Examining Interpretability of Machine Learning-based Models for Diabetes Prediction using LIME Explainable AI Technique

Tayyeba Sadaq

100611584

Supervised by Dr. Oluwarotimi Samuel

Discipline of Computing and Mathematics

School of Computing and Engineering

University of Derby

Submitted May 2025, in partial fulfilment of the conditions for the award of the degree of **BSc Computer Science**

# Abstract

Diabetes is a chronic disease characterised by the pancreas producing insufficient insulin or the body’s inability to utilise produced insulin effectively. The imbalance, as a result, causes a disruption in the blood sugar’s regulation – potentially leading to severe complications affecting the nervous systems and blood vessels. The prevalence of pre-diabetes (hyperglycaemia) is rising, the latest National Diabetes Audit reporting 3,615,330 individuals registered with a GP as having pre-diabetes in 2023 – an 18% increase from 2022 [1].

This study seeks to bridge the gap between traditional diagnostic methods and modern artificial intelligence (AI) techniques by developing a web and mobile application that integrates Explainable AI (XAI) frameworks with machine learning (ML) models. Specifically, it evaluates the performance of three models – Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GBM)– in conjunction with the LIME XAI framework.

The aim is to identify the most effective ML-XAI combination for delivering accurate, interpretable, and actionable diabetes diagnosis predictions. By fostering transparency and trust in AI driven healthcare solutions, the findings contribute to improving the usability and readability of AI assisted diabetes diagnosis.

**Keywords:** Machine Learning, Explainable AI, Diabetes, Hyperglycaemia, logistic regression, random forest, gradient boosting, LIME

# Acknowledgements

Table of Contents

[Abstract 2](#_Toc192602075)

[Acknowledgements 3](#_Toc192602076)

[Chapter 1: Introduction 8](#_Toc192602077)

[1.1 Introduction 8](#_Toc192602078)

[1.2 Motivation 8](#_Toc192602079)

[1.3 Research Aims and Objectives 8](#_Toc192602080)

[Chapter 2: Background and Literature Review 9](#_Toc192602081)

[2.1 Background of Diabetes and Diagnosis 9](#_Toc192602082)

[2.1.1 Traditional Diagnosis Methods 9](#_Toc192602083)

[2.1.2 Machine Learning and Diabetes Diagnosis 9](#_Toc192602084)

[2.2 Literature Review 9](#_Toc192602085)

[Chapter 3: Research Methodology 10](#_Toc192602086)

[3.1 Research Method 10](#_Toc192602087)

[3.2 Data Collection 10](#_Toc192602088)

[3.3 Data Analysis 10](#_Toc192602089)

[3.4 Ethics 10](#_Toc192602090)

[3.5 Limitations 10](#_Toc192602091)

[3.6 Conclusions 10](#_Toc192602092)

[Chapter 4: Design and Implementation 11](#_Toc192602093)

[4.1 Logistic Regression 11](#_Toc192602094)

[4.1.1 How LR works and Implementation 11](#_Toc192602095)

[4.2 Random Forest 11](#_Toc192602096)

[4.2.1 How RF works and Implementation 11](#_Toc192602097)

[4.3 Gradient Boosting Model 11](#_Toc192602098)

[4.3.1 How GBM works and Implementation 11](#_Toc192602099)

[Chapter 5: Results and Analysis 12](#_Toc192602100)

[5.1 Datasets 12](#_Toc192602101)

[5.2 Evaluation Metrics 12](#_Toc192602102)

[5.3 Experimental Setup 12](#_Toc192602103)

[5.4 Results and Analysis 12](#_Toc192602104)

[5.5 Evaluation Comparison 12](#_Toc192602105)

[Chapter 6: Conclusion 13](#_Toc192602106)

[6.1 Conclusion 13](#_Toc192602107)

[6.2 Limitations 13](#_Toc192602108)

[6.3 Future Work 13](#_Toc192602109)

[References 14](#_Toc192602110)

List of Abbreviations

AI Artificial Intelligence

ML Machine Learning

XAI Explainable Artificial Intelligence

LR Logistic Regression

RF Random Forest

GBM Gradient Boosting Model

LIME Local Interpretable Model-Agnostic Explanations

# Chapter 1: Introduction

## 1.1 Introduction

The World Health Organisation describes diabetes as a chronic disease where the pancreas does not produce enough insulin, or the body is unable to effectively use any insulin that is being produced. As insulin regulates the body’s blood sugar levels, when not controlled, it can lead to serious damage to various bodily systems including the nervous and blood vessels [2]. For both Type 1 and Type 2 diabetes, the condition’s long-term implications means that the condition’s risks can often be significantly mitigated. While there is no real methods of prevention, there are ways in which to adjust and manage one’s lifestyle in order to prevent and avoid premature death.

Despite advances in medical science, diabetes remains undiagnosed and undertreated. According to the World Health Organisation, around 830 million people have diabetes – with over half not receiving treatment or diagnosis [2]. This highlights a critical gap in global healthcare systems, especially in low-resource settings where [access to healthcare is more limited and the impacts of missed diagnosis and limited access to healthcare is more fatal to those affected]

## 1.2 Motivation

## 1.3 Research Aims and Objectives

# Chapter 2: Background and Literature Review

## 2.1 Background of Diabetes and Diagnosis

### 2.1.1 Traditional Diagnosis Methods

### 2.1.2 Machine Learning and Diabetes Diagnosis

## 2.2 Literature Review

# Chapter 3: Research Methodology

## 3.1 Research Method

## 3.2 Data Collection

## 3.3 Data Analysis

## 3.4 Ethics

## 3.5 Limitations

## 3.6 Conclusions

# Chapter 4: Design and Implementation

## 4.1 Logistic Regression

### 4.1.1 How LR works and Implementation

## 4.2 Random Forest

### 4.2.1 How RF works and Implementation

## 4.3 Gradient Boosting Model

### 4.3.1 How GBM works and Implementation

# Chapter 5: Results and Analysis

## 5.1 Datasets

## 5.2 Evaluation Metrics

## 5.3 Experimental Setup

## 5.4 Results and Analysis

## 5.5 Evaluation Comparison

# Chapter 6: Conclusion

## 6.1 Conclusion

## 6.2 Limitations

## 6.3 Future Work

# References

|  |  |
| --- | --- |
| [1] | England NHS, “NHS identifies over half a million more people at risk of type 2 diabetes in a year,” 12 June 2024. [Online]. Available: https://www.england.nhs.uk/2024/06/nhs-identifies-over-half-a-million-more-people-at-risk-of-type-2-diabetes-in-a-year/. [Accessed 9 December 2024]. |
| [2] | W. H. Organisation, “Diabetes,” 14 November 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed 2024 December 2024]. |