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

In the Name of Allah, the most Gracious, the Most Merciful.

All praise is due to Allah ﷻ, who has granted me the strength, patience and knowledge to complete this work. Without his guidance and countless blessings, this journey would not have been possible.

I am deeply grateful to my supervisor, Oluwarotimi Samuel, for their invaluable guidance, patience, and encouragement. Their insights and expertise have played a crucial role in shaping this research and I sincerely appreciate their support. May Allah reward them abundantly.

I extend my deepest gratitude to my parents, who’s endless duas, love and sacrifices, have been my greatest source of strength. This achievement is as much yours as it is mine. May Allah bless you with health, barakah and Jannah.

To my friends, who encouraged me through challenges and reminded me to trust in Allah’s plan, I am truly grateful. Your support made this journey lighter.

Lastly, I pray this work, a tool for diabetes diagnosis and support, benefit those in need and serves as a means of ease and healing. May Allah accept it as an effort toward seeking knowledge and serving others.

Ameen.

**Table of Contents**

[Abstract 2](#_Toc194593010)

[Acknowledgements 3](#_Toc194593011)

[Chapter 1: Introduction 9](#_Toc194593012)

[1.1 Introduction 9](#_Toc194593013)

[1.2 Motivation 9](#_Toc194593014)

[1.3 Research Aims and Objectives 9](#_Toc194593015)

[Chapter 2: Background and Literature Review 11](#_Toc194593016)

[2.1 Background of Diabetes and Diagnosis 11](#_Toc194593017)

[2.1.1 Traditional Diagnosis Methods 11](#_Toc194593018)

[2.1.2 Machine Learning and Diabetes Diagnosis 12](#_Toc194593019)

[2.2 Literature Review 12](#_Toc194593020)

[2.2.1 Importance of Explainable AI in Healthcare 12](#_Toc194593021)

[2.2.2 Applications of XAI in Diabetes Prediction 12](#_Toc194593022)

[2.2.3 XAI4Diabetes: A mobile Application for Diabetes Risk Prediction 13](#_Toc194593023)

[Chapter 3: Research Methodology 14](#_Toc194593024)

[3.1 Research Method 14](#_Toc194593025)

[3.2 Data Collection 15](#_Toc194593026)

[3.3 Data Analysis 15](#_Toc194593027)

[3.4 Ethics 16](#_Toc194593028)

[3.5 Limitations 16](#_Toc194593029)

[3.6 Conclusions – [debating keeping this section] 17](#_Toc194593030)

[Chapter 4: Design and Implementation 18](#_Toc194593031)

[4.1 Logistic Regression 18](#_Toc194593032)

[4.1.1 How LR works 18](#_Toc194593033)

[4.1.2 Implementation 18](#_Toc194593034)

[4.2 Random Forest 18](#_Toc194593035)

[4.2.1 How RF works 18](#_Toc194593036)

[4.2.2 Implementation 18](#_Toc194593037)

[4.3 Gradient Boosting Model 18](#_Toc194593038)

[4.3.1 How GBM works 19](#_Toc194593039)

[4.3.2 Implementation 19](#_Toc194593040)

[Chapter 5: Results and Analysis 20](#_Toc194593041)

[5.1 Datasets 20](#_Toc194593042)

[5.2 Evaluation Metrics 20](#_Toc194593043)

[5.3 Experimental Setup 20](#_Toc194593044)

[5.4 Results and Analysis 20](#_Toc194593045)

[5.5 Evaluation Comparison 20](#_Toc194593046)

[Chapter 6: Conclusion 21](#_Toc194593047)

[6.1 Conclusion 21](#_Toc194593048)

[6.2 Limitations 21](#_Toc194593049)

[6.3 Future Work 21](#_Toc194593050)

[References 22](#_Toc194593051)

**Table of Figures**

[Figure 1 - Global death rate from diabetes mellitus per 100,000 population (1980-2021) demonstrating a consistent upward trend in mortality over 4 decades [3] 11](#_Toc194581264)

