Heart Disease Risk Prediction System

1. Introduction

This project implements an end-to-end Heart Disease Risk Prediction System for CHUB Hospital. The system predicts five levels of heart disease risk using machine learning models and deploys the best-performing model through a Flask REST API with a responsive web interface.

2. Dataset Description

The dataset consists of 5000 patient records with 13 clinical, demographic, and diagnostic features. The target variable contains five classes: No Disease, Very Mild, Mild, Severe, and Immediate Danger.

3. Exploratory Data Analysis

EDA was performed to understand data distribution, missing values, outliers, and class balance. Statistical summaries, correlation heatmaps, boxplots, and class distribution charts were generated.

4. Data Preprocessing

Numerical features were imputed and scaled using StandardScaler. Categorical features were imputed using the most frequent value and encoded with OneHotEncoder. A ColumnTransformer combined both pipelines.

5. Model Training and Evaluation

Multiple models (MLP, Random Forest, SVM, KNN, Gradient Boosting) were trained using GridSearchCV. Models were compared based on cross-validation accuracy, test accuracy, and overfitting gap. The best-performing model was selected for deployment.

6. Deployment

The selected model was deployed using Flask as a REST API with prediction and information endpoints. A responsive HTML frontend allows clinicians to input patient data and view predictions with confidence scores.

7. Conclusion

This system provides an intelligent clinical decision support tool for early detection and risk assessment of heart disease, improving efficiency and consistency in clinical evaluation.