Diabetes Prediction using Machine Learning and Explainable AI Techniques

University of Derby

Course Subject: Independent Studies

Course Code: 6CC995

Student number: 100611584

Supervisor: Oluwarotimi Samuel

Academic Year: 2024-2025

Table of Contents

[Abstract 3](#_Toc187747284)

[Keywords 3](#_Toc187747285)

[Introduction 3](#_Toc187747286)

[Background of Diabetes prediction and Diagnosis 4](#_Toc187747287)

[Literature Review 5](#_Toc187747288)

[Importance of Explainable AI in healthcare 5](#_Toc187747289)

[Applications of XAI in Diabetes Prediction 6](#_Toc187747290)

[Broader Applications of XAI in Medical Diagnosis 8](#_Toc187747291)

[Challenges in XAI adoption 8](#_Toc187747292)

[Future Directions 8](#_Toc187747293)

[Methodology and Design 8](#_Toc187747294)

[Research design and approach 8](#_Toc187747295)

[Data Collection 8](#_Toc187747296)

[System and framework design 10](#_Toc187747297)

[Development process 10](#_Toc187747298)

[Verification and validation 10](#_Toc187747299)

[Implementation 10](#_Toc187747300)

[Tools and technologies 10](#_Toc187747301)

[Development steps 10](#_Toc187747302)

[Challenges 10](#_Toc187747303)

[Results and Analysis 11](#_Toc187747304)

[Findings 11](#_Toc187747305)

[Data analysis 11](#_Toc187747306)

[Discussion of results 11](#_Toc187747307)

[Evaluation and Reflection 11](#_Toc187747308)

[Assessing project outcomes 11](#_Toc187747309)

[Reflecting on methodology and design 11](#_Toc187747310)

[Future work 11](#_Toc187747311)

[Conclusion 11](#_Toc187747312)

[References 11](#_Toc187747313)

Abstract

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 being produced. Insulin is a regulator of the body’s blood sugar levels, which when not controlled, can lead to serious damage to various bodily systems, especially the nervous system and blood vessels. [1] The latest National Diabetes Audit shows that 3,615,330 people registered with a GP were found to have pre-diabetes (hyperglycaemia) in 2023 -an increase of 18% since 2022. [2]. This research aims to bridge the gap between traditional diagnostic methods and modern AI techniques by implementing a web and mobile application integrating Explainable AI (XAI) frameworks with machine learning (ML) models.

Specifically, this study evaluates the performance of three ML models—[TO BE DECIDED]—paired with two XAI frameworks (LIME and SHAP). The system also features a toggle option, allowing users to switch between XAI frameworks in real-time, fostering flexibility and accessibility. The findings aim to identify the most effective ML-XAI combination for delivering accurate, interpretable, and actionable diabetes predictions, contributing to enhancing trust and usability in AI-driven healthcare solutions.

Keywords

Machine Learning, Explainable AI, Diabetes, Hyperglycaemia, Disease prediction model,

Introduction

Diabetes is a chronic illness characterised by consistent hyperglycaemia. It arises from the pancreas’ inability to produce sufficient insulin or the body’s ineffectiveness in utilising the produced insulin. In both cases, the condition’s long-term implications underline a critical need for timely diagnosis. Timely intervention because of early diagnosis means that the condition’s risks can often be significantly mitigated, and while there’s no real prevention to the disease, there are ways to adjust and manage one’s lifestyle in order to prevent complications and avoid a premature death.

Despite the 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 a diagnosis. [1] This highlights a critical gap in the healthcare systems globally, especially in low-resource settings

Traditional diagnostic methods often rely on clinical tests such as fasting blood sugar levels, glucose tolerance tests and glycated haemoglobin (HbA1c) measurements in order to determine the presence and severity of diabetes. While effective, these are resource intensive methods and may not be as accessible to low-resource populations. Recent years have seen a development in Machine Learning (ML) and Artificial Intelligence (AI) as promising tools to enhance diabetes diagnosis and predictions, surpassing the traditional methods in both efficiency and scalability.

