Project Proposal: Medical Diagnosis Recommendation System

(AI-Powered Platform for Assisting Doctors in Disease Detection)

---

1. Introduction

In modern healthcare, doctors face increasing pressure to diagnose diseases quickly and accurately while managing large amounts of patient data. This project aims to develop an **AI-powered medical diagnosis recommendation system** that assists doctors in identifying possible diseases based on patient information.

The system analyzes data such as demographics, symptoms, medical history, and laboratory results to recommend the most probable diseases. By providing ranked suggestions and explanations for each recommendation, the platform enhances clinical decision-making, reduces human error, and improves patient outcomes.

Note: The dataset used in this project is sourced from **SyntheticMass**, a publicly available **synthetic healthcare dataset**. This ensures that all patient information is **artificially generated**, protecting privacy while preserving realistic clinical patterns for model development.

---

2. Project Objectives

The main objectives of this project are:

- To design and prepare a unified medical dataset suitable for building a disease recommendation model.

- To develop and train an AI model capable of recommending the most likely diseases based on patient data.

- To deploy the model through an interactive interface that allows doctors to input patient data and receive real-time recommendations.

---

3. Stakeholders

The development of the **AI-Powered Medical Diagnosis Recommendation System** involves a multidisciplinary team working collaboratively across data engineering, AI modeling, and visualization components. The key stakeholders and their responsibilities are outlined below:

Project Team

- **Data Engineering Team**
  Responsible for data acquisition, cleaning, transformation, and database design using the SyntheticMass dataset.
  Members:
    - **Fares Hussein Noaman**
    - **Youssef Waleed Samir**

- **AI & Machine Learning Engineering Team**
  In charge of model development, training, evaluation, and integration of the disease recommendation engine.
  Members:
    - **Mohamed Hany Abdelhadi**
    - **Mahmoud Ibrahim Farouk**

- **User Interface & Visualization Team**
  Handles the design and implementation of the initial Streamlit application interface and future dashboard development in Streamlit or Power BI.
  Members:
    - **Omar Abdelaal Bakr**
    - **Abdullah Mahmoud Fathy**

Supervision and Guidance

- **Kirolos** *– Academic Supervisor / Project Advisor*
  Provides technical guidance, ensures research and ethical standards are met, and oversees project deliverables.

Data Provider

- **SyntheticMass Project (MITRE Corporation)**
  Provides a publicly available **synthetic healthcare dataset** that simulates realistic patient data while preserving privacy.

---

4. Methodology and Milestones

Milestone 1: Data Collection & System Design

Objective:
Gather and prepare high-quality medical data and design the system's database structure.
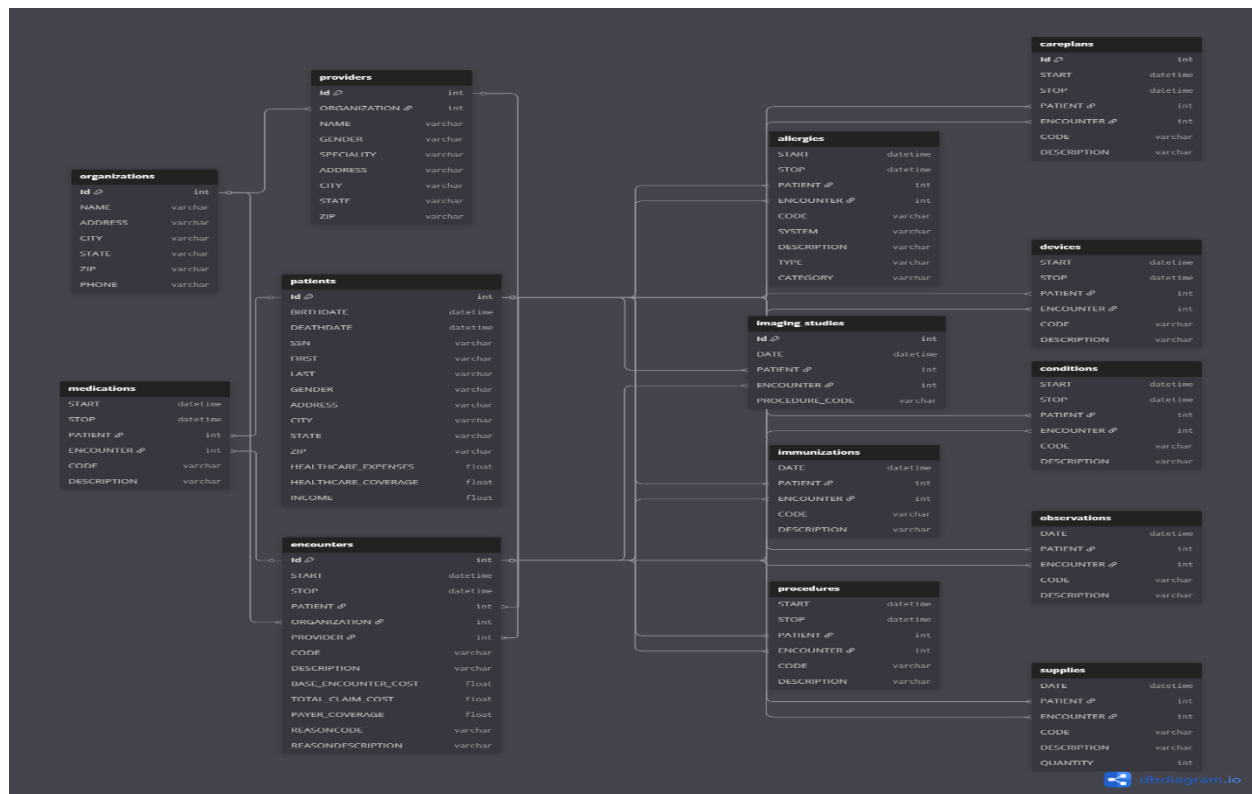
Activities:

- Use **SyntheticMass** data to simulate patient demographics, medical history, lab results, and other relevant records.

