109/ICAST55766.2022.10039533.

[9] A. H. Neehal, M. N. Azam, M. S. Islam, M. I. Hossain, and M. Z. Parvez, "Prediction of Parkinson's Disease by Analyzing fMRI Data and using Supervised Learning," 2020 IEEE Region 10 Symposium, TENSYMP 2020, pp. 362–365, Jun. 2020, doi: 10.1109/TENSYMP50017.2020.9230918.

[10] M. Patil, H. Bhadane, D. Shewale, M. Lahoti, D. Singh Rajawat, and R. Kumar Solanki, "Smart Machine Learning Model Early Prediction of Lifestyle Diseases," International Journal of Aquatic Science, vol. 14, p. 2023.

APPENDIX-A

PSUEDOCODE

```
#Import necessary libraries import pickle
import streamlit as st
from streamlit_option_menu import option_menu

# Load saved models
diabetes_model = pickle.load(open('path/to/diabetes_model.sav', 'rb')) heart_disease_model =
pickle.load(open('path/to/heart_disease_model.sav', 'rb')) parkinsons_model =
pickle.load(open('path/to/parkinsons_model.sav', 'rb'))

# Sidebar for navigation with st.sidebar:
selected = option_menu('Multiple Disease Prediction System',
['Diabetes Prediction', 'Heart Disease Prediction', 'Parkinsons Prediction'], icons=['activity', 'heart',
'person'],
default_index=0)

# Main application logic
if selected == 'Diabetes Prediction': # Page title
st.title('Diabetes Prediction using ML')


# Getting input data from the user # ... (similar code as in your script)

# Code for prediction diab_diagnosis = "

# Button for prediction
```

```
if st.button('Diabetes Test Result'): diab_prediction = diabetes_model.predict( )

if diab_prediction[0] == 1:
diab_diagnosis = 'The person is diabetic' else:
diab_diagnosis = 'The person is not diabetic' st.success(diab_diagnosis)
elif selected == 'Heart Disease Prediction': # Page title
st.title('Heart Disease Prediction using ML')

# Getting input data from the user # ... (similar code as in your script)

# Code for prediction heart_diagnosis = ''

# Button for prediction
if st.button('Heart Disease Test Result'): heart_prediction = heart_disease_model.predict( )

if heart_prediction[0] == 1:
heart_diagnosis = 'The person is having heart disease' else:
heart_diagnosis = 'The person does not have any heart disease' st.success(heart_diagnosis)

elif selected == "Parkinsons Prediction": # Page title
st.title("Parkinson's Disease Prediction using ML")

# Getting input data from the user # ... (similar code as in your script)

# Code for prediction parkinsons_diagnosis = ''

# Button for prediction
if st.button("Parkinson's Test Result"): parkinsons_prediction = parkinsons_model.predict( )

if parkinsons_prediction[0] == 1:
```

```
parkinsons_diagnosis = "The person has Parkinson's disease" else:
parkinsons_diagnosis = "The person does not have Parkinson's disease"
st.success(parkinsons_diagnosis)
# Import necessary libraries
import pickle
import streamlit as st
from streamlit_option_menu import option_menu
# Function to get user inputs for diabetes prediction
def get_diabetes_input():
  st.subheader('Diabetes Prediction Input')
  age = st.slider("Age", 20, 100, 40)
  glucose_level = st.number_input("Glucose Level (mg/dL)")
  # Add more input fields for other features
  return age, glucose_level  # Return user inputs
# Function to display and handle diabetes prediction
def predict_diabetes(diabetes_model):
  st.subheader('Diabetes Prediction using ML')
  age, glucose_level = get_diabetes_input()  # Get user inputs
  diab_diagnosis = ''
  if st.button('Diabetes Test Result'):
    diab_prediction = diabetes_model.predict([[age, glucose_level]])  # Include other features

    if diab_prediction[0] == 1:
      diab_diagnosis = 'The person is diabetic.'
    else:
      diab_diagnosis = 'The person is not diabetic.
    st.success(diab_diagnosis)
# Function to get user inputs for heart disease prediction
# Create similar functions for other diseases
# Load saved models
diabetes_model = pickle.load(open('path/to/diabetes_model.sav', 'rb'))
```

```
heart_disease_model = pickle.load(open('path/to/heart_disease_model.sav', 'rb'))
parkinsons_model = pickle.load(open('path/to/parkinsons_model.sav', 'rb'))
# Sidebar for navigation with st.sidebar
selected = option_menu('Multiple Disease Prediction System',
            ['Diabetes Prediction', 'Heart Disease Prediction', 'Parkinsons Prediction'],
            icons=['activity', 'heart', 'person'],
            default_index=0)


# Main application logic
if selected == 'Diabetes Prediction':
  predict_diabetes(diabetes_model)
elif selected == 'Heart Disease Prediction':
  # Function call for heart disease prediction
  elif selected == "Parkinsons Prediction":
  # Function call for Parkinson's disease prediction
```

