**Transforming healthcare with AI powered disease prediction based on patient data**

**Student Name:**Bavithira.S

**Register No:**731123104006

**Institution:** Government college of engineering,Erode

**Department:** Computer Science and Engineering

**Date of Submission:** April 26, 2025

**1. Problem Statement**

The early and accurate prediction of diseases can save millions of lives and reduce healthcare costs. However, manual diagnosis is time-consuming, prone to errors, and heavily dependent on specialist availability. This project aims to leverage artificial intelligence (AI) techniques to predict diseases based on patient data such as medical history, symptoms, laboratory results, and demographic information. By automating disease prediction, healthcare systems can improve patient outcomes, optimize resource allocation, and deliver personalized care.

**2. Objectives of the Project**

* Develop an AI model capable of predicting diseases based on patient data.
* Enhance early diagnosis and treatment planning.
* Minimize misdiagnosis through advanced machine learning techniques.
* Create a user-friendly interface to showcase predictions and explainable AI insights.

**3. Scope of the Project**

**Features to build/analyze:**

* Patient data preprocessing and feature engineering
* Disease prediction models
* Risk scoring and prioritization
* Basic visualization dashboards

**Limitations/Constraints:**

* Dataset limited to historical patient records.
* Focus primarily on predicting common diseases (e.g., diabetes, heart disease, cancer).
* Predictions are for academic purposes only, not for clinical use.
* No real-time hospital integration in this phase.

**4. Data Sources**

* **Dataset Source:** Kaggle (e.g., Heart Disease UCI dataset, Diabetes Prediction Dataset)
* **Dataset Type:** Public
* **Nature:** Static (downloaded and used for modeling)

**5. High-Level Methodology**

* **Data Collection:** Download structured patient datasets.
* **Data Preprocessing:** Handle missing values, encode categorical features, normalize data.
* **Exploratory Data Analysis (EDA):** Understand key variables, detect imbalances, visualize patterns.
* **Feature Engineering:** Derive new features (e.g., BMI from height and weight).
* **Model Building:** Train models like Logistic Regression, Random Forest, XGBoost, and Neural Networks.
* **Model Evaluation:** Use metrics like ROC-AUC, Precision, Recall, F1-Score.
* **Interpretation:** Use SHAP or LIME for explainable AI to interpret model decisions.
* **Visualization:** Create simple dashboards showing patient risk levels and top predictors.

**6. Tools and Technologies**

* **Programming Language:** Python
* **Notebook/IDE:** Google Colab / Jupyter Notebook
* **Libraries:**
  + Data Processing: pandas, numpy
  + Visualization: matplotlib, seaborn, plotly
  + Modeling: scikit-learn, XGBoost, TensorFlow/Keras
  + Explainable AI: SHAP, LIME

**7. Team Members and Roles**

| **Name** | **Role** |
| --- | --- |
| Rasika | Data Collection, Preprocessing |
| Bavithra | Exploratory Data Analysis, Feature Engineering | |
| Gopika sri | Model Building and Optimization |
| Bindhu | Model Evaluation, Explainable AI Analysis |
| Pavithra Malini | Visualization, Report Preparation |