Implementation of AI-Powered Medical Diagnosis System

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

Gorav Jindal, goravjindal86@gmail.com

Under the Guidance of

Pavan Sumohana

ACKNOWLEDGEMENT

Looking back on this thesis journey, it's impossible not to feel an overwhelming sense of gratitude. So many people have contributed, whether through direct guidance, a kind word of encouragement, or simply by being a supportive presence. We truly couldn't have reached this point without each and every one of you.

First and foremost, my deepest thanks go to my supervisor, Pavan Sumohana. Honestly, 'supervisor' feels like too formal a title for the incredible mentor you've been. From the very beginning, you fostered an environment where I felt safe to explore, to question, and even to stumble. Your door was always open, and you consistently made time to discuss ideas, troubleshoot challenges, and offer invaluable insights. It wasn't just about the technical aspects of the research; you took a genuine interest in my growth as a student and as a professional.

What truly set you apart, Pavan, was your ability to inspire. Your passion for the subject matter was contagious, igniting a similar spark within me. You didn't just provide answers; you encouraged me to think critically, to challenge assumptions, and to develop my own unique perspective. Those brainstorming sessions, where we bounced ideas back and forth, were some of the most stimulating and productive moments of this entire process. You had a knack for seeing potential where I saw roadblocks, and for turning seemingly insurmountable obstacles into opportunities for learning.

Beyond the academic guidance, your unwavering belief in my abilities was a constant source of motivation. There were times when I doubted myself, when the weight of the project felt overwhelming, but your confidence in me never wavered. That unwavering support was the anchor I needed to stay focused and persevere.

And it wasn't just during scheduled meetings. You were always accessible, whether it was a quick email, a spontaneous chat in the hallway, or a thoughtful piece of feedback on a draft. You went above and beyond, not only in guiding my research, but also in helping me navigate the complexities of the program as a whole. Your advice extended beyond the thesis, shaping my understanding of professional responsibility and ethical conduct. You've taught me not just how to conduct research, but how to be a thoughtful and impactful member of the academic community.

Working with you for the past year has been an absolute privilege. I've learned more than I ever imagined, both about my field of study and about myself. You've not just been a supervisor; you've been a mentor, a guide, and a role model. Thank you, Pavan, for your unwavering support, your invaluable guidance, and for believing in me every step of the way.

ABSTRACT

This project presents a Streamlit-based web application designed to facilitate the prediction of several common diseases, including diabetes, heart disease, Parkinson's disease, lung cancer, and hypothyroidism. Utilizing pre-trained machine learning models, the application enables users to input relevant health parameters and receive instant predictions regarding their risk for these conditions. The interface is built with Streamlit, providing a user-friendly and interactive experience, enhanced by a visually engaging design with a customized background and color scheme. The core functionality relies on pickled machine learning models trained on established datasets, ensuring reliable and efficient predictions.

The application aims to bridge the gap between complex medical diagnostics and accessible, preliminary health assessments. By simplifying the input process and delivering clear, concise results, it empowers users to proactively monitor their health. Each disease prediction module requires specific input parameters, tailored to the respective model's requirements, ensuring accurate and relevant predictions. The application's design prioritizes ease of use, making it accessible to individuals with varying levels of technical expertise.

This project would not have been possible without the unwavering support and guidance of numerous individuals. Firstly, I extend my deepest gratitude to my supervisor, Pavan Sumohana, whose mentorship was instrumental in the project’s success. His insightful feedback, continuous encouragement, and ability to inspire critical thinking were invaluable. His dedication to fostering my growth as a student and professional has been truly remarkable. Furthermore, I acknowledge the contributions of all those who provided assistance and encouragement throughout this journey. This application serves as a testament to the collaborative effort and dedication required to develop impactful healthcare tools, providing a valuable resource for preliminary health assessments.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Problem Statement:

The increasing prevalence of chronic diseases like diabetes, heart disease, Parkinson's, lung cancer, and thyroid disorders presents a significant challenge to healthcare systems globally. Early detection and timely intervention are crucial for effective management and improved patient outcomes. However, traditional diagnostic processes can be time-consuming, costly, and require specialized medical expertise. There's a need for accessible, preliminary diagnostic tools that can empower individuals to proactively monitor their health and seek timely medical attention.

1.2 Motivation:

This project was chosen to address the growing need for accessible and efficient preliminary health assessments. The motivation stems from the desire to leverage machine learning and web technologies to create a user-friendly platform that can assist in the early detection of common diseases. The potential applications include empowering individuals to take control of their health, providing a preliminary risk assessment tool for healthcare providers, and facilitating early intervention strategies. The impact lies in potentially reducing the burden on healthcare systems and improving patient outcomes through early detection.