However, the complexity of AI models means that there is a lack of transparency when it comes to offering predictive diagnosis, making it difficult for healthcare professionals and patients to trust and understand the outcomes of the model’s diagnosis. Explainable AI (XAI) frameworks such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) address this issue by providing an insight into model predictions. This means that we can foster trust and enable more actionable decision making for patients and healthcare professionals alike. By integrating these frameworks into a diabetes diagnostic tool, patients feel more empowered with a better understanding of their symptoms, and professionals are more encouraged to adopt AI solutions to aid their work.

This study specifically aims to bridge the gap between traditional diagnostic approaches and modern AI techniques by implementing a web and mobile application the leverages XAI framework alongside ML models. 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 help foster awareness of the condition and ways in which people can proactively manage their health. By addressing the strengths and limitations of XAI frameworks, it contributes to the development of a more transparent, patient-centered and accessible healthcare system.

Background of Diabetes prediction and Diagnosis

The rapid increase in the prevalence of diabetes has made early detection, diagnosis and management crucial in benefiting 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, being able to provide early accessible diagnosis is a step forward to improve millions of lives worldwide.

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 is easy, convenient and less expensive than other tests. [3] It works by measuring the levels of glucose in your blood after having fasted (no food or drink other than water) for at least 8 hours. The World Health Organisation defines 7.0 mmol/L or less to be normal and a level of 7 mmol/L or higher on two separate tests indicates diabetes. [1]

While traditional methods remain effective in the diagnosis of diabetes, they’re not without limitations. With a lack of consensus on the consistent standards and criteria for diagnoses, for example WHO and ADA offering different glucose thresholds, ADA having a lower threshold of 39-46 mmol/mol for diagnosing prediabetes while WHO sets a higher threshold, often starting at 42 mmol/mol. [4] The need to fast, reliance on laboratory infrastructure and potential delay in results, not to mention the possibilities of false diagnosis means there is clear areas for improvement to help provide more accessible diagnosis for patients and allow doctors to help people more effectively. Recent advancements in technology and data science have opened new pathways for diabetes diagnosis and prediction. Machine Learning (ML) techniques for instance, enabling the analysis of large datasets to identify patterns and correlations not immediately apparent. Predictive models can assess an individual’s risk based on factors such as age, weight, blood pressure and family history/genetics, thereby enabling an earlier intervention and handling.

Explainable AI (XAI) is a further advancement of this, given ML’s more “black box” nature in regards to traditional algorithms. This ambiguity of machine learning based diagnosis makes it challenging to gain insight into its internal mechanisms and means there’s a lack of full trust in system that would be responsible for critical and sensitive decision making. [5] XAI techniques provide an interpretability component that enable end-users to comprehend and interpret the outputs and predictions made by AI models. [6] Such transparency fosters trust in AI-driven diagnosis and support for a more informed decision making.

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 [7]

Literature Review

The prediction and diagnosis of diabetes has been extensively studied, with traditional medical approaches remaining as the cornerstone for clinical practices. However, advancements in machine learning (ML) and artificial intelligence (AI) have introduced more innovative methods that promise improved accuracy, scalability and early detection capabilities.

Importance of Explainable AI in healthcare

The growing prevalence of diabetes worldwide necessitates early and accurate diagnosis. Though traditional statistical methods are effective, they struggle with the complexity and volume of modern healthcare data. Machine learning models such as decision trees, random forests and neural networks have demonstrated superior predictive performance in handling such data as however their lack of interpretability hinders their clinical integration. Explainable AI is a means to bridge this gap and address the limitations by providing a transparent, user-friendly insight into how predictions are made. Using techniques like Local Interpretable Model Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) enhances a model’s interpretability and fosters trust among both patients and professionals when it comes to assisting and administering a diagnosis for such a prevalent, life-changing disease.