**List of Tables**

[Table 1: Application of the CRISP-DM Framework in This Study 15](#_Toc194583206)

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

FPG Fasting plasma glucose

WHO World Health Organisation

ADA American Diabetes Association

EDA Exploratory Data Analysis

# 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

Diabetes alone is a prevalent condition most people face, however only a small percentage of the population is often aware of their underlying condition and fewer still manage to live a less detrimental lifestyle as a result. This lack of knowledge and understanding of your own health conditions is prevalent in less developed countries and can often be a leading cause of death [back up with figures]. A condition that can often be managed through small shifts and changes in one’s lifestyle, diet and environment is something that ought to have accessible diagnosis available to patients. Creating an application that allows for people input their medical data safely and securely to provide several predictions in which to help them understand further steps is a useful and somewhat mandatory tool to help bridge the current gap in global healthcare systems. Especially those which are less developed and thereby offer less support and treatment.

## 1.3 Research Aims and Objectives

This study aims to bridge the gap between traditional diagnostic approaches and modern AI techniques by implementing a web and mobile application that leverages XAI frameworks alongside ML modes. By doing so, it seeks to improve the accessibility, interpretability and effectiveness of diabetes diagnosis to contribute to better health outcomes and a more proactive approach to[something]. fostering awareness to the diabetes and teaching ways in which people can proactively manage their health. Furthermore, addressing the strengths and limitations of XAI frameworks, the study contributes towards the development of more transparent, patient-centred and accessible healthcare.

# Chapter 2: Background and Literature Review

## 2.1 Background of Diabetes and Diagnosis

The rapid increase in the prevalence of diabetes has made early detection, diagnosis and management crucial to benefit the lives of individuals affected by the condition. As a leading cause of morbidity and mortality (Figure 1) due to the association with serious complications. The ability to provide early, accessible diagnosis is a further step taken to improve millions of lives worldwide.

A graph showing the number of patients with diabetes

Description automatically generated

Figure - Global death rate from diabetes mellitus per 100,000 population (1980-2021) demonstrating a consistent upward trend in mortality over 4 decades [3]

### 2.1.1 Traditional Diagnosis Methods

Traditionally, diabetes diagnosis involves clinical testing, requiring blood to be drawn by doctors, from patients, and tested in a medical lab. Fasting plasma glucose (FPG) is a preferred method by doctors, given it’s easy, convenient and minimal costs compared to other tests [4].Working by measuring the levels of glucose in the blood after fasting for 8 hours, WHO defines that 7.0 mmol/L or less is normal and a level of 7 mmol/L or higher on two separate FPG tests as indicative of diabetes [2].

While effective in the diagnosis of diabetes, the traditional methods are not without their limitations. With a lack of consensus on the consistent standards and criteria for diagnosis - for example WHO and ADA offering different glucose thresholds, the need to fast, reliance on laboratory infrastructure, the possibility of delayed results and false diagnosis. There are clear areas for improvement to help provide more accessible diagnosis for patients and allow professionals to aid people more effectively. Recent advancements in technology and data science have opened new pathways for diabetes diagnosis and prediction.

### 2.1.2 Machine Learning and Diabetes Diagnosis

ML techniques enable the analysis of larger datasets to identify patterns and correlations that may not be immediately apparent. Predictive models can assess an individual’s risk based on factors such as age, weight, blood pressure and family history/genetics. This allows for an earlier intervention to be staged and therefore better management of the condition to avoid possible reliance on heavy medication to handle their condition.

