Phase 1

Student Name: S.Thirulochine

Register Number: 620123106118

Institution: AVS Engineering College

Department: ECE

Date of Submission: 30-04-2025

---

1. Problem Statement

The healthcare industry often struggles with timely diagnosis due to the large volume of patient data and shortage of medical professionals. Delays in detecting diseases can lead to worsened health outcomes and increased healthcare costs. This project aims to leverage AI-driven predictive models to analyze patient data and forecast potential diseases, enabling early intervention and improved patient care. The solution addresses a pressing need for scalable, data-driven tools in the medical domain.

2. Objectives of the Project

- To develop an AI model that predicts diseases based on patient data.
- To identify key health indicators that significantly influence disease diagnosis.
- To improve accuracy and efficiency in early disease detection.
- To provide visual insights that aid healthcare professionals in understanding the predictions

3. Scope of the Project

- Features to Analyze/Build:

- Patient demographics and health indicators
- Machine learning classification algorithms
- Interactive dashboards for visualization

- Limitations/Constraints:

- Reliance on publicly available datasets
- Use of standard supervised learning models
- Initial deployment limited to notebook-based interfaces

4. Data Sources

- The project will use datasets from publicly available sources such as Kaggle or UCI Machine Learning Repository, which include patient demographics, symptoms, and diagnostic results. These datasets are static and will be downloaded once at the start of the project.

5. High-Level Methodology

- Data Collection: Download datasets from Kaggle/UCI repositories.
- Data Cleaning: Handle missing values, remove duplicates, normalize data formats.
- Exploratory Data Analysis (EDA): Use graphs, histograms, and correlation matrices to identify trends and anomalies.
- Feature Engineering: Create derived features like BMI from weight and height, age categories, etc. Model Building: Apply algorithms like Logistic Regression, Decision

- Trees, Random Forest, and SVM for prediction.

Model Evaluation: Use accuracy, precision, recall, F1-score, and ROC-AUC for performance evaluation.

Visualization & Interpretation: Use matplotlib and seaborn for data visualization; display prediction results in charts.

Deployment: The model will be integrated into a Streamlit-based web interface for demo purposes.

6. Tools and Technologies

Programming Language: Python

Notebook/IDE: Google Colab

Libraries: pandas, numpy, seaborn, matplotlib, scikit-learn, TensorFlow

Optional Tools for Deployment: Streamlit

7. Team Members and Roles

- S.Thirulochine: Project lead, responsible for model development and documentation.
- V.Sandhiya: Data preprocessing, EDA, and visualization tasks.
- M.Murshitha: Research on healthcare datasets and deployment using Streamlit