

# Research Proposal

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

Name: Zhang Yibo

Student ID: MCS241044

### 1. Introduction

Diabetes is a global health concern affecting over 500 million people worldwide. Early diagnosis is critical to prevent long-term complications, especially in low-resource environments. While many AI-based solutions exist, their complexity often limits clinical adoption. This study explores the performance of classical machine learning models for early diabetes detection using accessible clinical data.

### 2. Problem Statement

High-cost and high-complexity algorithms are impractical for widespread screening, particularly in under-resourced healthcare systems. There is a need for efficient and accurate traditional models that can deliver reliable predictions with minimal computational burden and clinical complexity.

### 3. Objectives

1. To compare the predictive performance of six traditional ML models on a benchmark diabetes dataset.
2. To analyze robustness to missing data and computational efficiency.
3. To recommend suitable models for different resource settings based on clinical applicability.

### 4. Methodology

Dataset: Pima Indians Diabetes dataset (UCI repository).

Preprocessing: KNN imputation, Z-score standardization, SMOTE.

Models: Logistic Regression, Decision Tree, KNN, Random Forest, Naive Bayes, SVM.

Evaluation: 5-fold cross-validation, AUC, Accuracy, Sensitivity.

Additional Analyses: ROC curves, SHAP feature importance, time & memory profiling.

## **5. Expected Outcome**

This study will identify which classical ML models offer a balance between predictive performance, simplicity, and practicality in different healthcare resource settings. It will also provide guidance on model deployment in real-world scenarios.

## **6. Timeline**

Week 1-2: Literature Review

Week 3-4: Dataset Collection & Preprocessing

Week 5-6: Model Training & Evaluation

Week 7: Result Interpretation (SHAP, ROC, etc.)

Week 8: Final Report & Presentation

## **7. Supervisor & Institution**

Supervisor: Dr. Shahizan Hassan

Institution: Faculty of Computer Science & Information Technology, Universiti Malaya