Explainable AI is a further advancement on predictive models. Given ML models are “black box” in nature, their ambiguity makes it challenging to gain insight into the internal mechanisms of the model, providing a lack of full trust in a system that would be responsible for critical and sensitive decision making [5]. XAI techniques and frameworks allow for an interpretability component to be integrated with ML model predictions, meaning end-users can comprehend and interpret outputs and predictions made by AI models [6].Transparency is the baseline needed to foster trust in AI-driven diagnosis and support for a more informed decision making. [add something on making AI more mainstream for healthcare]

## 2.2 Literature Review

The prediction of diabetes has been extensively studied. While traditional medical approaches remaining as the cornerstone for clinical practices, advancements in ML and AI have introduced more innovative methods. Promising improved accuracy, scalability and the promise of earlier detection capabilities. Not only improvements to what the clinical tests provide but a further enhancement and deeper understanding that can change how diabetes is managed and approached – even in pre-diabetic stages.

### 2.2.1 Importance of Explainable AI in Healthcare

There is a constant growing prevalence of diabetes worldwide that necessitates early and accurate diagnosis to [help people live better]. Though statistical methods are considered effective in identifying a patient’s [correlation to diabetes], it struggles to handle the complexity and volume of modern healthcare data. Comparing one patient’s data with the plethora available to diagnose diabetes requires time and financial support to

### 2.2.2 Applications of XAI in Diabetes Prediction

### 2.2.3 XAI4Diabetes: A mobile Application for Diabetes Risk Prediction

# Chapter 3: Research Methodology

## 3.1 Research Method

This study adopts a quantitative research approach, leveraging machine learning techniques to classify diabetes based on clinical data. A supervised learning method is employed, where labelled data is used to train the models into distinguishing between diabetic and non-diabetic individuals. Three models were chosen overall each with their own benefits and limitations as well as reasons for being selected for this research [mention the models chosen and a brief explanation as to why]. The models selected were as follows:

1. Logistic Regression (LR)
2. Random Forest (RF)
3. Gradient Boosting (GBM)

Logistic regression was chosen for its simplicity and interpretability. Useful in regard to understanding the linear relationship between features and diabetes risk. Meanwhile Random Forest can handle non-linearity, providing feature importance which improves the model’s robustness and explainability. Finally, Gradient Boosting was included for its high predictive accuracy and ability to capture complex patterns in data which helps determine a final diagnosis based on user inputs.

To ensure the mode’s interpretability and transparency in offered diagnosis, the LIME XAI was used along with feature importance analysis and [also textual explanation to help understand how the diagnosis came about] are integrated. These features help explain how different features contribute to the model’s decision-making process which improves the trust in AI-driven diagnosis. The use of multiple explainability functions also caters for both personal and professional use, making it so general users can clearly understand the decision-making process of the models when it gives a diagnosis. The research, and project itself follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework (Table 1)

|  |  |
| --- | --- |
| Process Step | Application to Research |
| Business Understanding | The objective is to classify diabetes using ML and enhance interpretability through XAI. The key risk factors (glucose levels, BMI, age, etc) are analysed. |
| Data Understanding | Exploratory Data Analysis (EDA) is performed to check feature distributions, correlations and missing values. |
| Data Preparation | Missing values are handled feature are standardised and necessary transformations are applied. |
| Modelling | Models are trained and hyperparameter tuning is performed to optimise their performance on user inputted data. |
| Evaluation | Performance is measured using accuracy, F1-score, ROC-AUC and explainability techniques. |
| Deployment | The model is incorporated into a “Diabetes Sense” app for user friendly access to predictions and explanations. (available for web and mobile) |

Table 1: Application of the CRISP-DM Framework in This Study

## 3.2 Data Collection

Data collection is a critical step in building reliable ML models – as the quality and relevance of the data has a direct influence on the accuracy and interpretability of predictions being made. For this study, the Pima Indians Diabates Dataset was taken from Kaggle. This is a publicly available dataset whose objective is to diagnostically predict whether a patient has diabetes, meaning it aligns strongly with the purpose and needs of this study. Originally from the National Institute of Diabetes and Digestive Kidney Diseases, the dataset explores the medical history and status of an indigenous community living in the southwest region of the United States [7]. The origin of the data and the people which it involves is an important factor to consider and can be considered a limitation of this study which will be explored later. It’s a widely used benchmark dataset for diabetes research as it focusses on populations with high prevalence of diabetes, highlighting critical risk factors and concise feature sets that provide an ideal foundation in understanding the relationship between clinical metrics and diabetes risk.

