Interpretable and Fair Machine Learning for Diabetes Risk Prediction Using Clinical Data

(Data Science project)

Prepared by:

(CS4-G1)

Murad Beshr	23160018
Abdullah Al-Kabos	23160045
Riyadh Marish	23164062
Abdulsalam Muzafar	23164099
Majed Al-Naayib	21160085

Date:

November 20th, 2025

1. Abstract

Diabetes is being recognized as a health problem of the modern world. Early detection of risks is crucial for timely intervention. While machine learning models are generally quite accurate at making predictions, many act like "black boxes," which limits their usefulness in a clinical setting. In our work, we propose a simple and efficient ML pipeline based on publicly available clinical datasets for the prediction of diabetes risk. We use models like Logistic Regression and XGBoost and integrate explainable AI techniques, such as SHAP, to provide clear insights into the main risk factors. To further enhance the clinical value of our approach, fairness of the model is studied for various demographic subgroups, and calibration is assessed to ensure that the risk scores are reliable. Our results show that even simple ML models can compete not only on predictive accuracy but also on the delivery of transparent and fair insights. These models yield actionable results, suitable for healthcare decision-making purposes. Our study establishes a reproducible and interpretability-oriented framework. This framework fills in some of the most important gaps within the literature and opens the way towards wider clinical acceptance.

Key words: Diabetes Risk Prediction, Machine Learning, Explainable AI(XAI), Fairness/Bais Analysis, Healthcare Decision Support