- Clean and preprocess the data to handle missing values, duplicates, and inconsistent formats.

- Engineer additional features such as symptom severity scores, comorbidity indicators, and risk factors.

- Design the **Entity Relationship Diagram (ERD)** — as created during this phase — to visualize how different data entities such as patients, encounters, medications, and procedures are interlinked.

- Build a master dataset that combines all patient information in a structured form for model training.

Deliverables:

- Clean, integrated synthetic dataset.

- ERD describing the database relationships.

- Documentation of the schema and feature descriptions.

ERD LINK: https://dbdiagram.io/d/Copy-of-Untitled-Diagram-690660c36735e11170cb7f31

Milestone 2: Model Development & Recommendation Engine

Objective:
Build and evaluate a machine learning model that recommends possible diseases based on patient characteristics.

Activities:

- Perform Exploratory Data Analysis (EDA) to identify correlations and patterns.

- Train and test multi-class and multi-label classification models for disease prediction.

- Implement a recommendation system that outputs the top probable diseases per patient.

- Evaluate performance using accuracy, precision, recall, and F1-score.

- Add explainability techniques to highlight key features influencing predictions.

Deliverables:

- Trained and validated recommendation model.

- Model evaluation and explainability reports.

- Codebase for reproducibility and analysis.

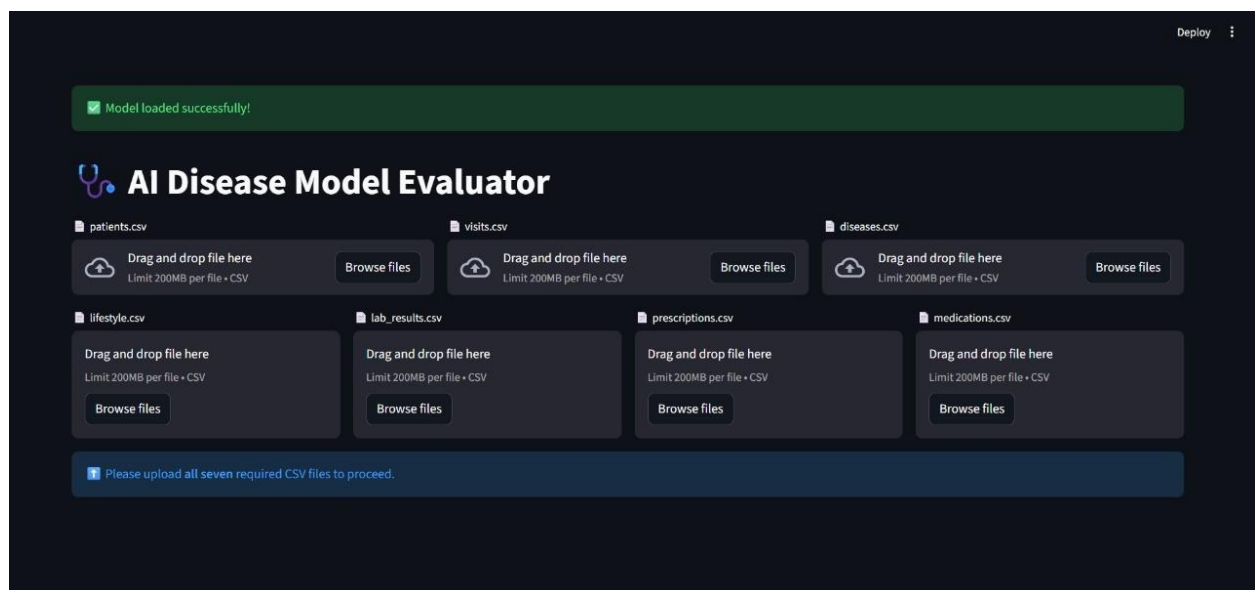**Milestone 3: Deployment & Visualization**

Objective:
Deploy the AI model with an interactive interface and build visual tools for insights.

Activities:

- Develop an initial **Streamlit-based web UI**, allowing users to upload patient data and view disease predictions.

- Integrate the model for real-time recommendations.

- Design plans for a **dashboard** (to be developed either in **Power BI** or **Streamlit**) to visualize key metrics such as disease frequency, patient distribution, and model performance.

- Include alert mechanisms for high-risk cases.

Deliverables:

- Initial Streamlit UI prototype.

- Plans and mockups for a full-featured analytical dashboard.

- Demonstration-ready deployment of the model.

5. Expected Outcomes

By the end of this project, we expect to deliver:

- An AI-powered disease diagnosis recommendation system built on synthetic, privacy-safe data.

- An interpretable, data-driven model that assists doctors in making faster, evidence-based decisions.

- An initial working UI built in Streamlit and a roadmap for developing a comprehensive Power BI or Streamlit dashboard.

---

6. Project Timeline

	Milestone Description	Main Deliverables
1	Data Collection & System Design	Synthetic dataset, ERD, documentation
2	Model Development & Recommendation Engine	Trained model, evaluation, explainability
3	Deployment & Visualization	Streamlit app (initial), dashboard plan, final demo

---

7. **Conclusion**

This project bridges the gap between artificial intelligence and healthcare expertise. Using **SyntheticMass synthetic data**, the system ensures ethical and secure model development while maintaining clinical relevance.
Through data-driven insights, explainable AI, and interactive visualization, this system aims to become a valuable decision-support tool for healthcare professionals — improving diagnostic efficiency and patient care quality.