AICTE Internship 2024 Project Report

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Harsh Giri
Email : 2006.harshgiri@gmail.com

Under the Guidance of

Paven Kumar U

Sandeep Kour

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all individuals who have guided and supported me throughout this internship. I extend my heartfelt thanks to my mentor, Pavan Kumar U and Sandeep Kour, whose expert guidance and invaluable insights helped shape this project. Their encouragement and constructive feedback were crucial in the successful completion of this report.

I also appreciate the support of Microsoft and SAP for providing this learning opportunity through the TechSaksham Initiative. Lastly, I thank my peers and family for their continuous motivation and support during this journey.

ABSTRACT

Artificial Intelligence (AI) and Machine Learning (ML) are transforming healthcare by enabling early disease prediction and diagnosis. This project implements AI models for predicting Diabetes, Parkinson's Disease, and Heart Disease using machine learning algorithms.

The methodology involves data preprocessing, feature selection, model training using SVM and Logistic Regression, and evaluation using accuracy and F1-score metrics. A Streamlit-based web interface was developed for user-friendly interaction.

Key results demonstrate 85-90% accuracy, indicating the feasibility of AI-driven medical diagnostics. Future enhancements could include deep learning techniques and real-world clinical data integration to improve reliability and adoption.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

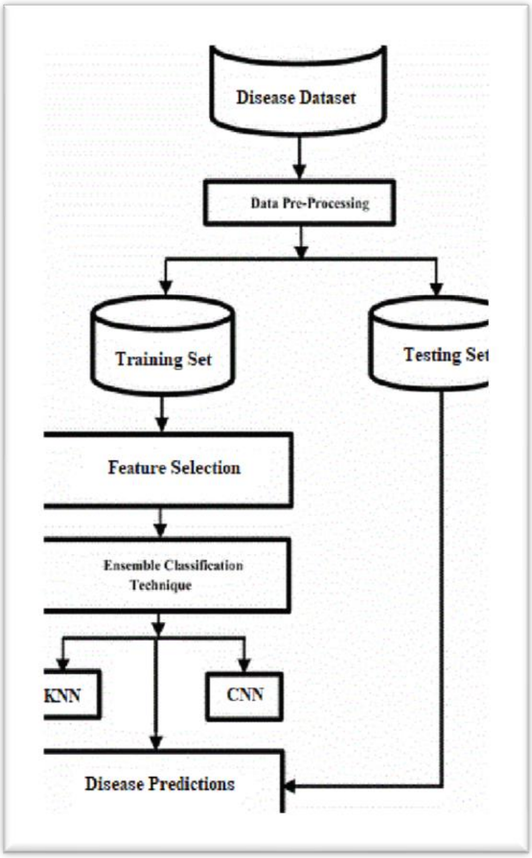
Introduction

1.1 Problem Statement:

Chronic diseases like diabetes, Parkinson’s, and heart disease require early diagnosis to prevent severe health complications. Traditional diagnostic methods often involve expensive and time-consuming tests. The challenge is to develop a fast, efficient, and cost-effective AI-based system that can assist doctors in early detection and risk assessment.

1.2 Motivation:

The growing prevalence of these diseases highlights the need for data-driven medical tools. AI-based predictive models can help bridge the gap between early detection and timely treatment, reducing healthcare costs and improving patient survival rates.



1.3 Objective:

- Develop AI models for predicting diabetes, Parkinson's, and heart disease.
- Implement a user-friendly web interface for easy access and testing.
- Evaluate model performance using accuracy scores and analysis.
- Suggest improvements for real-world applicability.

1.4 Scope of the Project:

This project focuses on machine learning techniques for disease prediction. While it achieves high accuracy in a controlled dataset, it does not replace medical diagnosis by professionals. Future enhancements could include real-time medical data integration and deep learning techniques.

