Transformation of Healthcare with AI-Powered Disease Prediction

Based on Patient Data

Phase-2 Submission

- **TRANSFORMATION HEALTHCARE WITH AI-POWERED DISEAS PREDICTION BASED ON PATIENT DATA**

1.STUDENTS NAME & REGISTER NUMBER:

K. SOFIYADEVI(422223106032)

2. COLLEGE NAME:

SURYA GROUP OF INSTITUTION

3. DEPARTMENT NAME:

ELECTRONIC &COMMUNICATION ENGINEERING

4. DATE OF SUBMISSION:

03.05.2025

GitHub Repository Link

- https://github.com/sofiyadevi/sofiyadevi

Abstract

- Using AI models to predict diseases based on patient data.
- Helps in early diagnosis and better healthcare.
- Uses ML techniques like Random Forest, Decision Tree, SVM.

Problem Statement

- The project addresses the need for early and accurate disease prediction using artificial intelligence models trained on real patient data, which includes health records, lab results, and wearable sensor information.
- **Type of Problem:** Classification (e.g., Disease risk prediction), Regression (e.g., predicting future biomarker levels)
- **Impact and Relevance:** AI-powered disease prediction improves early diagnosis, enhances patient outcomes, reduces hospital burden, and enables personalized healthcare.

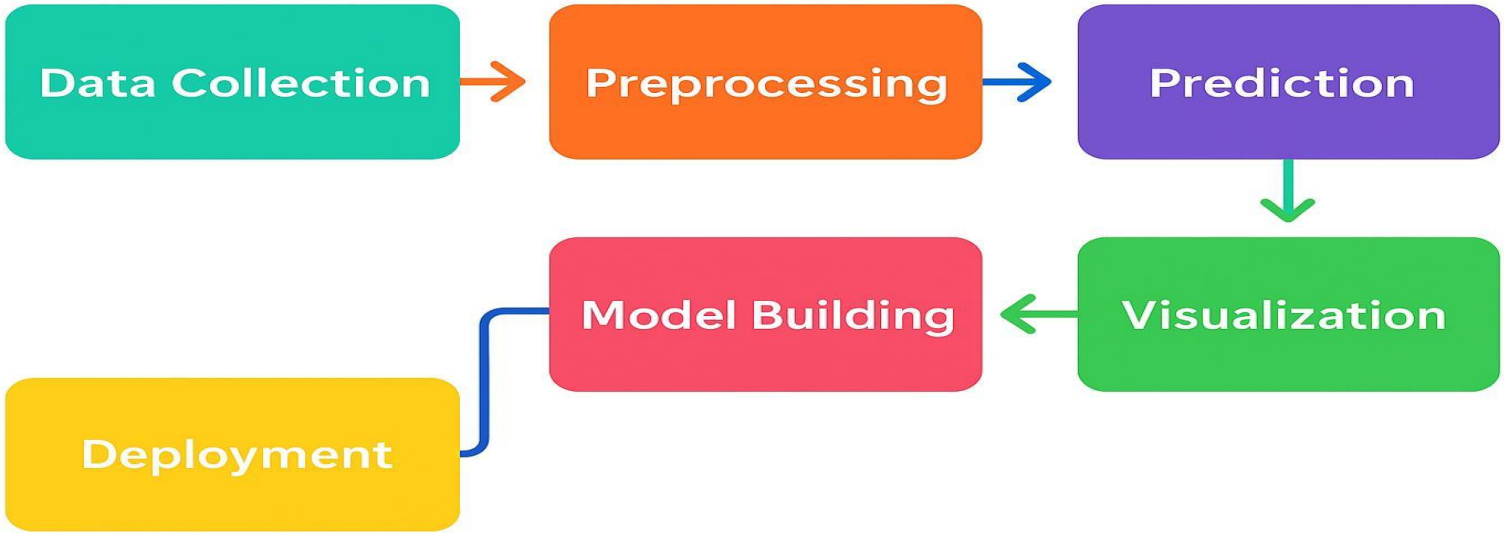
Objectives

- **Technical Objectives:**
- Collect and preprocess healthcare datasets (EHRs, sensor data).
- Build models for disease classification and risk prediction.
- Evaluate the model for accuracy and reliability.
- **Model Aims:**
- Achieve high prediction accuracy for diseases like diabetes, heart conditions, and cancer.
- Ensure clinical interpretability and practical utility of predictions.

Methodology

- Collect patient data from open-source databases and simulations.
- Clean and preprocess the data (handle missing values, normalize features).
- Apply feature engineering techniques.
- Build machine learning models (Random Forest, Decision Tree, SVM).
- Evaluate model performance using accuracy, precision, recall.
- Deploy the best model for real-time disease prediction.

System Architecture



Data Collection: Gathering patient details like medical history and lab reports.
Data Preprocessing: Cleaning, normalizing, and feature engineering.
Model Building: Training ML models like Random Forest, Decision Tree, etc.
Prediction: Predicting disease based on new patient data.
Visualization: Displaying prediction results via dashboard.
Deployment: Making the model accessible through APIs for hospitals.

Data Collection and Preprocessing

- **Data Sources:** Online public healthcare datasets, simulated patient records.
- **Preprocessing Steps:** Handling missing and incorrect data entries
- Normalizing numeric fields like blood pressure, glucose levels.
- Encoding categorical variables like gender, smoker status.

Model Development

- **Algorithms Used:**
- Random Forest Classifier
- Decision Tree Classifier
- Support Vector Machine (SVM)
- Logistic Regression (optional)
- **Model Evaluation Metrics:**
- Accuracy
- Precision
- Recall
- F1-Score
- Best model selected based on highest F1-Score.

Results and Discussion

- Random Forest model achieved the highest accuracy (example: 92%).
- Decision Tree performed well but slightly lower accuracy (example: 87%).
- Visualization of confusion matrix and ROC curve proved model efficiency.
- Future scope includes real-time data integration and continuous learning models.

Tools and Technologies

- **Programming Language:** Python
- **Libraries:** Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn
- **Platform:** Google Colab / Jupyter Notebook
- **Version Control:** GitHub
- **Data Storage:** Local Storage / Cloud

Team Members & Contributions

1. [M. Irusammal] – Project Lead & ML Engineer

- Led data collection, model building, and evaluation.
- **2. [N. Jayapratha] – Data Analyst & EDA Specialist**
- Conducted data cleaning, EDA, and visualization.
- **3. [K. Sofiyadevi] – Healthcare Domain Analyst**
- Helped interpret data features and align model results with medical relevance.

Thank You