PSUEDOCODE FOR DIABETES DATA SET

```
# Import necessary libraries
import numpy as np
import pandas as pd from sklearn.model_selection
import train_test_split from sklearn
import svm from sklearn.metrics
import accuracy_score
import pickle
# Data Collection and Analysis
# Load the dataset diabetes_dataset = pd.read_csv('/content/diabetes.csv')
# Explore the dataset
# Display the first 5 rows of the dataset
diabetes_dataset.head()
# Get the shape of the dataset (number of rows and columns)
diabetes_dataset.shape
```

```
# Get statistical measures of the data
diabetes_dataset.describe()
# Count the distribution of outcomes
diabetes_dataset['Outcome'].value_counts()
# Group data by outcome to observe means of features for each outcome
diabetes_dataset.groupby('Outcome').mean()
# Separating the data into features (X) and labels (Y)
X = diabetes_dataset.drop(columns='Outcome', axis=1)
Y = diabetes_dataset['Outcome']
# Train Test Split X_train, X_test, Y_train,
Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)
# Training the Model
classifier = svm.SVC(kernel='linear') classifier.fit(X_train, Y_train)
# Model Evaluation
# Calculate accuracy score on training and test data
X_train_prediction = classifier.predict(X_train) training_data_accuracy =
accuracy_score(X_train_prediction, Y_train)
X_test_prediction = classifier.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
# Making Predictions
input_data = (5, 166, 72, 19, 175, 25.8, 0.587, 51)
input_data_as_numpy_array = np.asarray(input_data)
input_data_reshaped = input_data_as_numpy_array.reshape(1, -1)
prediction = classifier.predict(input_data_reshaped)
if (prediction[0] == 0):
print('The person is not diabetic')
else:  print('The person is diabetic')
# Saving the trained model filename = 'diabetes_model.sav' pickle.dump(classifier, open(filename,
'wb'))
# Loading the saved model and making predictions
```

loaded_model = pickle.load(open('diabetes_model.sav', 'rb')) prediction =
loaded_model.predict(input_data_reshaped) if (prediction[0] == 0):
print('The person is not diabetic') else: print('The person is diabetic')
# Display the columns of feature data (X) for column in X.columns: print(column)