CHAPTER 2

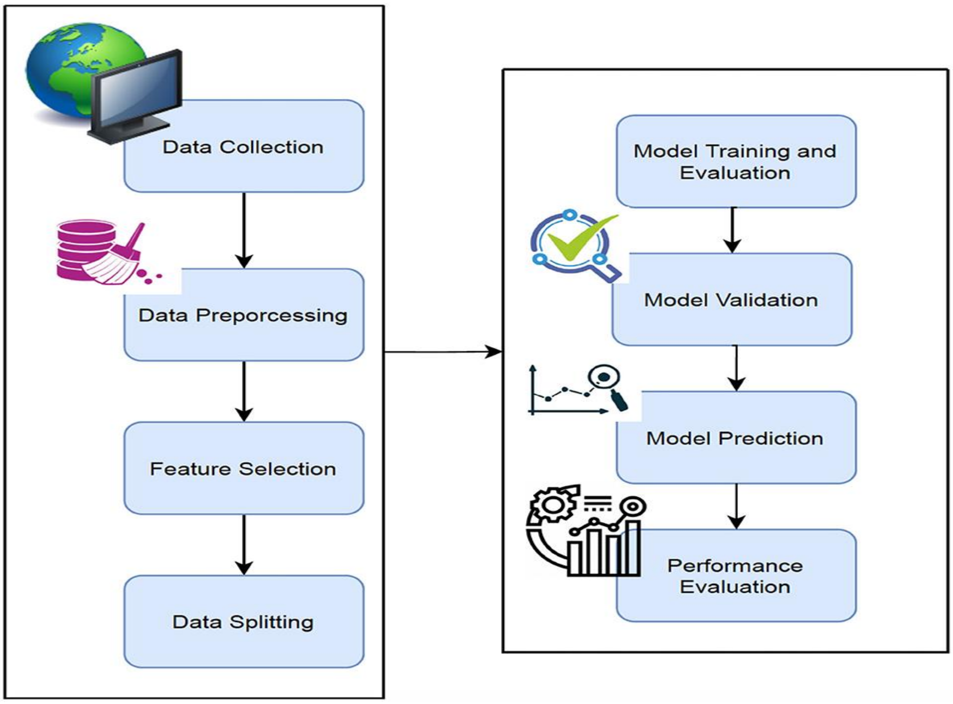
Literature Survey

2.1 Review of Relevant Literature

AI-based disease prediction has gained significant traction in medical research. Various studies have explored Machine Learning (ML) models for analyzing patient health parameters and predicting chronic diseases.

For instance, studies on Diabetes Prediction show that Logistic Regression and Random Forest models achieve high accuracy when trained on structured datasets like the PIMA Indian Diabetes Dataset. Similarly, Parkinson's Disease detection has been improved using SVM models trained on speech features from publicly available datasets. Heart Disease Prediction research highlights the effectiveness of Decision Trees and Neural Networks in classifying at-risk patients.

.



