Project Report

AI Health Assistant: Disease Risk Prediction using Machine Learning and Streamlit

---

Title Page

Project Title: AI Health Assistant – Disease Risk Prediction with Streamlit
Author: Kashif Ali
Institution: University of Gujrat
Date: October 2025

Abstract

This project presents the development of an **AI-powered Health Assistant** that predicts the risk of common diseases such as **Diabetes** and **Heart Disease**. The system leverages **machine learning models** trained on medical datasets and integrates them into a user-friendly interface using **Streamlit**. The assistant provides predictions along with **confidence scores**, enabling better interpretability of results. While not intended for medical diagnosis, this project demonstrates how AI can support **preventive healthcare** by offering early risk assessment tools.

---

1. Introduction

The growing prevalence of chronic diseases like diabetes and heart disease has made **early detection and prevention** increasingly important. With the rise of **machine learning** and **artificial intelligence**, it is now possible to build systems that assist individuals in understanding their health risks.

This project aims to create an **AI Health Assistant** that predicts disease risks using trained models and provides a simple, modern interface for end-users. The system is designed for **educational and research purposes**, showcasing the application of AI in healthcare.

---

2. Literature Review

Several studies highlight the effectiveness of AI in healthcare:

- Machine learning models such as **Logistic Regression, Decision Trees, and Random Forests** have been widely used for disease prediction.

- Streamlit has emerged as a popular framework for deploying machine learning applications due to its simplicity and clean UI.
- Previous works emphasize the importance of **explainability** in medical AI systems to ensure trust and transparency.

This project builds upon these insights by combining **predictive modeling** with a **user-friendly dashboard**.

---

3. Methodology

The methodology consists of three main stages:

3.1 Dataset

- **Diabetes Dataset**: Contains patient features such as glucose, BMI, insulin levels, etc.
- **Heart Disease Dataset**: Includes attributes like age, cholesterol, chest pain type, blood pressure, etc.

3.2 Preprocessing

- Data cleaning and handling missing values.
- Normalization of input features.
- Splitting into training and testing sets.

3.3 Model Training

- Models: **Logistic Regression, Random Forest, SVM (tested), final models chosen based on accuracy**.
- Evaluation Metrics: **Accuracy, Precision, Recall, F1-Score**.
- Best performing models were saved using **Joblib**.

---

4. Implementation

The project was implemented in **Python 3.9+** using the following tools:

- **Scikit-learn** for training and saving models.
- **Streamlit** for creating a modern, interactive web application.
- **NumPy & Pandas** for data processing.

Features Implemented:

1. **Diabetes Risk Prediction**
    - Inputs: Glucose, BMI, Insulin, etc.
    - Output: Risk level + Probability bar.

2. **Heart Disease Risk Prediction**
    - Inputs: Age, Blood Pressure, Cholesterol, etc.
    - Output: Risk assessment + Confidence score.
3. **User-Friendly UI**
    - Clean layout using Streamlit.
    - Screenshots of prediction results included.

---

5. Results

The trained models achieved the following:

- **Diabetes Model**: Accuracy ~ 82%
- **Heart Disease Model**: Accuracy ~ 85%

The Streamlit app successfully integrates both models and provides interactive risk assessments.

Screenshots of the UI were added to demonstrate functionality (see Figures 1–6).

---

6. Discussion

This project shows the potential of AI in providing **accessible and interpretable health tools**. While the system should not be used for real medical diagnosis, it demonstrates how predictive models can support **preventive healthcare awareness**.

Key strengths:

- Easy-to-use UI for non-technical users.
- Confidence/probability scores improve transparency.
- Modular design allows future extension.

Limitations:

- Dataset size was limited.
- Not clinically validated.

---

7. Future Work

Planned improvements include:

- Adding more diseases (kidney disease, cancer, etc.).
- Integrating a **RAG-based Chatbot** for health advice.

- Developing a **health dashboard** to track progress over time.
- Mobile app version for broader accessibility.
- Cloud deployment (Streamlit Cloud, HuggingFace Spaces, Heroku).

---

8. Conclusion

The **AI Health Assistant** successfully demonstrates the application of machine learning in predicting disease risks with a modern interface. It provides an educational tool that highlights how AI can be used for **early risk awareness** in healthcare. With further development and validation, such systems can play an important role in supporting **preventive medicine**.

---

9. References

1. Pedregosa et al. (2011). *Scikit-learn: Machine Learning in Python*. JMLR.
2. Hunter, J.D. (2007). *Matplotlib: A 2D graphics environment*. Computing in Science & Engineering.