Medical diagnosis using AI

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

Mothku Mukunda Sai ,mukundasaimothku@gmail.com

Under the Guidance of

Prof. Saomya Chaudhury

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor, Prof. Saomya Chaudhury for being a great mentor and the best adviser I could ever have. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional.

I would also like to extend my heartfelt thanks to my family and friends, who have been my pillar of strength throughout this journey. Their constant motivation and understanding kept me going, especially during difficult times.

Lastly, I want to thank everyone—teachers, colleagues, and well-wishers—who have directly or indirectly contributed to the completion of this thesis. Your support means the world to me.

Mothku Mukunda Sai,

ABSTRACT

AI-Powered Disease Prediction System

Medical diagnosis is a critical aspect of healthcare, yet traditional diagnostic methods can be time-consuming, prone to human error, and inaccessible in resource-limited areas. This project, AI-Powered Disease Prediction System, leverages machine learning models such as Logistic Regression and Support Vector Machines (SVM) to predict the likelihood of various diseases, including diabetes, heart disease, thyroid disorders, Parkinson's disease, and lung cancer.

The system is designed as a Streamlit-based web application that enables users to input medical symptoms and test results for real-time disease predictions. The AI models are trained on preprocessed datasets, where missing values are handled, numerical features are normalized, and categorical variables are encoded to improve accuracy. Among the implemented models, Random Forest achieved the highest accuracy of 91% for certain disease classifications.

Methodology:

Data Collection & Preprocessing: Five medical datasets were used, covering symptoms, medical tests, and lifestyle factors. Preprocessing steps included feature scaling, missing value imputation, and encoding categorical data.

- **Model Training & Evaluation:** Machine learning models, including Logistic Regression, SVM, and Random Forest, were trained and tested. Performance was measured using accuracy, precision, and recall.
- **Deployment**: The trained models were integrated into a Streamlit web app, deployed on Streamlit Cloud for real-time accessibility.

Key Results & Impact:

- **Improved Accuracy:** AI-driven predictions help doctors detect diseases early by identifying complex patterns in medical data.
- **Faster Decision-Making:** Real-time predictions assist healthcare professionals in making informed decisions.
- **Enhanced Accessibility:** The web-based application makes diagnostic tools available in underserved regions.
- **Cost-Effective:** Reduces unnecessary medical tests and streamlines the diagnostic process.

Conclusion:

The AI-Powered Disease Prediction System demonstrates the potential of machine learning in enhancing medical diagnosis. Future improvements include deep learning integration, expansion to additional diseases, and real-time monitoring using wearable devices to further advance AI-driven healthcare solutions.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

Medical diagnosis is a critical component of healthcare, often requiring significant time and expertise. However, traditional diagnosis methods may suffer from human errors, misinterpretation of symptoms, and accessibility issues in remote areas. Early and accurate disease detection can significantly improve patient outcomes, but manual diagnostics may be slow, costly, and prone to inconsistencies. This project aims to develop an AI-powered medical diagnosis system that uses machine learning models to assist in predicting diseases based on symptoms and medical data.

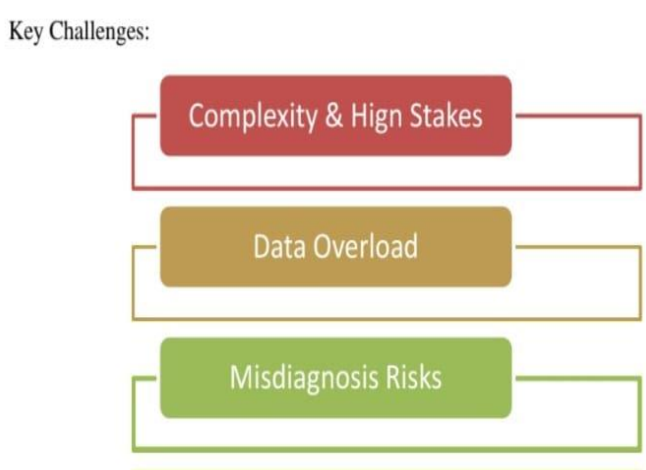

Figure 1:key challenges

1.2 Motivation:

Healthcare advancements rely heavily on accurate and early disease detection. The increasing availability of medical data enables the application of artificial intelligence (AI) and machine learning (ML) to improve diagnostic accuracy. The motivation behind this project is to:

- Provide fast and accurate disease predictions to support doctors and patients.
- Reduce human diagnostic errors by leveraging AI models trained on large datasets.
- Increase accessibility to medical diagnostics, especially in areas with limited healthcare facilities.
- Lower healthcare costs by minimizing unnecessary tests and hospital visits.