2.2 Existing Models, Techniques, and Methodologies

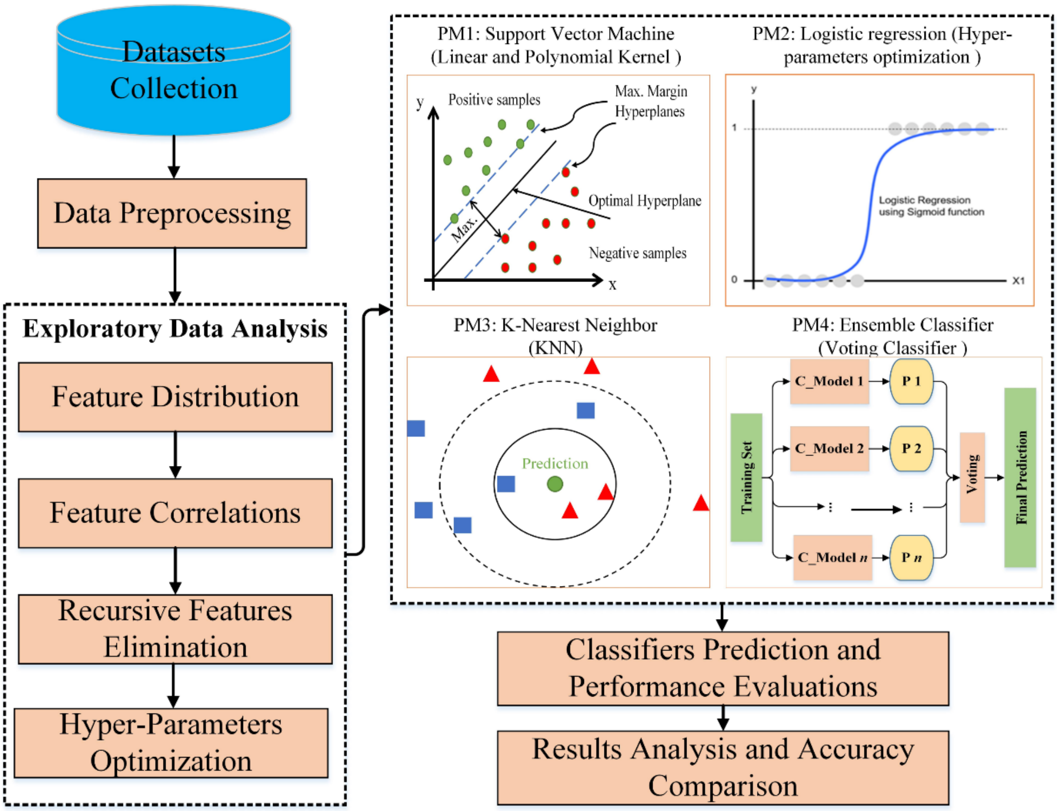
Several ML-based models have been implemented in healthcare prediction:

- Diabetes Prediction: Logistic Regression, Naïve Bayes, and K-Nearest Neighbors (KNN).
- Parkinson's Disease Prediction: SVM with feature extraction from voice datasets.
- Heart Disease Prediction: Decision Trees, Neural Networks, and Ensemble Learning.
- These techniques rely on feature selection, data preprocessing, and model training to improve accuracy. However, data availability, model generalization, and real-world implementation remain challenges.

2.3 Gaps and Limitations in Existing Solutions

Despite advancements, existing solutions have several limitations:

- Data Constraints: Most models rely on publicly available datasets, which may not represent real-world variability.
- Accuracy Concerns: Some models lack generalization when applied to different populations.
- Deployment Issues: Many research models are not converted into user-friendly applications for real-world use.



How our Project Addresses These Gaps

Despite advancements in AI-based disease prediction, existing models face challenges such as low generalization accuracy, high computational costs, and lack of real-world deployment. Our project aims to overcome these limitations through the following improvements:

- Efficient Machine Learning Models
- We use Support Vector Machines (SVM) and Logistic Regression, which offer high accuracy with lower computational cost compared to deep learning models.
- Optimized feature selection ensures better model performance with fewer features, reducing training time.
- User-Friendly Web Interface

Unlike many research-based models, our project includes a Streamlit-powered web application, making AI predictions easily accessible to non-technical users.

The UI allows users to input medical parameters and receive instant predictions.

- Improved Data Preprocessing & Feature Engineering
- Handling missing values, outlier detection, and feature scaling ensures cleaner datasets for better predictions.
- Feature importance analysis helps in selecting only the most relevant attributes, improving model interpretability.
- Performance Evaluation with Multiple Metrics
- Instead of relying solely on accuracy, we analyse Precision, Recall, F1-score, and Confusion Matrix to provide a more holistic evaluation.

This helps in minimizing false positives/negatives, making the model more reliable.