Applications of XAI in Diabetes Prediction

Research has extensively explored the integration of Explainable AI techniques into diabetes prediction systems to bridge the gap between technical sophistication and practical usability. Traditional ML models often excel in the accuracy of their predictions but fall flat when providing insights into the decision-making process, thereby limiting their application and adoption into high stake fields such as healthcare. XAI addresses this by offering interpretable insights that enhance trust and make the system more accessible and valuable for healthcare professionals and patients alike.

XAI4Diabetes: A Mobile Application for Diabetes Risk Prediction

Hendawi et al introduced XAI4Diabetes, a mobile application designed to predict risk while offering interpretable and transparent explanations for predictions. By utilising knowledge base, knowledge matching, prediction and interpretation, XAI4Diabetes forecasts diabetes risk and provides valuable insights into the prediction process and outcomes. [8]

The knowledge base module organises diabetes related data through the use of ontologies and knowledge graphs which provides a structured and extensible representation of complex medical in information. The prediction model utilises machine learning models, specifically deep neural network (DNN), random forest (RF) and decision tree (DT) to generate risk predictions based on the data of patients. The interpretation module employs XAI tools like LIME and SHAP to deliver explanations at local and global levels to enable healthcare professionals to better understand the underlying reasons for each prediction. Their evaluation module assesses the app’s usability and effectiveness through user studies which ensured that the system met the practical needs of medical professionals, a key component to include in order to help foster trust from the professionals with AI tools to improve and support their work.

Structured, survey-based user studies were also conducted as a part of XAI4Diabetes’ evaluation to assess the app’s usability and influence on participant comprehension of machine learning predictions in a real-world patient scenario [8]. These demonstrated a significant improvement in the trust and understanding of healthcare providers in AI driven prediction. The application highlighted features such as BMI, blood glucose levels and polyuria as critical predictors and offered clear and actionable interpretations of risk factors considered when offering a diagnosis. There was a utilisation of user-friendly interfaces (Figure 2 and 3) and the leverage of visuals to simplify complex data. However, despite the strengths of their research, the small sample size of participants limited the generalisability of findings, additionally the technical barriers made explaining the intricacies of the ML model, meaning it required specialised knowledge for interpretation [8]. The app’s design itself lacked features tailored to patient specific needs such as language accessibility and cultural relevant context which is an important consideration when making the application more accessible globally.

Looking forward, the researchers proposed several enhancements to the application. This includes integration of animation to the UI and adding multilingual support in order to conduct larger scale studies with more diverse user groups. Such improvements would aim to address the current limitations while further establishing the app as a robust and user-centric tool for diabetes risk prediction and management. With the broader implication to make explainable and transparent AI in various industries [8].

Screens screenshot of a phone

Description automatically generated

Figure - Screenshots of a patient’s basic information (left) and examination information (right). [8]

A screenshot of a cell phone

Description automatically generated

Figure - Screenshots of prediction result (left) and explanation based on patients’ symptoms (right). [8]

Systematic Approaches in XAI Integration

Broader Applications of XAI in Medical Diagnosis

Challenges in XAI adoption

Future Directions

Methodology and Design

Research design and approach

Data Collection

<https://www.kaggle.com/datasets/andrewmvd/early-diabetes-classification>

<https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>

<https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset/data>

<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database/data>

Data collection is a critical step in building a reliable Machine Learning model as the quality and relevance of the data directly influences the accuracy and interpretability of the predictions being made. For this project, four publicly available datasets were selected from Kaggle, each containing medical and demographic information related to diabetes diagnosis. They vary in size, features and data sources, which provides a comprehensive foundation for developing and testing ML models.

The datasets chosen include medical metrics such as glucose levels, insulin levels, Body Mass Index (BMI) and blood pressure. All which are essential in identifying patterns associated with diabetes. By combining the insights from multiple datasets, we aim to enhance the model’s performance and ensure that robust predictions are being provided.