APPENDIX-B

SCREENSHOTS

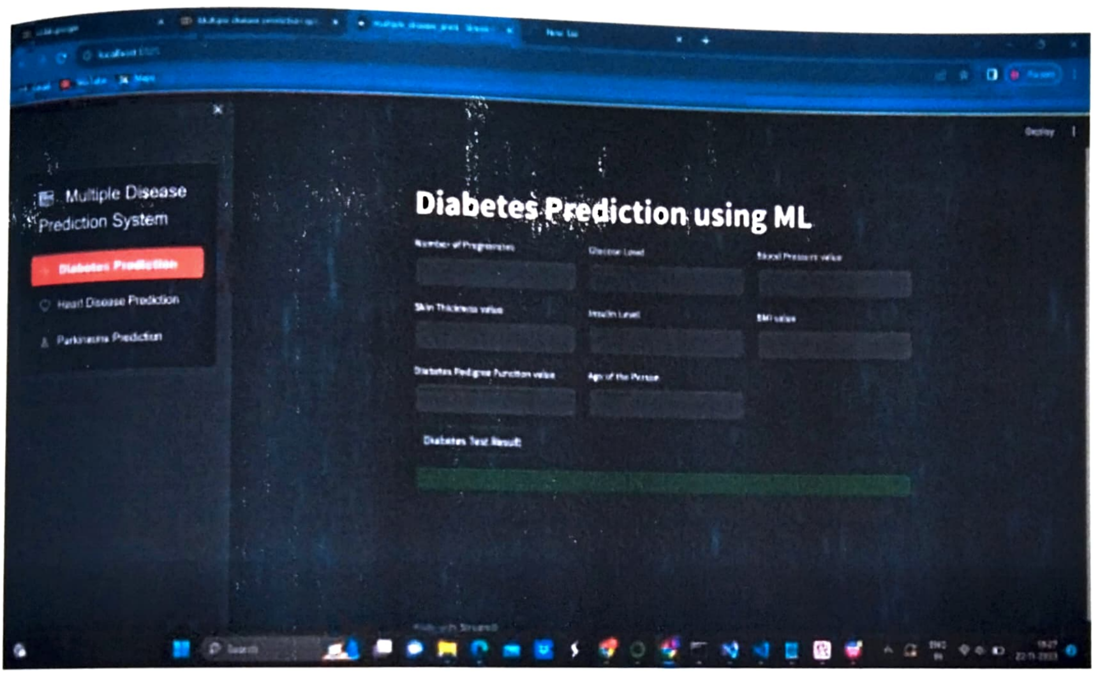

Figure 8: Interface of the Disease Prediction System

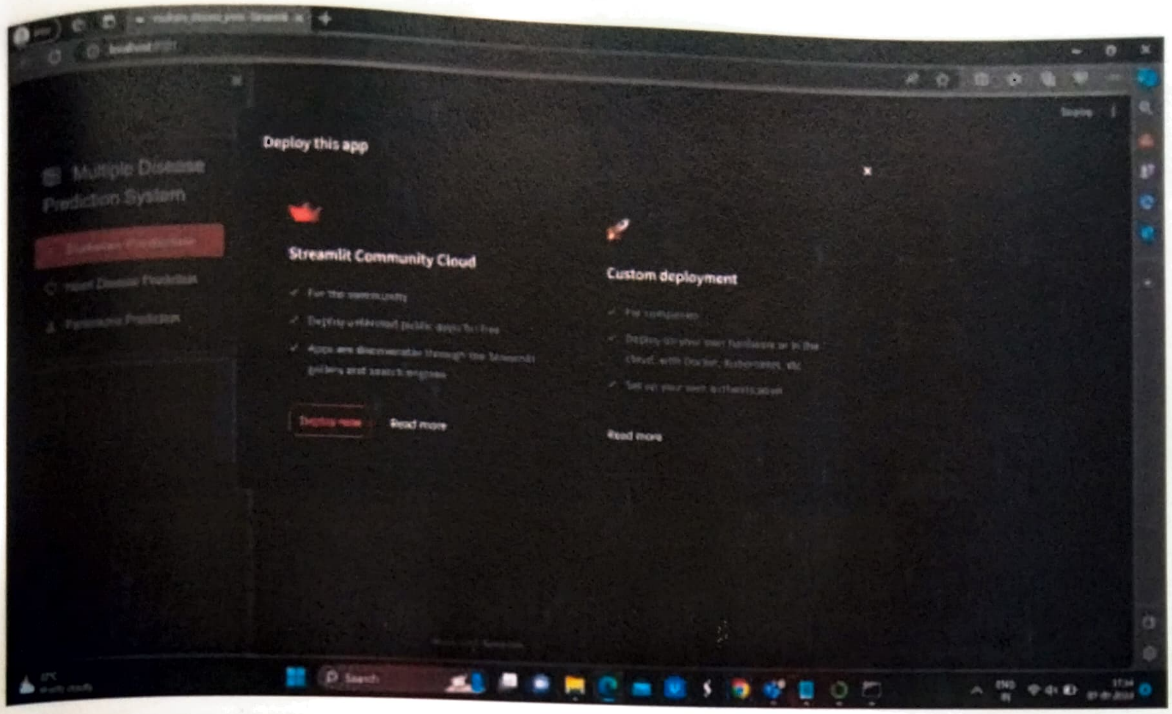

Figure 9: Different Deployment Options

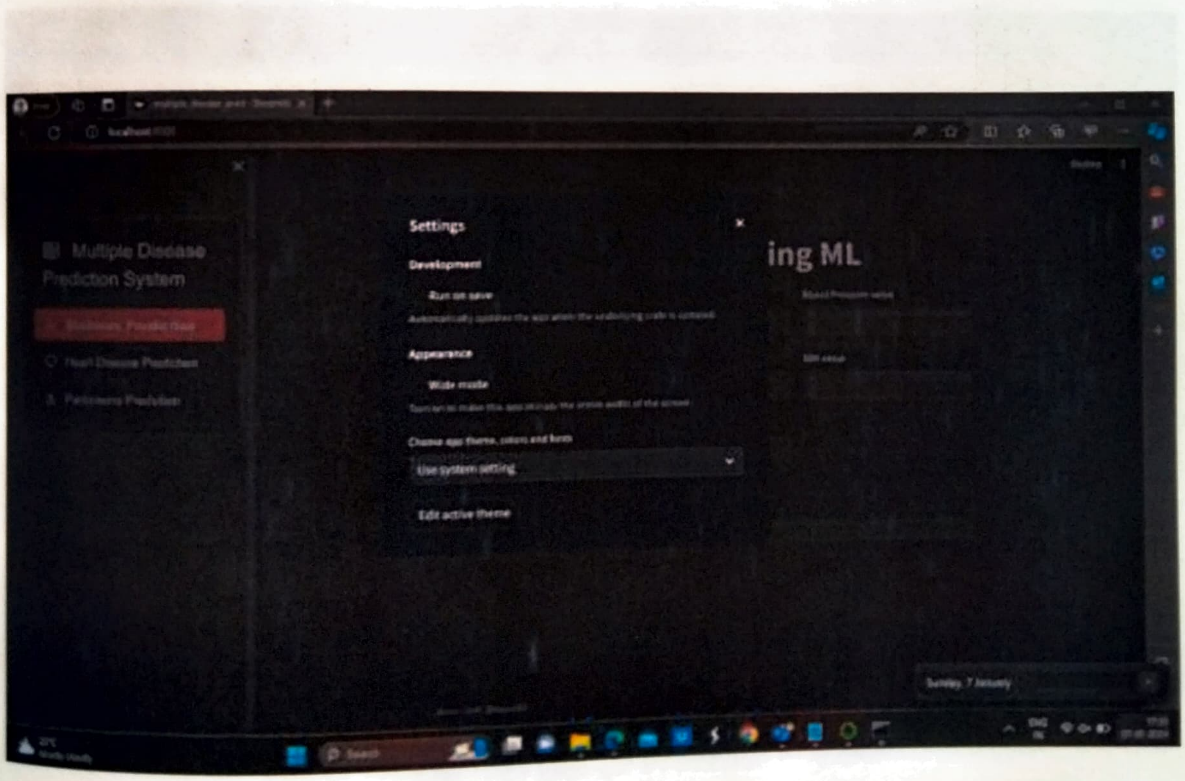

Figure 10: Settings for Modifications

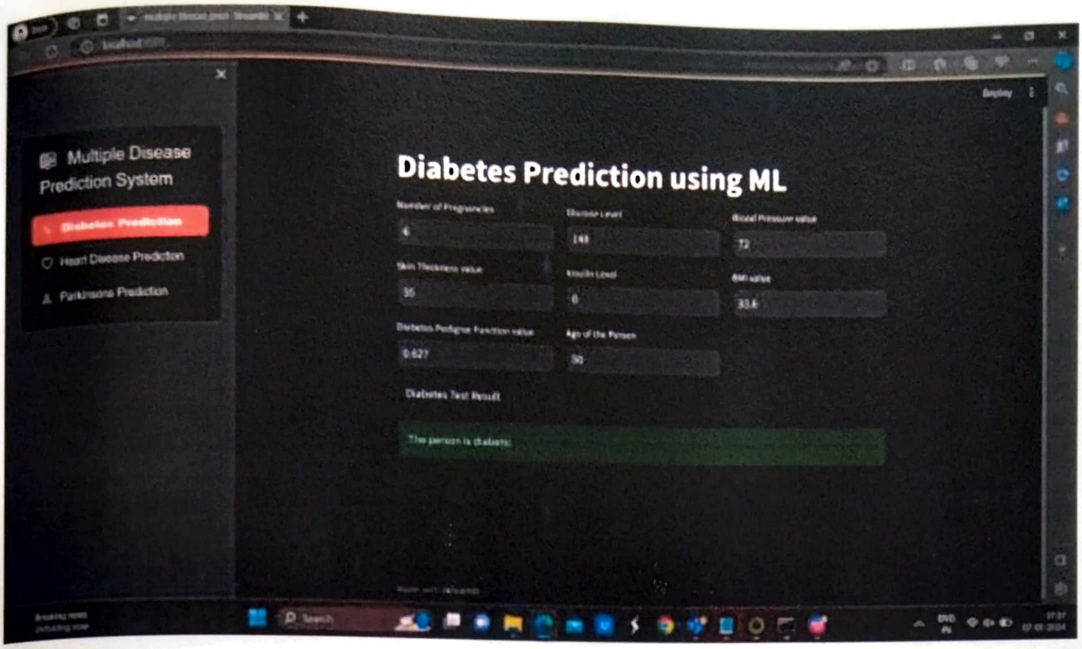

Figure 11: Diabetes Disease Prediction positive

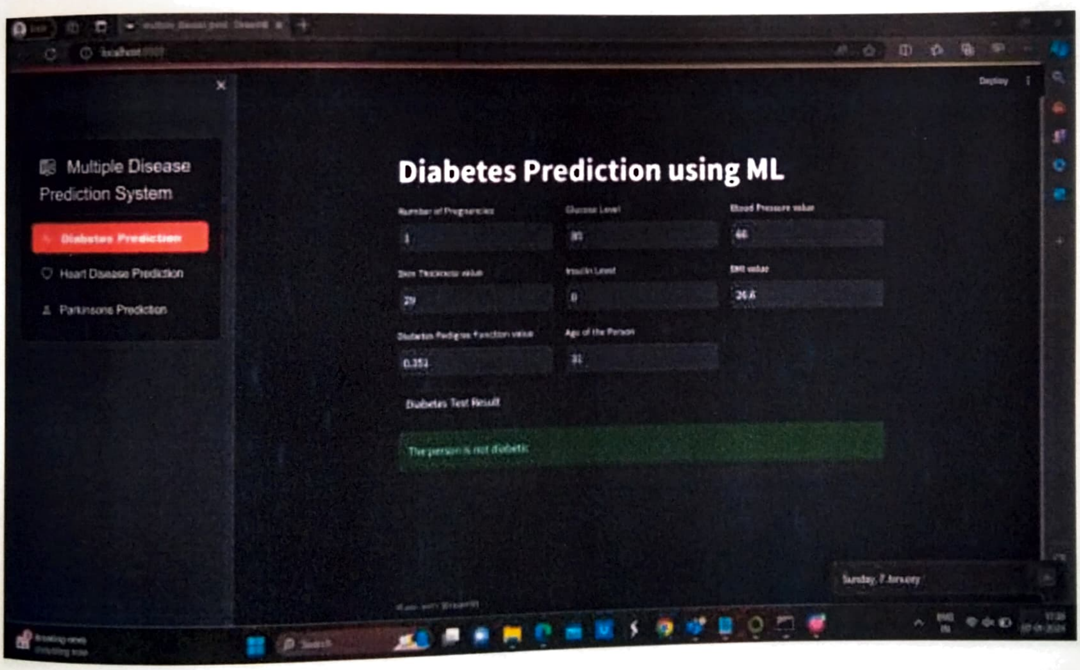

Figure 12: Diabetes Disease Prediction negative

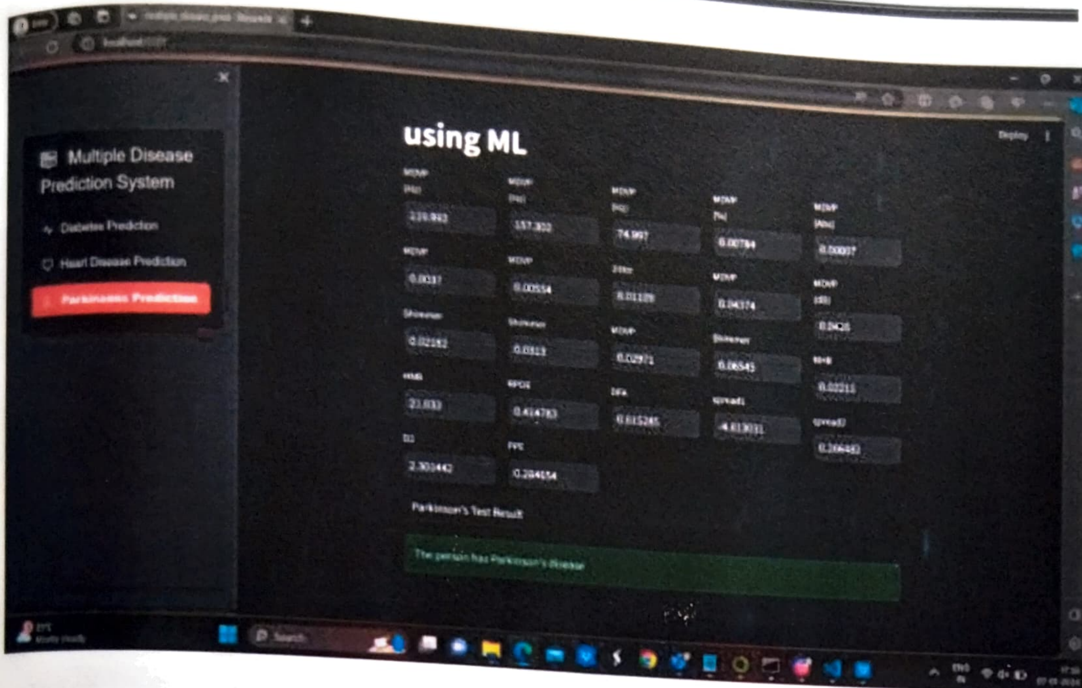

Figure 13: Parkinson's Disease Prediction positive

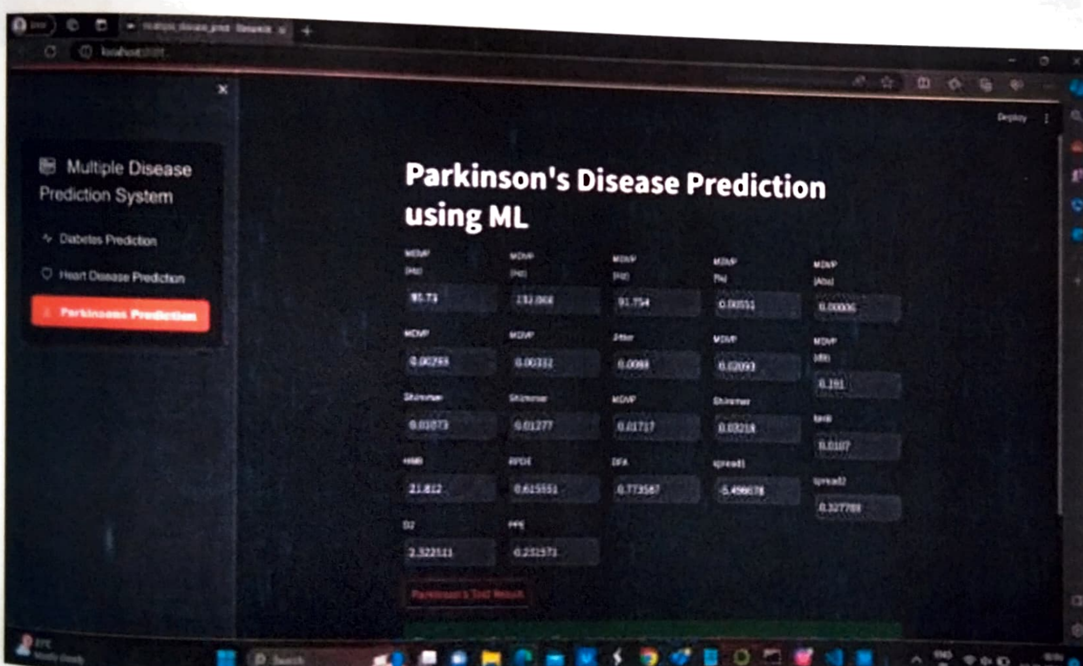

Figure 14: Parkinson's Disease Prediction negative

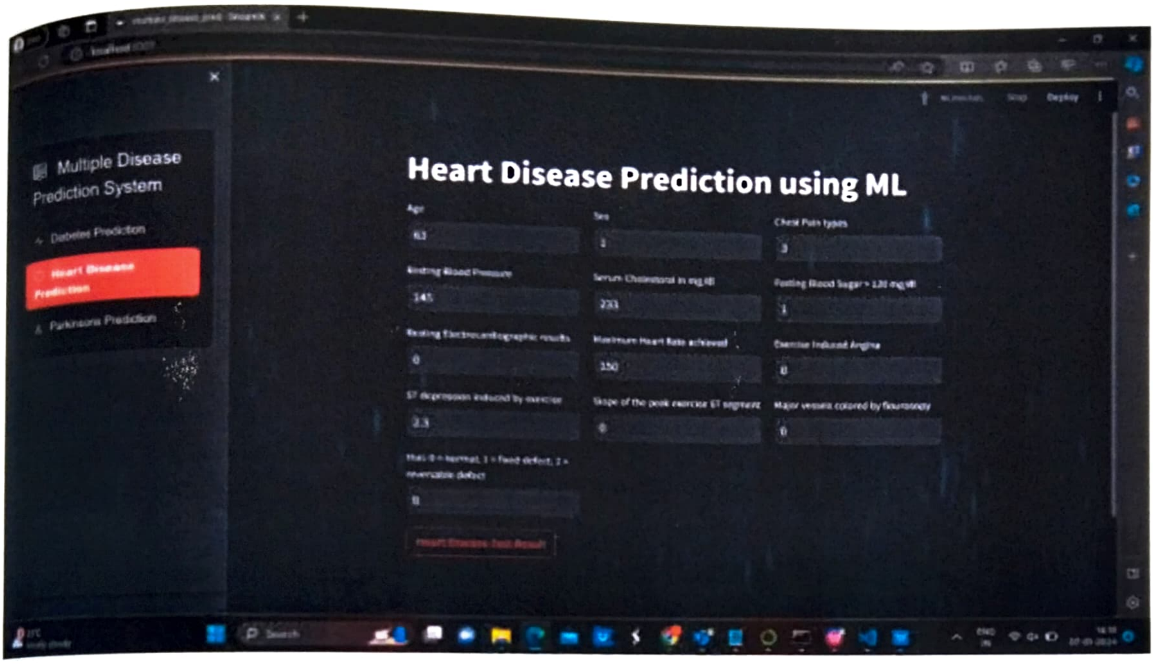

Figure 15: Heart Disease Prediction

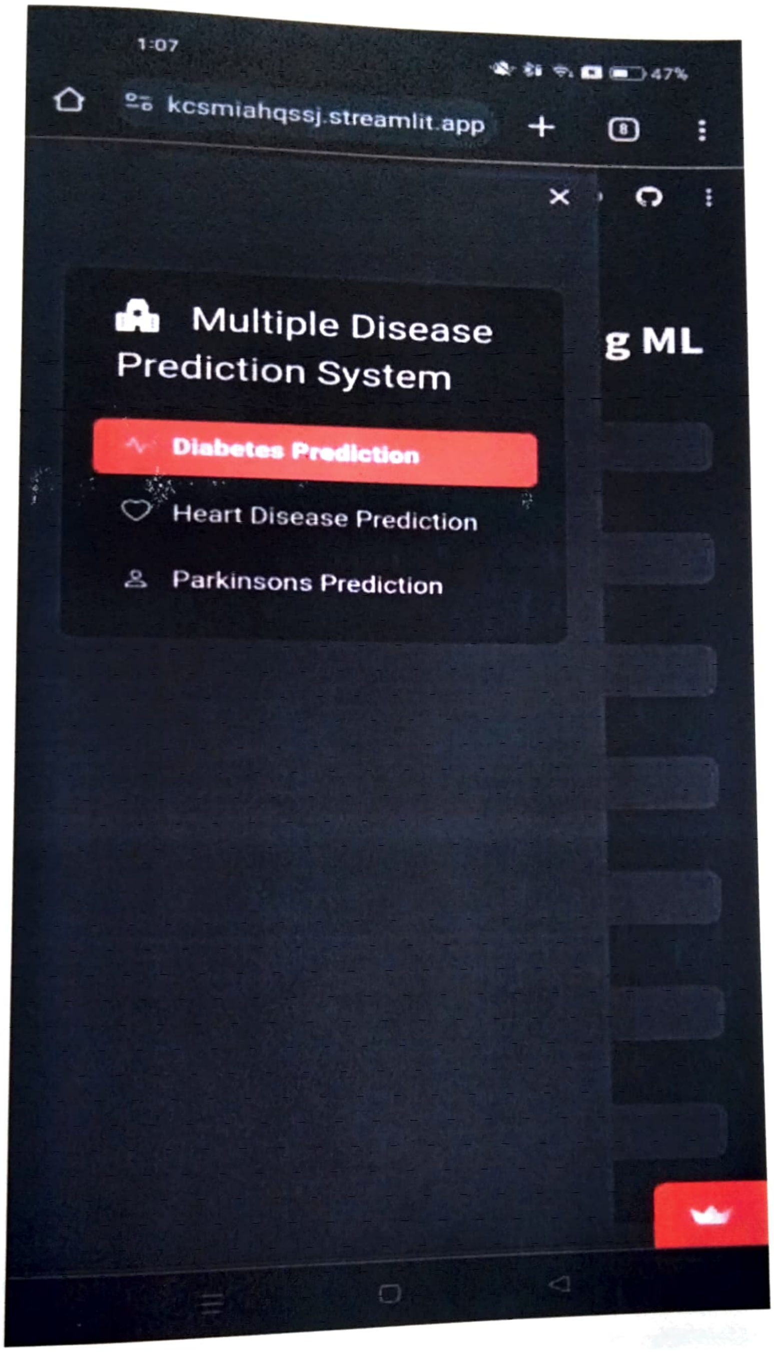

Figure 16: Mobile view Web App

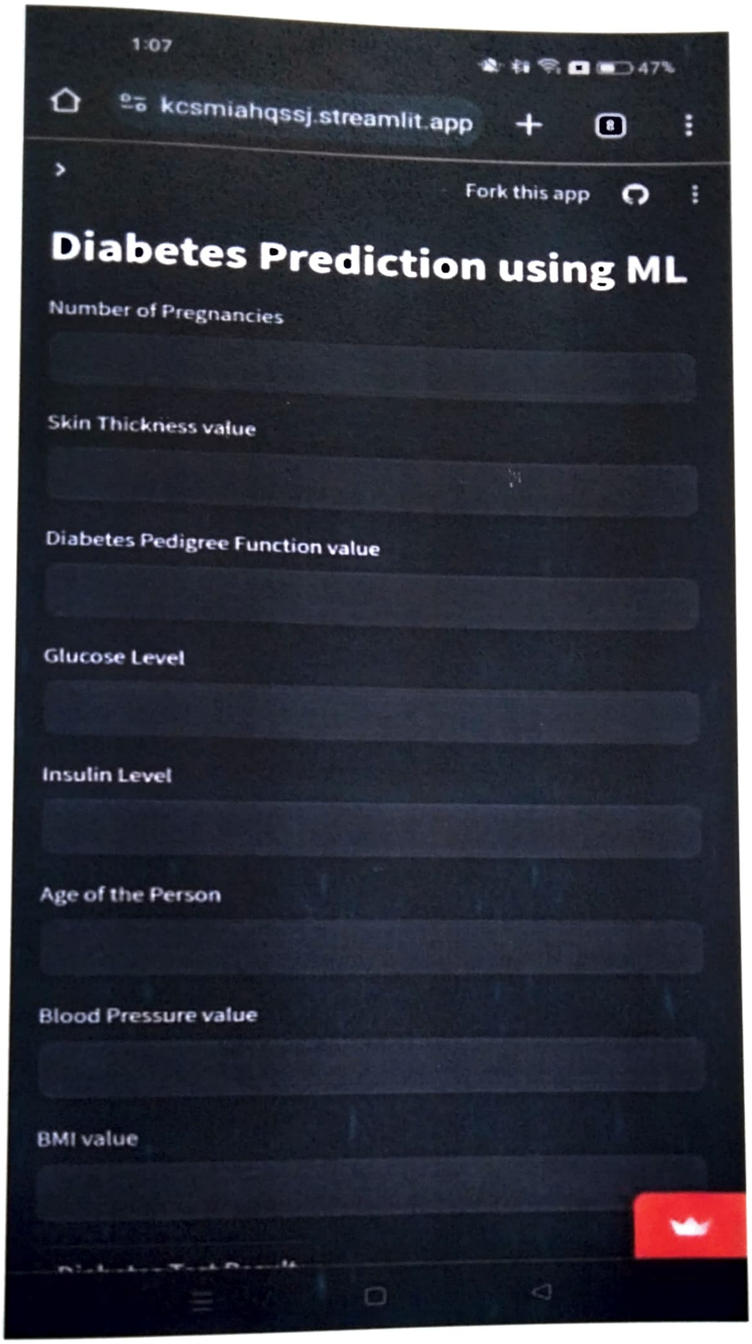

Figure 17: Mobile view for Diabetes Prediction input

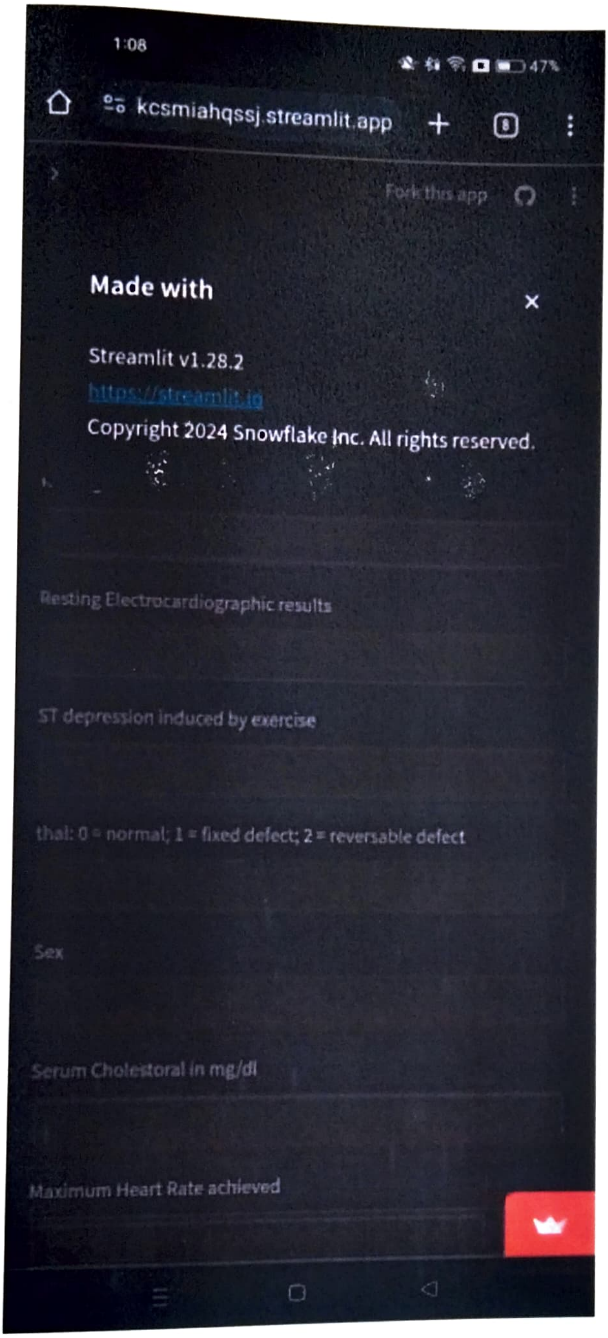

Figure 18: Mobile view settings

APPENDIX-C

ENCLOSURES

Sukruthi - UP-II Report

ORIGINALITY REPORT