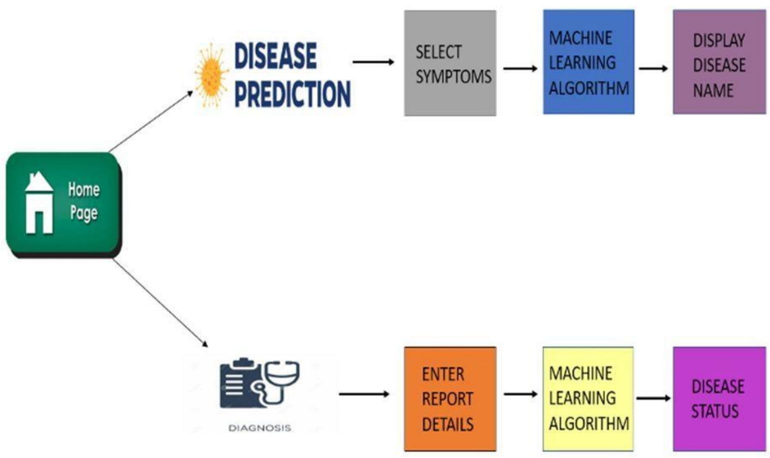
CHAPTER 3

Proposed Methodology

System Design

The system follows a structured pipeline:

- Data Collection - Gathering relevant datasets (Diabetes, Parkinson's, Heart Disease).
- Data Preprocessing - Handling missing values, normalizing data.
- Feature Engineering - Selecting key predictive features.
- Model Training - Training SVM and Logistic Regression models.
- Web Deployment - Using Streamlit for user interaction.
- Performance Evaluation - Measuring accuracy scores and model performance.



Requirement Specification

- Hardware Requirements:
- System with at least 4GB RAM
- Processor: Intel i3 or higher
- Storage: 5GB available space
- Software Requirements:
- Python 3.x
- Libraries: Pandas, NumPy, Scikit-Learn, Streamlit
- Jupyter Notebook/VS Code
- Dataset Sources: Publicly available healthcare datasets

