**Software Requirements Specification (SRS)**

**SehatAI**

AI-powered Healthcare Support System for Rural Pakistan  
**Version:** 1.0  
**Date:** September 2025

**📑 Table of Contents**

1. Introduction
2. Problem Statement
3. Project Overview (What, Why, Where, How)
4. Proposed Solution
5. Types of Users
6. Functional Requirements
7. Non-Functional Requirements
8. Project Scope
   * In-Scope
   * Out-of-Scope
9. Technology Stack & Resources
   * Primary Choice
   * Secondary Choice (Alternative)
10. Data Sources
11. Project Timeline (14 Days Plan)
12. Summary

**1. Introduction**

SehatAI is an **AI-powered healthcare assistant** designed to support rural clinics and public health facilities in Pakistan. It focuses on three core tasks:

1. **AI-based X-ray analysis** (detecting TB, pneumonia).
2. **Prescription digitization** (OCR for Urdu & English handwritten text).
3. **Risk scoring for common diseases** (diabetes, heart disease).

The project uses **open-source datasets, libraries, and tools** only, in compliance with the Techathon guidelines

guidelines\_uraan\_AI

.

**2. Problem Statement**

* Rural clinics lack diagnostic equipment and expert doctors.
* Prescriptions are handwritten, making them error-prone and difficult to store or validate.
* Health records are mostly paper-based, leading to inefficiency and mismanagement.
* Predictive monitoring of diseases like **diabetes and heart disease** is almost non-existent in rural setups.

This results in **delayed diagnosis, poor patient management, and increased health risks** for millions of people.

**3. Project Overview**

**What (Objective):**

Build an AI tool that:

* Analyzes chest X-rays for early disease detection.
* Converts handwritten prescriptions into digital format.
* Predicts health risks based on demographic and lifestyle data.

**Why (Purpose):**

* Improve **access to healthcare** in underserved rural areas.
* Enable **digital health records** for every citizen.
* Provide doctors with **decision support tools** for faster diagnosis.

**Where (Usage Areas):**

* Rural health clinics.
* Public sector hospitals.
* Government health monitoring programs.
* Future: NADRA e-Health integration.

**How (Method):**

* Use **pre-trained open-source AI models** for X-rays and OCR.
* Train lightweight models on public datasets.
* Deploy simple web/mobile interfaces for doctors and clinics.

**4. Proposed Solution**

**SehatAI = “Doctor’s AI Assistant”**

Features:

1. **X-ray Analyzer:** CNN model that classifies X-rays for TB/pneumonia.
2. **Prescription Reader:** OCR pipeline for Urdu/English text recognition.
3. **Risk Scoring:** Logistic regression / ML model to calculate health risk scores.
4. **Dashboard:** Displays patient reports, risk levels, and digitized prescriptions.

**5. Types of Users**

* **Primary Users:**
  + Doctors in rural/public clinics.
  + Paramedical staff who handle X-rays and prescriptions.
* **Secondary Users:**
  + Patients (to view their reports via ID).
  + Government health officers (for disease monitoring).

**6. Functional Requirements**

1. Upload chest X-ray → get TB/pneumonia result (accuracy ≥ 85%).
2. Upload prescription image → OCR converts into editable text (accuracy ≥ 80%).
3. Enter demographic data → system predicts disease risk score (AUC ≥ 0.8).
4. Dashboard to manage patients, prescriptions, and results.
5. Export patient data in PDF/CSV.

**7. Non-Functional Requirements**

* **Usability:** Simple, bilingual (English + Urdu).
* **Performance:** Models must run on CPU (Phase-1 requirement).
* **Security:** Basic authentication for data access.
* **Scalability:** Phase-2 can expand to cloud/GPU deployment.
* **Compliance:** Use only open-source datasets/libraries.

**8. Project Scope**

**In-Scope**

* AI models for chest X-rays, OCR, and risk scoring.
* Dashboard/web app prototype.
* Use of open datasets only.
* Documentation, demo video, and GitHub submission.

**Out-of-Scope**

* Full integration with NADRA.
* Real-time hospital deployment.
* Advanced GPU training (reserved for Phase-2).

**9. Technology Stack & Resources**

**AI/ML Frameworks**

* **Best Choice:** PyTorch (open-source, easy deployment).
* **Alternative:** TensorFlow/Keras.

**OCR for Prescriptions**

* **Best Choice:** Tesseract OCR (supports Urdu + English).
* **Alternative:** EasyOCR (open-source, multilingual).

**X-ray Analysis (CNN Models)**

* **Best Choice:** ResNet / DenseNet (pre-trained on ImageNet, fine-tuned).
* **Alternative:** VGG16 or EfficientNet.

**Risk Scoring Models**

* **Best Choice:** Logistic Regression / Random Forest (Scikit-learn).
* **Alternative:** LightGBM / XGBoost.

**Dashboard / Frontend**

* **Best Choice:** Streamlit (fast prototyping, easy for Phase-1 demo).
* **Alternative:** Flask/Django (for web APIs).

**Data Handling**

* **Best Choice:** Pandas + NumPy.
* **Alternative:** Apache Arrow.

**10. Data Sources (Open & Public)**

1. **Chest X-rays:** NIH Chest X-ray Dataset, Kaggle Tuberculosis Dataset.
2. **Prescriptions (OCR):** Synthetic Urdu + English handwritten text datasets, Urdu handwritten dataset (UCOM).
3. **Risk Scoring:** Pakistan Demographic and Health Survey (PDHS), WHO global health survey datasets.

**11. Project Timeline (14 Days Plan)**

**Day 1–2:** Finalize datasets, assign roles.  
**Day 3–5:** Train baseline CNN (X-ray), test OCR on sample prescriptions.  
**Day 6–7:** Implement risk scoring model with survey data.  
**Day 8–9:** Build prototype dashboard (Streamlit).  
**Day 10:** Integrate models into dashboard.  
**Day 11–12:** Optimize accuracy, refine datasets.  
**Day 13:** Prepare technical report, demo video.  
**Day 14:** Final submission (GitHub, docs, executive summary).

**12. Summary**

SehatAI is a **practical, socially impactful, and feasible project** for the Uraan AI Techathon. It uses **public datasets, open-source tools, and lightweight AI models** to provide:

* Early disease detection (X-rays).
* Prescription digitization (OCR for Urdu/English).
* Predictive health monitoring (risk scoring).