

Title:

**Early Detection of Common Disease in Afghanistan using Machine Learning
and Health Data**

Group Members:

Hashmatullah Asady

Samiullah Gulzar

Homa Akrami

Baiqra Muradi

Fardaws Karimi

1. Project Idea:

We propose a system for **Early Detection of Common Diseases in Afghanistan using ML and Health data**. Our main goal is to review the health data and risk factors for malnutrition, diarrhea and respiratory infections as the most frequent illnesses among Afghan citizens. By improving patient outcomes and reducing pressure on the healthcare system, our system allows health providers to make earlier decision in patient care. Additionally, it will raise public awareness about these common diseases, encouraging the adoption of preventive measures and prioritization of health.

2. Relevance to Sustainable Development Goals (SDGs):

- **SDG 3: Good Health and Well-being**, reducing mortality and improving healthcare access in underserved communities.
- **SDG 1: No Poverty**, by lowering healthcare costs for families.
- **SDG 10: Reduced Inequalities**, by providing equitable healthcare solutions in remote areas.

3. Literature Examples:

- *"A Machine Learning Approach to Detect Dehydration in Afghan Children"* (Momand et al., 2023): This study leverages various classifiers to develop a predictive model using a dataset of sick children from the Afghanistan Demographic and Health Survey, aiming to determine the dehydration status of children under 5 years.
- *"Machine Learning Based Tuberculosis (ML-TB) Health Predictor Model: Early TB Health Disease Prediction with ML Models for Prevention in Developing Countries"* (S. Gupta, R. Kumar, and P. Singh): This paper presents a machine learning-based health predictor model for early tuberculosis prediction, emphasizing its application in developing countries.

4. Describe Your Data:

This project analyzes 10 years of health records with over 50,000 entries across Afghanistan, health records from public health surveys, global health organizations (e.g., **WHO**) and **OpenMRS**. The data will be in formats like CSV and JSON, containing features such as patient demographics, symptoms, diagnoses, outcomes, and environmental factors (e.g., pollution level). Tackling data issues with preprocessing, ensuring privacy via anonymization and encryption per regulations like GDPR.

5. Approach (Machine Learning or Deep Learning):

Machine learning algorithms such as Random Forest (for robustness and handling missing data), Logistic Regression (for binary classification with low computational costs), and deep learning models like LSTMs (for sequential data) will be employed to predict diseases. The goal is to reduce mortality rates and treatment costs. Practical implementation involves collaboration with Ministry of Public Health and NGOs and providing user-friendly dashboards for real-time use by healthcare workers.