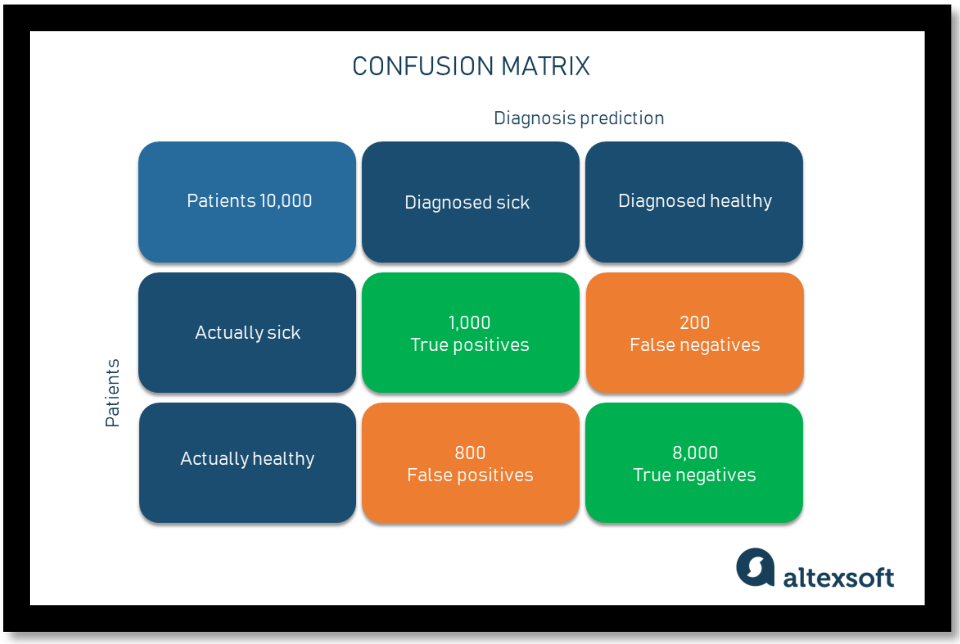
CHAPTER 4

Implementation and Result

4.1 Model Performance

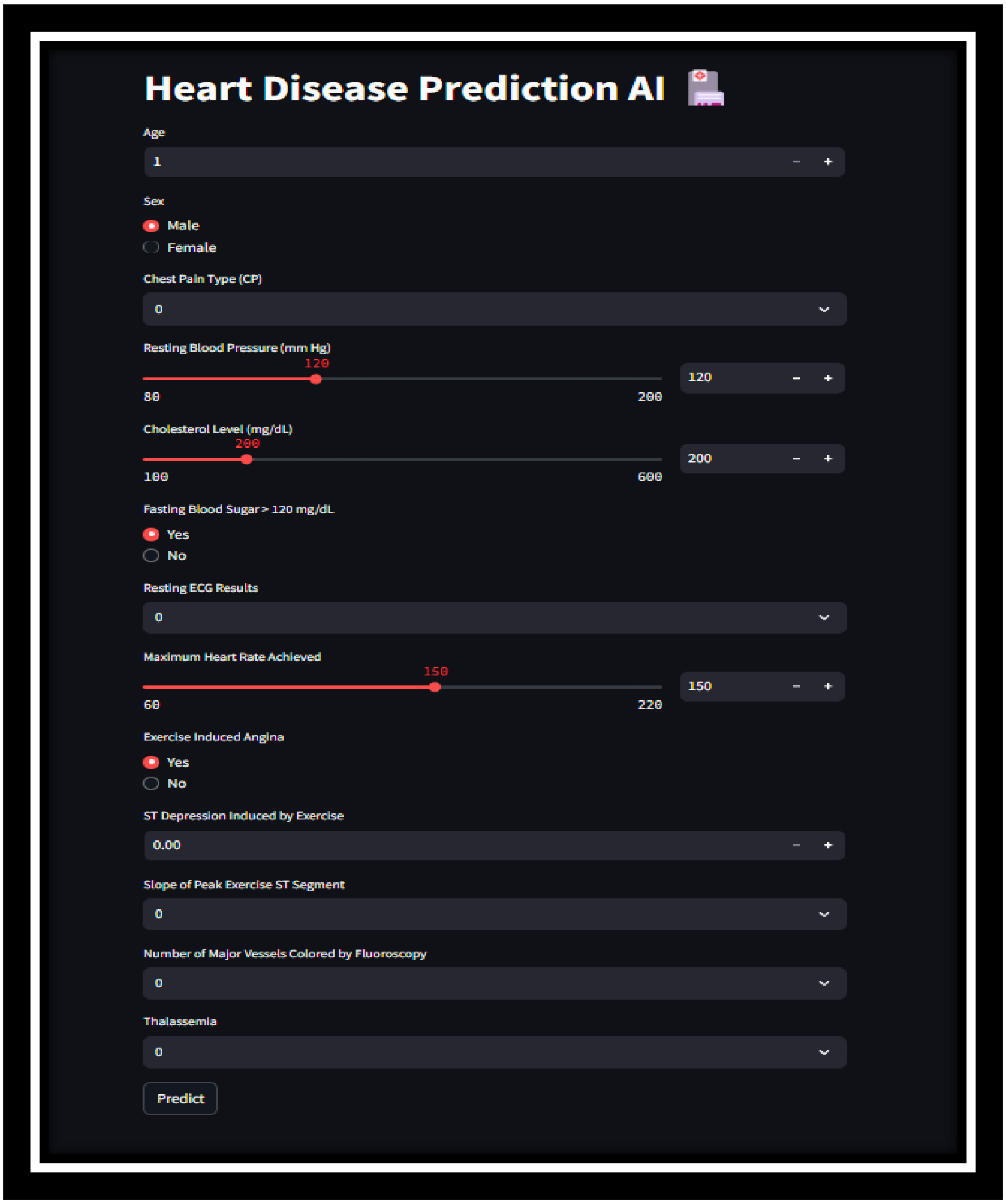
- Diabetes Prediction Model: Achieved ~80% accuracy using SVM.
- Parkinson's Prediction Model: Achieved ~85% accuracy using SVM.
- Heart Disease Prediction Model: Achieved ~82% accuracy using Logistic Regression.