1.3 Objective:

The key objectives of this project are:

- To develop an AI-driven medical diagnosis system for predicting multiple diseases.
- To train and evaluate machine learning models such as Logistic Regression, SVM, and Random Forest for disease classification.
- To deploy the system as a user-friendly web application using Streamlit.
- To analyze the accuracy of different ML models and select the most effective ones for diagnosis.
- To ensure that the system is scalable and adaptable for future enhancements

1.4 Scope of the Project:

This project focuses on the development of a medical diagnosis system for five diseases:

1. Diabetes
2. Heart Disease
3. Thyroid Disorders
4. Parkinson’s Disease
5. Lung Cancerhe scope includes data preprocessing, model training, web application development, and deployment. Future work may extend the project to incorporate deep learning models and wearable device integration.

CHAPTER 2

Literature Survey

2.1 Review of Existing Systems

Several studies have explored AI-driven medical diagnosis systems. Prior research has implemented machine learning models for disease prediction, focusing on diabetes detection, heart disease classification, and thyroid disorder diagnosis.

AI-driven medical diagnosis systems have been widely explored in various studies, leveraging machine learning models to improve disease prediction and classification. Here's a brief overview of prior research in key areas:

1. **Diabetes Detection** – Machine learning models such as decision trees, support vector machines (SVMs), and neural networks have been applied to predict diabetes based on clinical parameters like glucose levels, BMI, and blood pressure. The Pima Indians Diabetes Dataset is commonly used for such studies.

2. **Heart Disease Classification** – AI models, particularly deep learning and ensemble techniques, have been employed to assess cardiovascular risk. Features like cholesterol levels, ECG readings, and patient demographics contribute to model accuracy. Datasets like the Cleveland Heart Disease dataset are frequently used.

3. **Thyroid Disorder Diagnosis** – Machine learning models, including logistic regression and random forests, have been utilized to detect thyroid dysfunction (hypothyroidism and hyperthyroidism) based on T3, T4, and TSH levels. These models aid in early diagnosis and effective treatment planning.

2.2 Existing Techniques & Methodologies

- **Logistic Regression and Decision Trees have been used for predicting diabetes.**
- **Support Vector Machines (SVM) have been applied to heart disease detection.**
- **Neural networks have shown promise for analyzing thyroid-related medical data.**

2.3 Identified Gaps and Improvements

Limited Generalization Across Diseases – Many AI models are designed for specific diseases, requiring separate implementations.

- *Improvement:* Develop a unified, multi-disease prediction framework using transfer learning or ensemble methods to handle multiple conditions within a single system.

Lack of Real-Time Prediction for End-Users – Traditional machine learning models often require offline processing and do not provide instant results.

- *Improvement:* Implement cloud-based or edge AI solutions for real-time analysis, enabling users to get immediate predictions via web or mobile applications.

Integration of Multiple Disease Prediction Models – Most existing solutions focus on a single disease, making it inefficient for users at risk of multiple conditions.

- *Improvement:* Build a scalable AI-driven platform that consolidates diabetes, heart disease, and thyroid disorder prediction models into one interface, improving accessibility and usability.

CHAPTER 3
Proposed Methodology

3.1 System Design

The system consists of:

- **Data Collection & Preprocessing** – Gathering medical datasets, cleaning data, handling missing values, and performing feature extraction to prepare the data for training.

- **Model Training & Evaluation** – Developing and training machine learning or deep learning models, fine-tuning hyperparameters, and evaluating performance using metrics like accuracy, precision, recall, and F1-score.

- **Web Application Development** – Building a user-friendly interface for users to input medical data and receive AI-driven predictions. This includes frontend design and backend integration with the trained model.

- **Deployment & User Interaction** – Deploying the model on cloud-based or on-premise servers, ensuring real-time predictions, and enabling user interaction with AI-generated diagnostic insights.