20%	17%	11%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to M S Ramaiah University of Applied Sciences Student Paper	9%
2	Submitted to University of Wales Institute, Cardiff Student Paper	2%
3	www.coursehero.com Internet Source	1%
4	github.com Internet Source	1%
5	fastercapital.com Internet Source	1%
6	huggingface.co Internet Source	<1%
7	www.sei.cmu.edu Internet Source	<1%
8	Mowafaq Salem Alzboon, Mohammad Subhi Al-Batah, Muhyeeddin Alqaraleh, Ahmad Abuashour, Ahmad Fuad Hamadah Bader.	<1%

Plagiarism report for the complete project report

Link for Research paper:

https://drive.google.com/file/d/1vw0Cw4sHiLX92bzrSolWHL7rDpffm0Da/view?usp=drivesdk

Journal of Xidian University

An UGC-CARE Approved Group - 2 Journal (Scopus Active Journal)

ISSN NO: 1001-2400

Scientific Journal Impact Factor - 5.4

ACCEPTANCE LETTER TO AUTHOR

Dear Author,

With reference to your paper submitted "Leveraging Machine Learning for Early Prediction of Lifestyle Diseases: A Data- Driven Approach", we are pleased to accept the same for publication in Journal of Xadian University.

Manuscript ID: JXU-R10025

Please send the payment receipt for an online maintenance / processing fee of 2000+18%=2360 INR per paper. Please note that the amount we are charging is very nominal & only an online maintenance and processing fee.

The Fee includes:

- Online maintenance and processing charge.
- Soft copy of certificate for each author.
- No limitation of number of pages.
- Editoral fee.

Note:

- Paper will be published within 48 hours after receiving the payment confirmation.
- Once the paper is published online, corrections are not allowed. So send your final paper before publication.
- Fee paid for publication of the paper will not be refunded under any circumstances.
- In case of any query please do not hesitate to contact us at editorjxu@gmail.com. Early reply is appreciated.

DATE
8-Jan-24

Sincerely,
Best regards,
Jenny Corbett

Jenny

http://www.xadzkjdx.cn/

Letter of Acceptance

UGC CARE GROUP - 2 JOURNAL - Journal of Xidian University: Volume 18, Issue 1, 2024

Manuscript ID(s): JXU-R10025