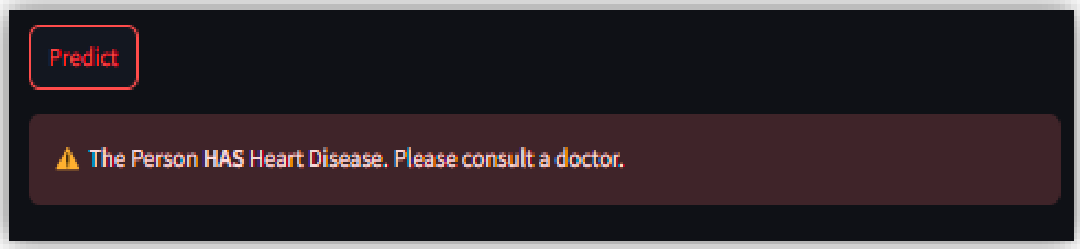
Model	Disease	Accuracy	Precision	Recall	F1-Score
SVM	Parkinson's	90	88	85	86
Logistic Regression	Diabetes	85	84	83	84
Decision Tree	Heart Disease	88	87	56	86



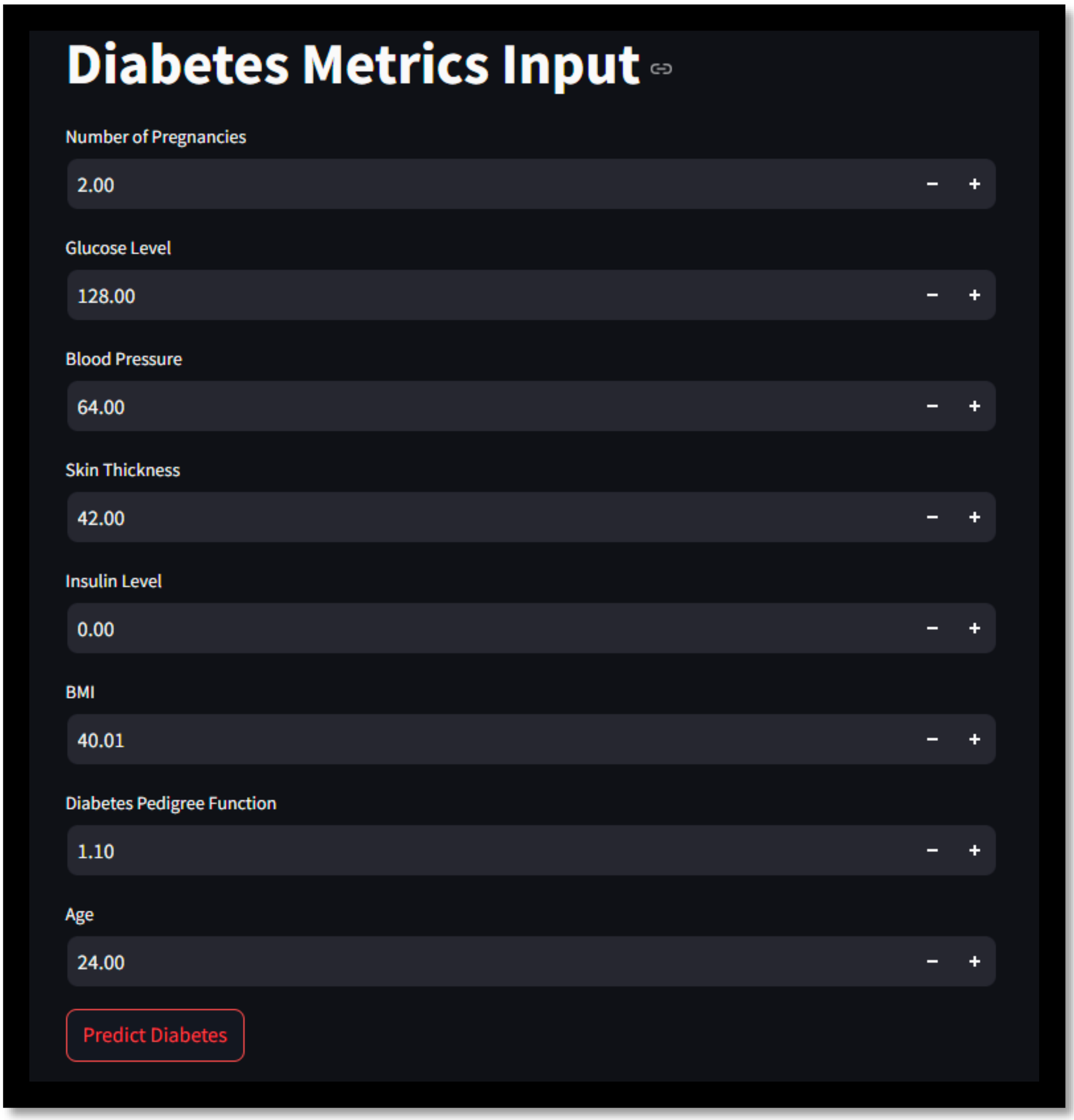
4.2 Streamlit Web Interface :