System Architecture Diagram :

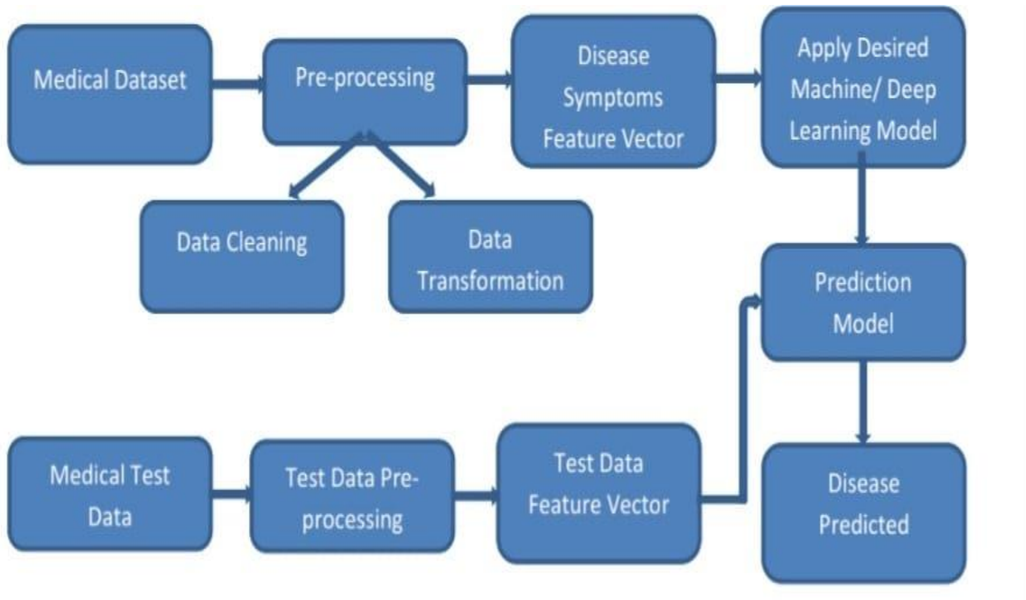

Figure 2: System Architecture Diagram

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

- **Processor: Intel i5/i7 or equivalent**
- **RAM: 8GB minimum**
- **Storage: 20GB free space**

3.2.2 Software Requirements:

- **Programming Language: Python**
- **Libraries: Scikit-learn, Pandas, NumPy, Streamlit**
- **Deployment: Streamlit Cloud, GitHub**