1.3 Objectives:

- **Develop a Streamlit-based web application for predicting diabetes, heart disease, Parkinson's, lung cancer, and hypothyroidism.**
- **Implement machine learning models for each disease using pre-trained and pickled models.**
- **Create a user-friendly interface that simplifies the input of health parameters.**
- **Provide clear and concise prediction results to users.**
- **Deploy the application for easy access and usability.**

1.4 Scope of the Project:

The scope of this project is limited to the development of a web application for preliminary disease prediction using pre-existing machine learning models. The application focuses on five specific diseases: diabetes, heart disease, Parkinson's, lung cancer, and hypothyroidism. Limitations include the reliance on pre-trained models, the absence of real-time data integration, and the focus on preliminary risk assessment rather than definitive diagnosis. The application is intended for informational purposes and does not replace professional medical advice.

CHAPTER 2

Literature Survey

2.1 Review of relevant literature includes studies and research on machine learning applications in medical diagnostics, specifically for diabetes, heart disease, Parkinson's, lung cancer, and thyroid disorders.

2.2 Existing models, techniques, and methodologies include logistic regression, support vector machines, random forests, and neural networks used for disease prediction. Web application frameworks like Streamlit have been explored for creating user-friendly interfaces.

2.3 Gaps in existing solutions include the need for more accessible and integrated platforms for preliminary health assessments. This project addresses these gaps by providing a unified web application that simplifies the prediction process for multiple diseases, making it accessible to a wider audience.

CHAPTER 3

Proposed Methodology

3.1 System Design:

- **Figure 1: System Architecture Diagram**
  - **User Interface (Streamlit) -> Input Parameters -> Machine Learning Models (Pickled) -> Prediction Results -> User Interface (Streamlit)**
  - **Explanation: The user interacts with the application through a Streamlit-based interface, providing input parameters. These parameters are fed into pre-trained machine learning models that generate prediction results, which are then displayed to the user through the interface.**

3.2 Requirement Specification:

3.2.1 Hardware Requirements:

- **Table 1: Hardware Requirements**
  - **Standard personal computer with sufficient RAM and processing power.**

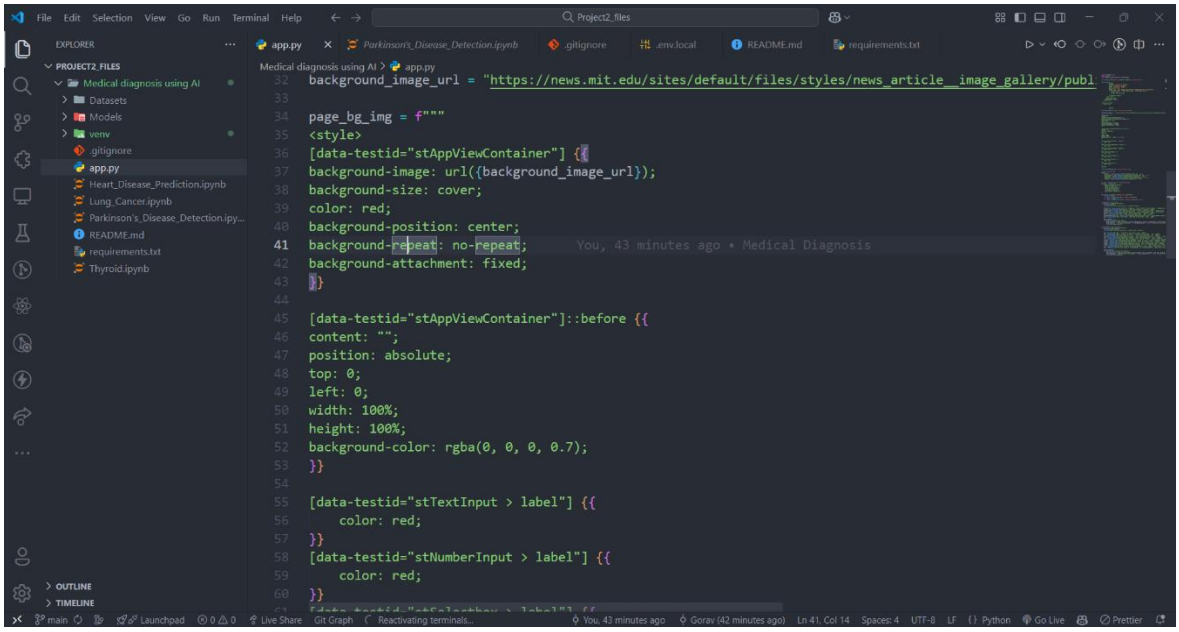
3.2.2 Software Requirements:

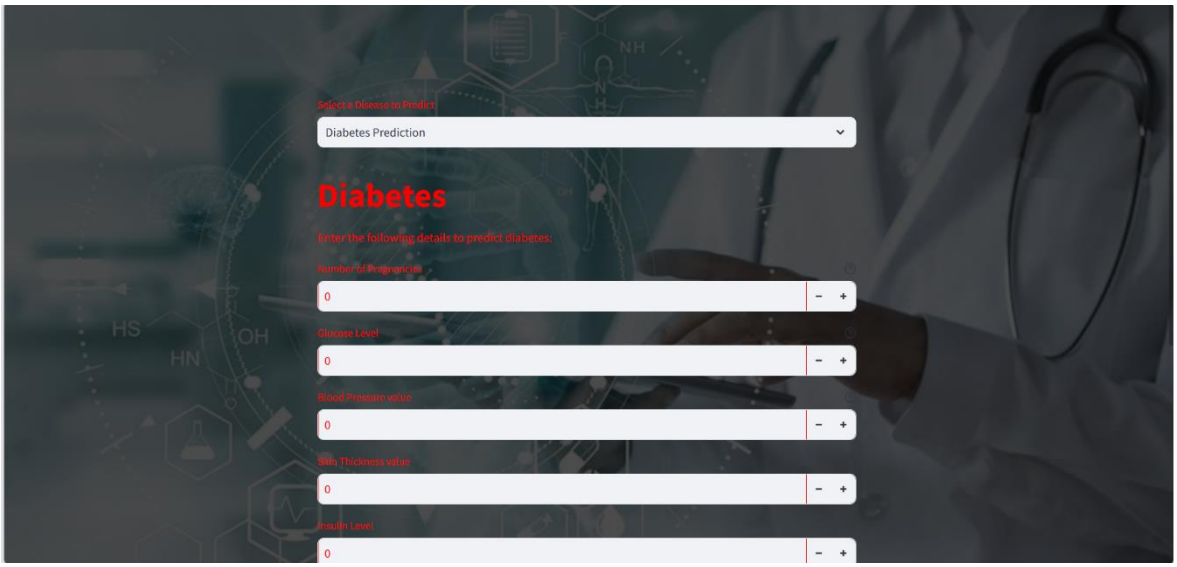
- **Table 2: Software Requirements**
  - **Python 3.x**
  - **Streamlit**
  - **Scikit-learn**
  - **Pandas**
  - **NumPy**
  - **Pickle**

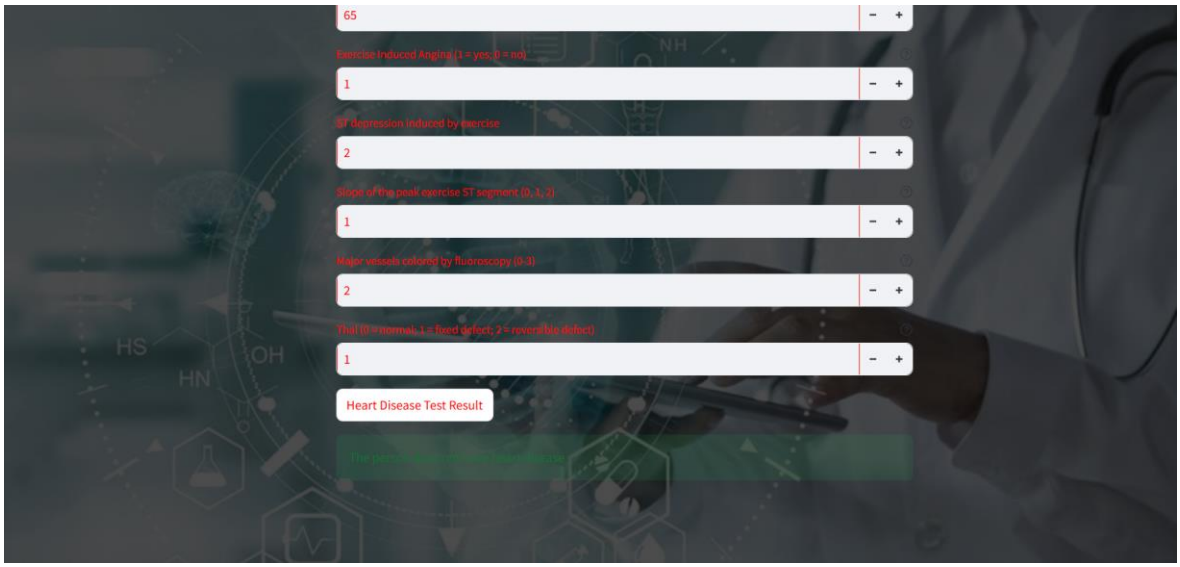
CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:







4.2GitHub Link for Code: [Gorav22/Medical-Diagnosis]

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future improvements include integrating real-time data from wearable devices, expanding the database with more comprehensive datasets, and enhancing the models with advanced machine learning techniques. Addressing unresolved issues may include improving the accuracy of predictions and incorporating user feedback for continuous improvement.

5.2 Conclusion:

This project successfully developed a Streamlit-based web application for preliminary disease prediction, providing a user-friendly and accessible tool for health monitoring. The application demonstrates the potential of machine learning in healthcare and contributes to the growing field of digital health. The project has successfully met its objectives, providing a valuable resource for preliminary health assessments.

REFERENCES

[1]. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002. (Note: Replace this with more relevant Medical machine learning based references.)

[2]. Scikit-learn documentation.

[3]. Streamlit documentation.

[4]. Research papers on the specific machine learning models you used.

[5]. Papers about the datasets you used.