**Functional Document**

1. Introduction

The **Disease Prediction System** is a clinical decision support tool that leverages **Machine Learning (ML)** and **Deep Learning (DL)** to provide accurate disease detection and prediction. This sprint focuses on building and integrating prediction models for **Chronic Kidney Disease, Diabetes (Diagnosis and Readmission), Heart Disease, Liver Disease, Parkinson’s, and Sepsis**, using a unified **Streamlit** dashboard. Each disease module is tailored to medical relevance and dataset availability, ensuring effective clinical support and explainability for healthcare professionals

1. Product Goal

The overarching goal is to deliver an end-to-end, AI-powered healthcare platform that provides:

* Accurate disease risk predictions using ML/DL tailored to each condition
* Real-time interaction through a user-friendly web interface (Streamlit)
* Model transparency, with explanations and clinical insight visualizations
* Modular design to allow scalability across diseases and regions

Each prediction module aligns with preventive care, resource optimization, and personalized treatment planning.

1. Demography (Users, Location)

**Users**

* **Target Audience:**
  + Physicians (Nephrologists, Cardiologists, Endocrinologists, Neurologists, ICU doctors)
  + Nurses and medical support staff
  + Hospital administrators and researchers
* **User Characteristics:**
  + Clinicians with basic or intermediate digital literacy
  + Need for quick risk stratification at point-of-care
  + Operating in environments with or without advanced diagnostic tools

**Location**

* **Primary Use Regions:**
  + Urban hospitals and specialty clinics
  + Rural healthcare centers with limited diagnostic resources
  + Remote telemedicine setups
  + Research institutions and academic medical centers

1. Business Processes

The system integrates with hospital workflows to support the following processes:

1. **Patient Intake and Data Entry**
   * Lab reports and vitals are entered manually or imported via EHR systems.
   * Includes both static (e.g., age, sex) and dynamic (e.g., lab test results) parameters.
2. **Disease-Specific Risk Assessment**
   * Backend ML/DL models process input data to provide real-time probability scores or classifications.
   * Each model corresponds to a specific disease, tailored by clinical features.
3. **Prediction Visualization and Feedback**
   * Risk outputs are visualized as colored progress bars, numerical scores, or classification tags (e.g., “High Risk”, “No Risk”).
   * Models are explainable via diagrams (e.g., LSTM/GNN architecture) and feature importance plots.
4. **Clinical Action**
   * Results guide physicians in initiating further testing, medication adjustments, or patient education.
   * Supports early detection, reducing unnecessary procedures (e.g., biopsies or angiograms).
5. **Model Training/Updating (Admin-Level)**
   * First-time system setup includes automatic training of all models.
   * Admin users can retrain models with updated datasets if required.
6. Features

**5.1 Feature #1: Disease Prediction Dashboard (Streamlit UI)**

**1. Description**  
A centralized dashboard built using **Streamlit** that integrates prediction models for various diseases. Each disease module appears as a tab or section, offering dedicated forms for input and displaying results in real-time.

**2. User Story**  
*As a physician, I want to access a single web interface where I can switch between disease-specific modules and receive predictions after entering patient data.*

**5.2 Feature #2: LSTM-Based Disease Prediction**

**1. Description**  
Implementation of **Long Short-Term Memory (LSTM)** models for diseases where time-series or sequential dependencies are crucial, such as:

* Diabetes Readmission Risk Prediction
* Heart Disease (variant using LSTM)
* Parkinson’s Disease Progression

**2. User Story**  
*As a data analyst in a hospital, I want to use models that capture time-related patterns (e.g., prior admissions, symptom progression) to better understand chronic patient risks.*

**5.3 Feature #3: Graph Neural Network (GNN) Prediction Module**

**1. Description**  
A module that uses **GNN-based models** for predicting **total\_UPDRS** in Parkinson’s Disease and **Heart Disease** classification. Graph structures are applied to represent patient health parameters as nodes and their interdependencies as edges.

**2. User Story**  
*As a healthcare researcher, I want to use GNN-based models to assess Parkinson’s severity more accurately and explore feature relationships visually.*

**5.4 Feature #4: Clinical Explainability Tools**

**1. Description**  
Explainability is integrated into each module using:

* Feature importance visualizations (e.g., SHAP, attention weights)
* Visual architecture diagrams for each model (LSTM, GNN, etc.)
* Color-coded risk indicators (e.g., green for low, red for high)

**2. User Story**  
*As a clinician, I want to understand which factors contributed most to a prediction, so I can make informed medical decisions and explain them to the patient.*

**5.5 Feature #5: Automatic Model Training on First Use**

**1. Description**  
On system initialization, all models are trained automatically from respective datasets. Model files and preprocessing tools (e.g., scaler.pkl, encoders.pkl) are saved and reused across future sessions.

**2. User Story**  
*As an admin, I want the system to automatically train models the first time it runs, so I don't have to configure training pipelines manually.*

**5.6 Feature #6: Model Architecture Diagrams**

**1. Description**  
Each prediction module includes a static or interactive visualization of the model architecture (e.g., GNN layers, LSTM flow) to provide insights into how predictions are generated.

**2. User Story**  
*As a healthcare researcher, I want to see how a model is architected to ensure transparency and scientific validity.*

**5.7 Feature #7: Modular Disease Forms with Custom Inputs**

**1. Description**  
Each disease prediction module includes its own tailored input form within the dashboard. Input fields match the clinical dataset features (e.g., blood glucose, creatinine, ECG results), making it easy to replicate real-world diagnosis.

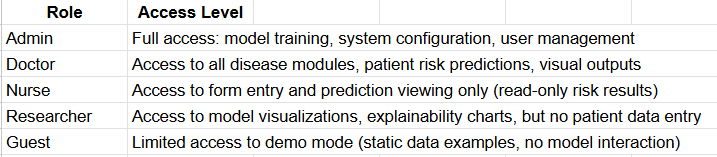
**2. User Story**  
*As a doctor, I want to input specific lab values and health indicators depending on the disease, so that predictions are accurate and clinically meaningful.*

**5.8 Feature #8: Backend Prediction API (Optional / Extendable)**

**1. Description**  
Though the Streamlit UI is the primary interface, models are structured in a way that backend APIs can be easily added for future integration with hospital systems or third-party apps.

**2. User Story**  
*As a hospital IT admin, I want to plug this system into our existing patient portal in the future, using secure backend endpoints.*

1. Authorization Matrix



1. Assumptions

 The Streamlit-based interface will remain compatible with all models during the sprint.

 Each model will function independently and not rely on interdependent services during inference.

 All model training scripts will execute correctly with the available datasets on first use.

 Clinicians will input clean and pre-validated data; basic input validation will be implemented on forms.

 Admin users have access to the environment with necessary libraries (TensorFlow, XGBoost, PyTorch Geometric, etc.).

 Hospital IT infrastructure supports deploying this as a standalone web application or within an intranet.

 No real-time EHR integration is assumed in this sprint; patient data is entered manually.