Comparison of Traditional Machine Learning Models for Early Diabetes Screening in Resource-Constrained Settings

Chapter 1: Introduction

1.1 Background

(WHO)type II diabetes mellitus is a common chronic disorder which affects about 40 times the adults around the world in 2025, compared with those in 1990. According to the World Health Organization (WHO), the total cases of type II diabetes mellitus increased worldwide, from about 4% in 1990 to nearly 14% today. If left untreated or inadequately controlled, the risk of serious complications, like nephropathy, neuropathy and cardiovascular disease is very high. Therefore, an early and timely detection is important for patients who have a higher chance of developing severe complications.

(ML)parallel to the expanding scope of data-driven health care, there is the promising application of machine learning (ML) models for prediction of type 2 diabetes mellitus during early screening. While most of the existing studies have focused on complex or ensemble-based algorithms, very little work has been done to evaluate traditional ML models, such as Logistic Regression and Decision Trees, when applied in real-world circumstances where resource constraints play a major role.

1.2 Problem Statement

In many community clinics or rural areas，clinicians face severe constraints in terms of data availability，computational resources，and technical expertise.Thus，a key question arises:

Which traditional machine learning model is most appropriate for early diabetes screening in resource-constrained clinical settings?

This study aims to provide a rigorous comparative analysis of six commonly used traditional ML algorithms，not only in terms of predictive accuracy but also in robustness，efficiency，and clinical applicability.

1.3 Research Objectives

-To compare the performance of six traditional ML models on early diabetes detection using the Pima Indian dataset.

-To evaluate models under the lens of prediction，interpretability，robustness to missing data，and efficiency.

-To recommend suitable models for different levels of clinical infrastructure(e.g.，primary care vs.tertiary hospitals).

1.4 Significance of the Study

Unlike prior studies focused solely on accuracy or advanced architectures，this research provides a practical framework for model selection in low-resource healthcare environments.It bridges the gap between machine learning research and real-world clinical deployment by factoring in missing data，model interpretability，and computational cost.

1.5 Chapter Organization

This thesis is structured as follows:

-Chapter 2 reviews prior literature related to ML models in diabetes prediction.

-Chapter 3 outlines the dataset，preprocessing steps，and experimental methodology.

-Chapter 4 presents results from model evaluation and clinical validation.

-Chapter 5 concludes the study and proposes future research directions.

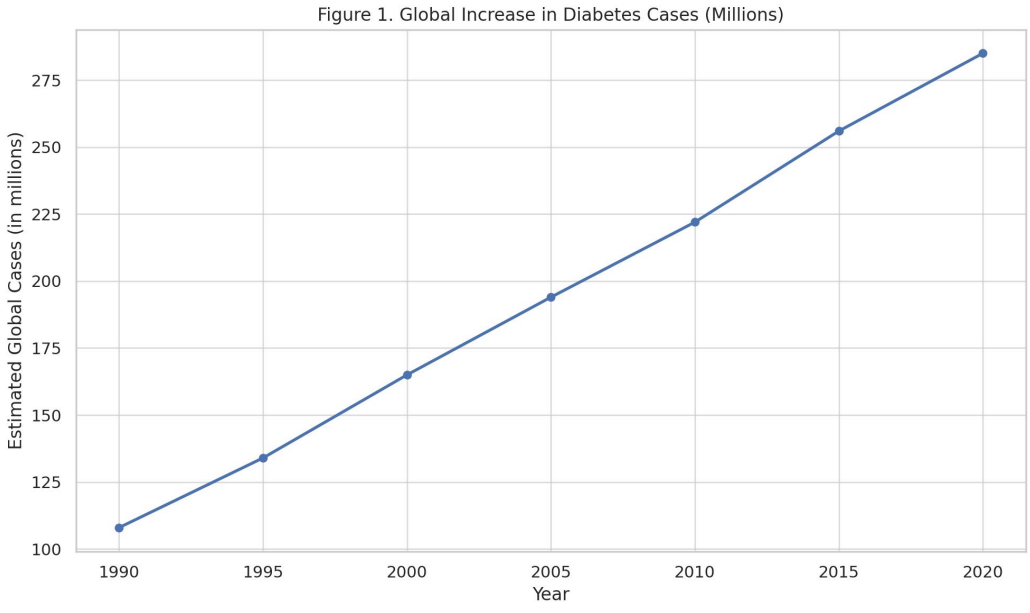

Figure 1. Global Increase in Diabetes Cases (Millions)