**Transforming Healthcare with AI-powered Disease Prediction**

**Phase1**

**STUDENT NAME:E AAKASH**

**REGISGER NUMBER:620523243001**

**INSTITUTION:CMS COLLEGE OF ENGINEERING**

**DEPARTMENT:ARTIFICIAL INTELLIGENCE AND DATA SCIENCE**

**DATE OF SUBMISSION:**

**problem Statement**

**Modern healthcare systems often face challenges in early and accurate disease detection due to limited access to real-time diagnostics, human error, and high patient loads. Delayed diagnosis can lead to increased morbidity, mortality, and healthcare costs**

**Objective of the Project**

**To develop an AI-powered system that predicts diseases based on patient data (e.g., symptoms, medical history, lab results) to assist healthcare providers in making accurate and timely diagnoses.**

**Scope of the Project**

**- Focus on predictive modeling for common diseases (e.g., diabetes, heart disease, cancer, COVID-19)**

**.- Use structured patient data for model training and testing.- Implement a user-friendly interface for clinicians.**

**- Evaluate the models accuracy, precision, recall**

**Features to Analyse**

**- Patient demographics (age, gender, etc.)**

**- Medical history and family history**

**- Vital signs (blood pressure, heart rate, etc.)**

**- Laboratory test results- Reported symptoms and duration**

**- Lifestyle factors (smoking, alcohol use, physical activity**

**Constraints and Limitations**

**- Data availability and privacy concerns (HIPAA, GDPR)**

**- Bias in historical data**

**- Interpretability of AI models (especially deep learning)**

**- Limited generalization to rare or emerging diseases**

**- Dependence on high-quality, labeled data**

**Data Sources**

**- Public health datasets (e.g., UCI Machine Learning Repository, Kaggle datasets)**

**- WHO and CDC databases**

**- Hospital EHRs (Electronic Health Records) if accessible under compliance**

**- Government health department**

**High-level Methodology**

**1. Data Collection Acquire and consolidate relevant patient datasets**

**2. Data Preprocessing Clean, normalize, and encode features**

**3. Exploratory Data Analysis (EDA) Visualize and understand trends**

**4. Feature Engineering Derive and select meaningful features**

**5. Model Selection Train ML models (e.g., Random Forest, XGBoost, Neural Networks)**

**6. Evaluation Use metrics like accuracy, AUC, precision/recall**

**7. Deployment Integrate into a clinical decision support system**

**Tools and Techniques**

**- Languages: Python, R**

**- Libraries: Scikit-learn, TensorFlow, Keras, Pandas, NumPy**

**- Visualization: Matplotlib, Seaborn, Plotly**

**- Platforms: Jupyter Notebook, Google Colab, AWS/GCP for deployment**

**- Frameworks: Flask/Django (for web interface), Streamlit (for dashboarding**