## 3.3 Data Analysis

The dataset undergoes preprocessing to enhance the performance of the selected models. This includes handling missing values, outliers and scaling features. [talk about preprocessing techniques used, why, screenshots/graphs that showcase it working etc].

Exploratory Data Analysis (EDA) was then conducted to identify patterns and correlations in the dataset between features and to the target variable. [methods used]. Alongside this, visualisations such as histograms, correlation matrices and boxplots were used to examine feature distributions and help detect outliers in features. [graphs and what they show]

## 3.4 Ethics

Ethical considerations are crucial in AI-driven healthcare research, particularly with public datasets. This study uses the Pima Indian’s Diabetes Dataset, which originates from the National Institute of Diabetes and Digestive and Kidney Diseases. While publicly available on Kaggle, ethical responsibility remains, needing to ensure that the data is used appropriately and doesn’t contribute to bias or misrepresentation. There are 4 key ethical considerations to take into account: data privacy and anonymity, bias and fairness, transparency and explainability, and responsible AI use.

These considerations were met accordingly, some through the dataset itself and others through how the project was managed and handled. The dataset doesn’t include personally identifiable information, ensuring the patients retain their privacy and remain anonymous. The dataset is focused on an indigenous population, meaning that the model may not generalise well to other demographic groups. [how was this handled]. To promote ethical AI usage, this study has a strong focus on the use of the LIME XAI technique alongside feature importance analysis and a text-based explanation of diagnosis for both users and professionals to help understand the model’s decision-making process and foster trust between user and AI. In addition to this, the ‘Diabetes Sense’ app provides predictions as a tool to assist users but cannot replace professional medical advice. As such, clear disclaimers are included throughout the app/website to emphasise this and inform users that to act on/take any diagnosis further, they are to consult a medical professional. [Furthermore, that any information provided on the website is a collection of publicly available knowledge and general advice]

## 3.5 Limitations

Prior to implementing this study, several limitations were identified that could impact the study’s effectiveness and applicability. The first major constraint, mentioned earlier, is the dataset itself. While widely used for diabetes research it holds a bias as to representing a specific population that’s susceptible to higher diabetes rates which can limit its generalisability to other demographic groups. While efforts were made to mitigate this [mention techniques like feature selection etc that was used], the inherent limitations of population specific data remain.

Furthermore, the dataset doesn’t include certain important diabetes risk factors such as diet, physical activity and environmental factors, which could enhance the predictive accuracy however were unavailable in the dataset. However, the app is designed with a help and advice section to offer users insights into how these can contribute to diabetes and how to balance and manage your lifestyle to avoid being as severely impacted if diagnosed positively.

Finally, machine learning models inherently caries the risk of bias and overfitting, particularly with smaller datasets. Measures were taken to optimise the performance through hyperparameter tuning and validation techniques, but some performance variation is always to be expected given the model is being applied to varying real-world data. [add some figures here to further explain]

## 3.6 Conclusions – [debating keeping this section]

This chapter outlined the research methodology employed in the study, detailing the data collection, preprocessing, model selection and evaluation process. A quantitative approach was used, leveraging ML techniques to classify diabetes while integrating XAI to enhance the model’s transparency and user trust. A CRISP-DM framework was followed to ensure a structured approach from data understanding to deployment.

Key ethical considerations, like bias, fairness, and responsible AI usage, were acknowledged and addressed as much as possible to ensure the model remains interpretable and provides reliable insights without leading people to believe it can replace professional medical advice. Identified limitations, such as the dataset’s representativeness and missing lifestyle-related factors were acknowledged, shaping how the model itself and the app was designed and evaluated.

