# Predicting Hospital Readmission Risk in Diabetic Patients Using Machine Learning

## 1. Introduction

Hospital readmissions are a major concern in healthcare systems worldwide, particularly among diabetic patients who require continuous care and monitoring. Frequent readmissions not only increase healthcare costs but also indicate ineffective treatment or follow-up. With the growing availability of healthcare data, machine learning (ML) offers an efficient way to analyze complex patient information and predict which individuals are at high risk of readmission.

This project applies ML techniques to predict whether a diabetic patient will be readmitted within 30 days after discharge, helping hospitals take early preventive actions.

## 2. Objectives

- To analyze hospital data of diabetic patients and identify key features influencing readmission.
- To develop predictive models using Logistic Regression and Random Forest Classifier.
- To evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
- To assist healthcare providers in identifying high-risk patients and improving patient care quality.

## 3. Dataset Description

The dataset used is a Diabetic Hospital Readmission Dataset, containing over 100,000 patient records and more than 40 features.
Key attributes include:

- Demographics: age, gender
- Clinical details: number of diagnoses, A1C results, medications
- Hospital visit data: time in hospital, number of inpatient/outpatient visits
- Target variable: readmitted (Yes/No)

## 5. Methodology

**Data Preprocessing:**

- Handle missing values and remove irrelevant columns.
- Encode categorical variables (e.g., gender, discharge type) into numerical form.
- Normalize or scale continuous features.
- Split the dataset into training and testing sets (e.g., 80:20).

**Model Building:**

- Implement two supervised machine learning algorithms:
- Logistic Regression: for interpretable, probability-based classification.
- Random Forest Classifier: for robust, high-accuracy prediction.

**Model Evaluation:**

- Evaluate using accuracy, precision, recall, F1-score, and confusion matrix.
- Compare models to identify the best-performing algorithm.

## 6. Expected Outcome

- A machine learning model capable of predicting whether a diabetic patient is at risk of readmission.
- Identification of the most important factors influencing readmission.
- An interpretable and accurate predictive system that can assist hospitals in reducing readmission rates and improving patient outcomes