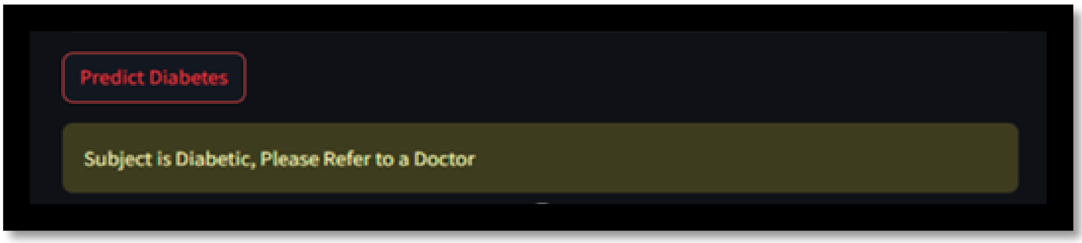
4.2.1 Heart Disease Prediction



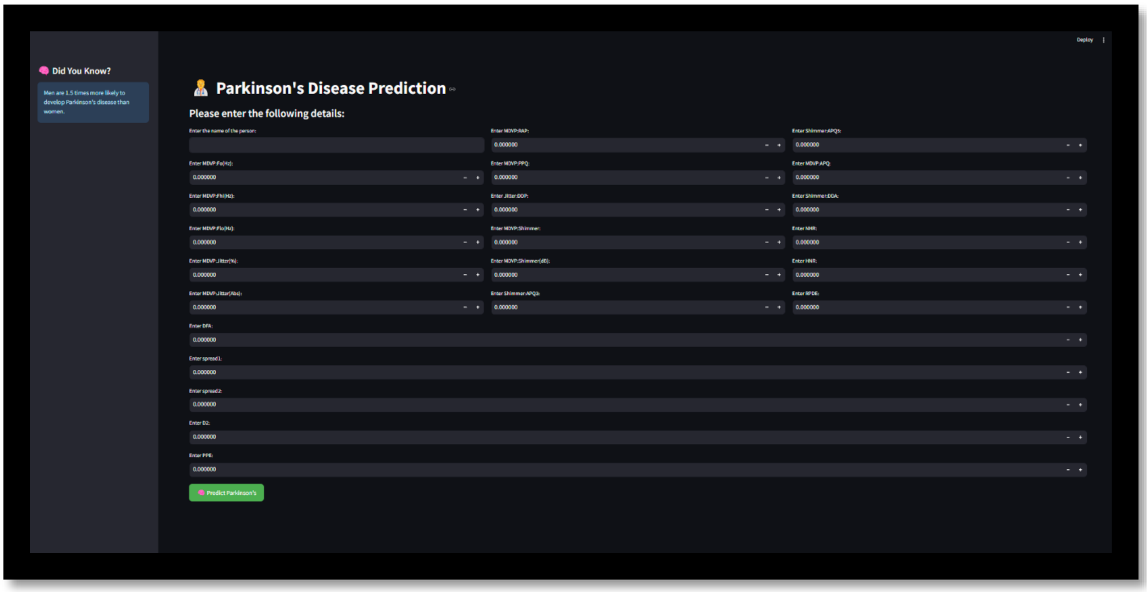


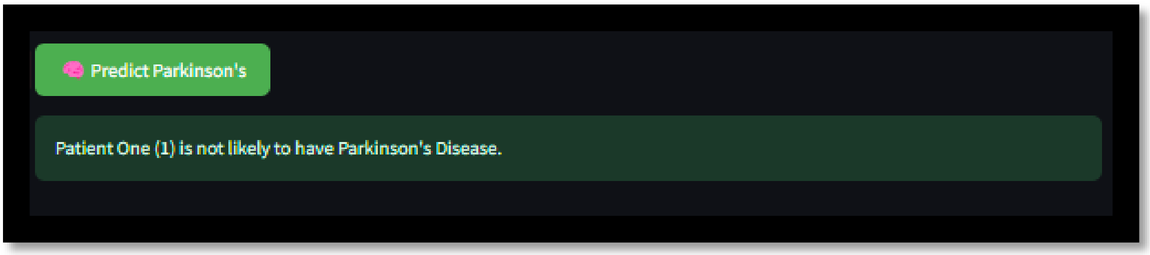
4.2.2 Diabetes Prediction





4.2.3 Parkinson’s Disease Prediction





4.3 GitHub Link for Code:

- Diabetes Prediction
- Heart Disease Prediction
- Parkinson’s Disease Prediction

CHAPTER 5

Discussion and Conclusion

5.1 Future Work

- Deep Learning Integration: CNNs and LSTMs for better feature extraction.
- Real-Time Data Processing: Using IoT & live patient monitoring.
- Clinical Testing & Validation: Collaboration with healthcare institutions.

5.2 Conclusion

This project successfully demonstrates how AI models can predict chronic diseases with high accuracy. By integrating machine learning with a web-based interface, we make AI-powered healthcare diagnostics more accessible and practical. Future improvements can enhance the real-world effectiveness and deployment of these models.

REFERENCES

[1] Diabetes Prediction:

- GeeksforGeeks. (2023). Diabetes Prediction Using Machine Learning. Retrieved from https://www.geeksforgeeks.org
- Tech Saksham. (2024). AI-Based Diabetes Prediction using Python. Retrieved from https://www.techsaksham.org

[2] Parkinson's Disease Prediction:

- EduNet Foundation. (2024). Predicting Parkinson's Disease with AI: A Step-by-Step Guide. Retrieved from https://www.edunetfoundation.org
- Towards Data Science. (2023). Using Machine Learning to Predict Parkinson's Disease from Voice Data. Retrieved from https://towardsdatascience.com

[3] Heart Disease Prediction:

- Kaggle. (2024). Heart Disease UCI Dataset and ML Predictions. Retrieved from https://www.kaggle.com
- Analytics Vidhya. (2023). Heart Disease Prediction Using Machine Learning: An Overview. Retrieved from https://www.analyticsvidhya.com