Review Observations

Journal of XIDIAN University

ISSN: 1001-2400

Paper Title: Leveraging Machine Learning for Early Prediction of Lifestyle Diseases: A Data- Driven Approach

Author(s): Sukruthi Rao, S Harish, R Saisaran

Dear Author(s),

Please find the paper review reflections for your paper below:

Clear Problem Statement: The paper effectively addresses the pressing issue of lifestyle diseases and underscores the importance of early prediction. The clarity in defining the problem creates a strong foundation for the subsequent discussion on predictive models and their application in healthcare.

Methodological Transparency: The authors demonstrate transparency in their methodology, particularly in the development of predictive models using scikit-learn. The integration of logistic regression, streamlined workflow within Anaconda, and data integrity measures with MD5 are well-explained. This clarity aids readers in understanding the technical aspects of the proposed system.

Comprehensive Literature Review: The literature survey provides a comprehensive overview of the landscape, citing relevant studies and technologies in disease prediction. The inclusion of machine learning, genomics, and wearable technologies reflects a nuanced understanding of the diverse approaches to early disease identification.

Structured Work Plan and Key Concepts: The work plan is well-structured, outlining steps from model loading to user engagement. The key concepts section effectively summarizes the core components of the multiple disease prediction system, offering readers a quick grasp of the implemented approach.

Clarification on Model Loading: The paper briefly mentions concerns about script loading saved models without corresponding training procedures. It would be beneficial to elaborate on this point, providing a clearer explanation and perhaps suggesting solutions. This ensures that readers, especially those less familiar with the technical aspects, understand the potential issues and their implications on the application's functionality.

Thanks,
Editor

Review Observations

SUKRUTHI - Leveraging Machine Learning for Early Prediction of Lifestyle Diseases: A Data-Driven Approach

ORIGINALITY REPORT

6%	4%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.biorxiv.org Internet Source	1%
2	Submitted to University of Huddersfield Student Paper	1%
3	Submitted to Napier University Student Paper	1%
4	cambridgespark.com Internet Source	1%
5	Submitted to Arts, Sciences & Technology University In Lebanon Student Paper	<1%
6	Submitted to University of Greenwich Student Paper	<1%
7	eprints.ucm.es Internet Source	<1%
8	Submitted to Liverpool John Moores University Student Paper	<1%

Plagiarism for the Research Paper

SUSTAINABLE DEVELOPMENT GOALS

The Project work carried out here is mapped to SDG-3 Good Health and Well-Being.

The project work carried here contribute to the well-being of the human society. This can be used for analyzing and detecting Diabetes, Heart Disease, Parkinson's disease in the early stages so that the required medication can be started early to avoid further consequences which might result in mortality.