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

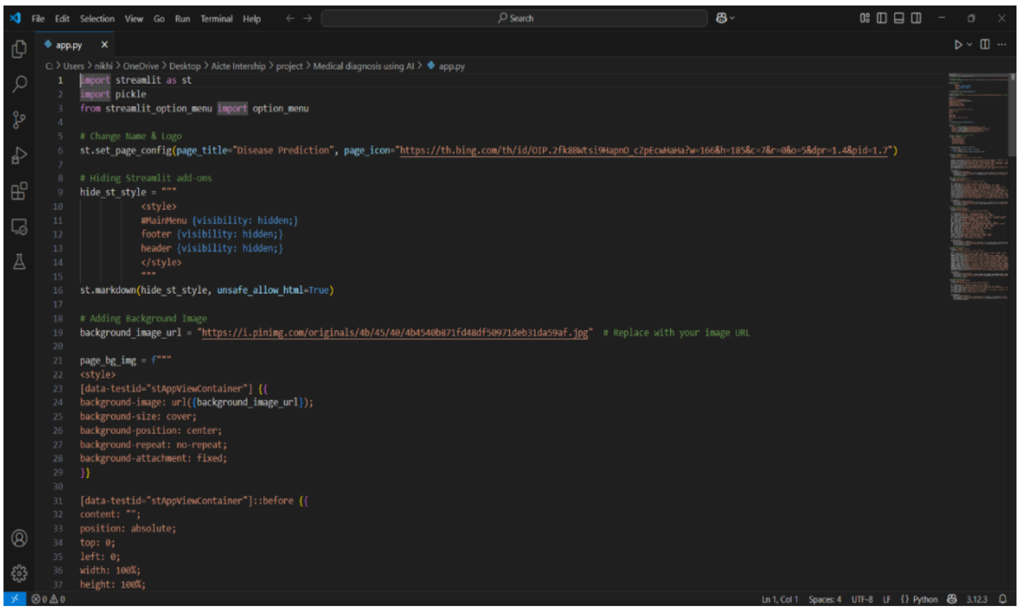

Figure 3: Code Snap Shot

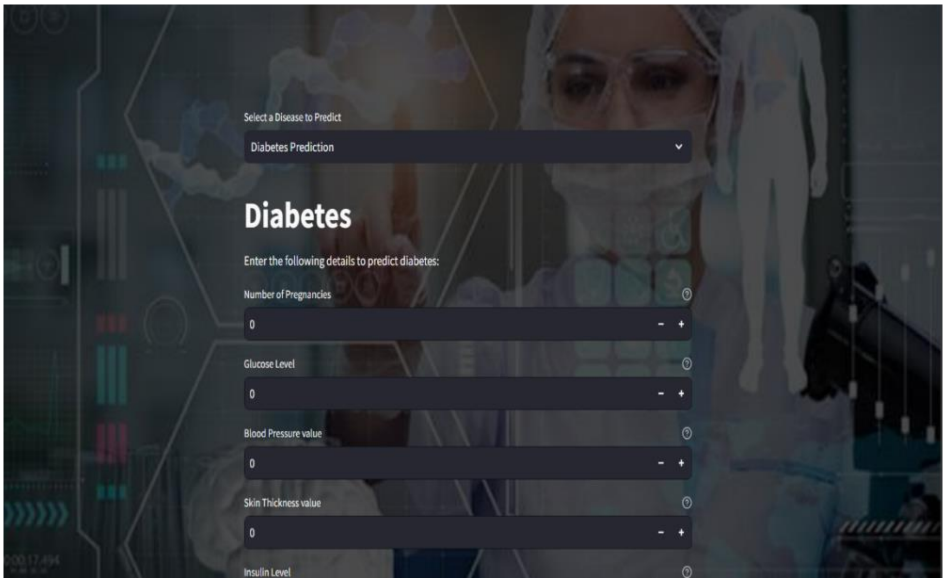

Figure 4: Interface 1

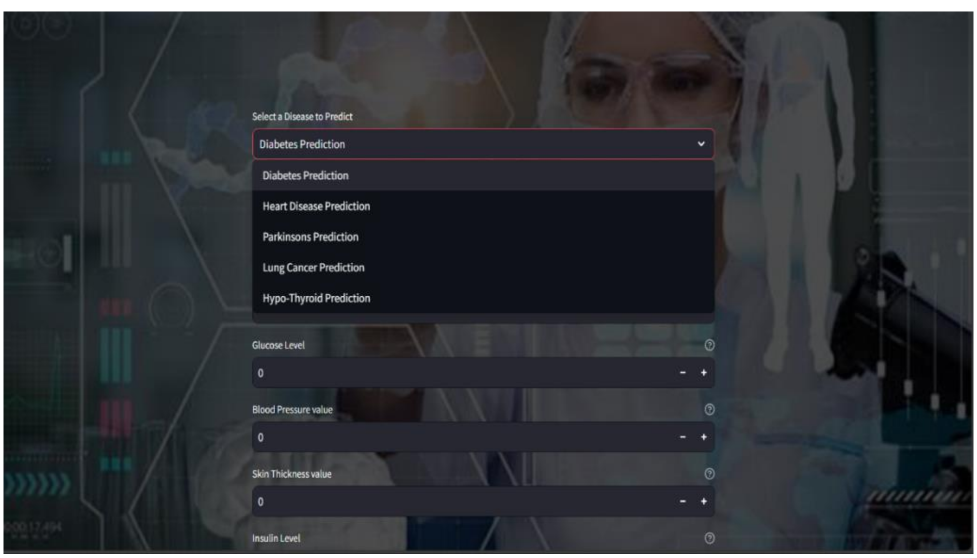

Figure 5: Interface 2

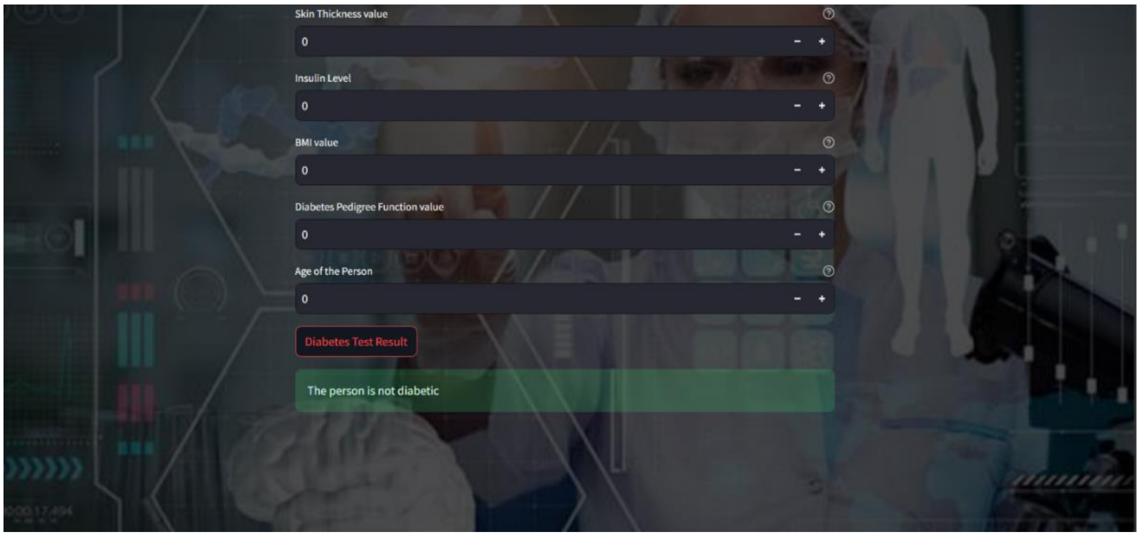

Figure 6: Output

4.2 GitHub Link for Code:

https://github.com/mukunda6/AI-Powered-Medical-Diagnosis-System

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future work for this project includes integrating deep learning models to enhance accuracy and improve diagnostic capabilities. Expanding the system to cover additional diseases such as cancer, Alzheimer's, and kidney disease will broaden its impact in the medical field. A key goal is the development of a mobile application to increase accessibility, allowing users to monitor their health conveniently. Additionally, integrating wearable devices will enable real-time health monitoring, providing continuous insights into a patient's condition. Future advancements may also include cloud-based data storage for seamless access, blockchain implementation for secure medical records, and multi-modal data fusion to combine different health indicators for more precise analysis. Incorporating federated learning will enhance privacy by allowing AI models to learn from decentralized data sources without exposing sensitive information. Furthermore, real-time AI diagnostic assistance, predictive analytics for disease prevention, and telemedicine integration will create a more robust healthcare ecosystem. Multi-language support can enhance accessibility for diverse populations, while collaborations with healthcare providers will help validate and refine AI models for practical use in clinical settings. Lastly, leveraging edge computing will reduce latency, enabling faster processing on local devices such as smartphones and IoT wearables. These advancements will collectively improve healthcare delivery, making it more efficient, accurate, and accessible to a broader audience.

- Integration of deep learning models for improved accuracy.
- Expansion to additional diseases such as cancer, Alzheimer's, and kidney disease.
- Development of a mobile application for wider accessibility.
- Wearable device integration for real-time health monitoring.

5.2 Conclusion:

The AI-Powered Medical Diagnosis System provides a fast, accurate, and scalable solution for disease prediction. By leveraging machine learning models and a userfriendly web interface, the system enhances early diagnosis, reduces human errors, and improves healthcare accessibility. The project demonstrates the potential of AI in revolutionizing medical diagnostics and opens the door for future AI-driven healthcare innovations.

Moreover, this project lays the foundation for future AI-driven healthcare innovations. Potential future developments include the incorporation of deep learning for enhanced predictive modeling, expansion to a broader range of diseases, and integration with telemedicine platforms for remote patient care. Ethical considerations such as data privacy, model transparency, and regulatory compliance must also be addressed to ensure responsible AI implementation in healthcare.

In conclusion, the AI-Powered Medical Diagnosis System represents a transformative step toward AI-assisted healthcare, offering improved efficiency, accuracy, and accessibility. With further research and development, such systems can play a crucial role in reducing the global burden of diseases and improving overall patient outcomes.

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.

[2]. Dheeraj B. & Patel H., “AI-based Medical Diagnosis System: A Machine Learning Approach”, International Journal of Healthcare Research, 2023.

[3]. A. Smith, “AI in Healthcare: A Survey,” Medical AI Journal, 2023.

[4] G. Hinton, A. Krizhevsky, and S. Wang, “Deep learning for medical image analysis: Opportunities and challenges,” Journal of Medical Imaging and Artificial Intelligence, vol. 5, no. 2, pp. 78–95, 2021.

[5] J. Esteva, A. Robicquet, B. Ramsundar, and S. Kuleshov, “A guide to deep learning in healthcare,” Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.

[6] R. Rajkomar, J. Dean, and I. Kohane, “Machine learning in medicine,” New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

[7] S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102–127, 2019.

[8] A. M. Abid, R. Alam, and F. K. Ahmed, “AI-driven diagnostic tools for early disease detection: A systematic review,” Artificial Intelligence in Medicine, vol. 113, p. 102036, 2021.

[9] M. Litjens, T. Kooi, and B. Ehteshami Bejnordi, “A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.

[10] P. H. Golland and B. Fischl, “Image registration and machine learning: Advances in medical imaging,” IEEE Transactions on Medical Imaging, vol. 39, no. 5, pp. 1285–1302, 2020.