Overall, this methodology establishes a strong foundation for developing an AI driven diabetes prediction tool to balance accuracy and interpretability. The next chapter will focus on the design and implementation of the system, detailing how the models and explainability techniques were integrated into the ‘Diabetes Sense’ application.

# Chapter 4: Design and Implementation

## 4.1 Logistic Regression

### 4.1.1 How LR works

Logistic Regression is a widely used statistical methods for binary classification tasks. It models the relationship between input features and the possible binary outcome, using the sigmoid function (Figure 2). This maps the linear combination of input features to a value between 0 and 1, which represents the probability of belonging to a specific class. A threshold, typically of 0.5, is applied in order to determine the final class label. LR is particularly useful for problems in which relationships between features and target variables are approximately linear.

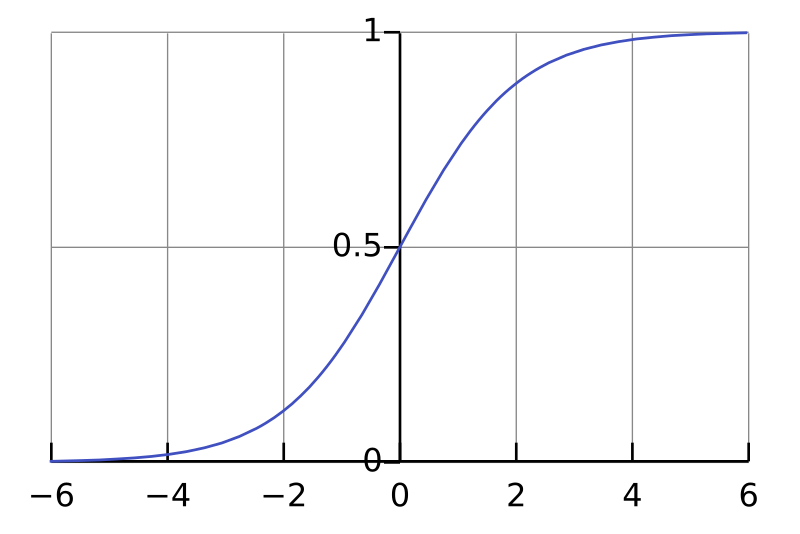


Figure 2: the logistic curve [8]

### 4.1.2 Implementation

In this project, LR was implemented with the ‘LogisticRegression’ class from the ‘scikit-learn’ library. The dataset was pre-processed by standardising the features using ‘StandardScaler’ to ensure all the variables were on the same scale.

## 4.2 Random Forest

### 4.2.1 How RF works

### 4.2.2 Implementation

## 4.3 Gradient Boosting Model

### 4.3.1 How GBM works

### 4.3.2 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]. |
| [3] | Network, Global Burden of Disease Collaborative, “Global Burden of Disease Study 2021 (GBD 2021) Results,” Institute for Health Metrics and Evaluation (IHME), 2022. [Online]. Available: https://vizhub.healthdata.org/gbd-results. [Accessed 8 January 2025]. |
| [4] | E. Mahoney, Diabetes: diagnosis and management, New York: Lucent Press, 2018. |
| [5] | S. A. Tanim, A. R. Aurnob, T. E. Shrestha, F. Mridha, R. I. Emon and S. U. Miah, “Explainable deep learning for diabetes diagnosis with DeepNetX2,” *Biomedical Signal Processing and Control,* vol. 99, p. 106902, 2025. |
| [6] | Z. Sadeghi, R. Alizadehsani, M. Akif, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, A. Rami, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hladík and Nahavandi, “A review of Explainable Artificial Intelligence in healthcare,” *Computers and Electrical Engineering,* vol. 118, p. 109370, 2024. |
| [7] | J. Radin, ““Digital Natives”: How Medical and Indigenous Histories Matter for Big Data,” *Data Histories,* vol. 32, no. 1, pp. 43-64, 2017. |
| [8] | Q. (talk), “Sigmoid function,” 2 07 2008. [Online]. Available: https://commons.wikimedia.org/w/index.php?curid=4310325. |