Dataset Overviews

Pima Indians Diabetes Dataset – UCI ML Repository

This is a widely used benchmark dataset for diabetes research. Focussing on populations with a high prevalence of diabetes, it highlights critical risk factors and a concise feature set that provides an ideal foundation in understanding the relationship between clinical metrics and diabetes risk. The simplicity and historical use of the dataset in ML research makes it valuable as a baseline for evaluating the performance and interpretability of XAI-integrated models, ensuring that the study’s approach aligns with established diabetes research methodologies.

Diabetes Prediction Dataset by Mustafa TZ

This large-scale dataset emphasises predictive diabetes diagnosis, having records with a rich mix of demographic, clinical and behavioural data. This scale and diversity allow for a robust model training and validation that addresses the study’s goal of building a scalable and efficient AI model. The inclusion of traditional and diagnostic metrics as well as lifestyle indicators strengthens both it’s relevance in development and adaptability to resource settings. Its large sample size enhances model accuracy and generalisability and makes it ideal for building scalable diagnostic solutions for underrepresented populations.

Early Diabetes Classification Dataset by Andrew MVD

This dataset provides a foundation for identifying early-stage diabetes risks, focusing on a combination of clinical indicators and lifestyle variables. Early detection aligns with the study's aim of mitigating diabetes-related complications by enabling proactive intervention. The inclusion of features like glucose levels, BMI, and physical activity provides insight into critical factors contributing to hyperglycaemia and supports the development of AI models capable of delivering timely predictions. The dataset’s focus on early indicators is well-suited for researching predictive diagnosis, offering data to train models for detecting diabetes at an early, more manageable stage.

Diabetes Health Indicators Dataset by Alex Teboul

This dataset uses health survey data to focus on predictors of diabetes, capturing essential health and behavioural indicators that are accessible in low-resource environments. Its categorical and numerical features align with the study’s aim to explore scalable and accessible diagnostic methods and can be used to develop models that provide predictions using readily available data, enhancing the applicability of the study's web and mobile application in resource-constrained settings. By bridging the gap between clinical and lifestyle factors, this dataset is highly relevant in focussing on accessible and interpretable AI solutions for diabetes management.

Data Selection Criteria

The datasets for this study were selected for their relevance, diversity, and accessibility. Each dataset includes critical attributes for diabetes diagnosis, such as glucose levels, BMI, insulin levels, and blood pressure, aligning with the study’s goal of accurate and explainable predictions. Diversity was prioritized by incorporating demographic, clinical, and lifestyle factors to develop adaptable models for various populations and scenarios. Publicly available datasets ensure transparency and reproducibility, while their size and feature distribution provide a robust foundation for training, validation, and generalizability. The inclusion of the Pima Indians dataset, a benchmark in diabetes research, aligns with established methodologies and advances research in XAI frameworks. By leveraging this dataset alongside others, the study evaluates the effectiveness of XAI methods in delivering transparent ML model explanations, fostering trust, and improving decision-making for healthcare professionals and patients.

Limitations of the Data

While the datasets provide valuable insights, they’re not without their limitations. One of the key challenges in sourcing is the class imbalance, with some datasets exhibiting a disproportionate radio of positive and negative diabetic cases. This imbalance could bias the model’s performance and requires careful handling in training. Missing and incomplete data, specifically for insulin levels and physical activity is also something which will be handled using data imputation in order to maintain data integrity.

Furthermore, the datasets primarily consist of cross-sectional data, meaning it lacks the temporal direction needed to capture the trends of diabetes over time. Additionally, the population bias of the datasets is another limitation. The Pima Indians diabetes set focusses on a specific ethnic group which limits its generalisability to other populations. The self-reported nature of lifestyle and behavioural data in some of the datasets also introduces the risk of inaccuracies which can potentially affect the reliability of the model.

Finally, while the datasets do aim to align with low-resource settings, as a result they can lack certain clinical measurements that are more commonly available in high-resource environments. Yet despite these limitations, they all offer a robust starting point for developing an interpretable AI model that can address the critical need for accessible and accurate diabetes diagnosis.

System and framework design

Development process

Verification and validation

Implementation

Tools and technologies

Development steps

Challenges

Results and Analysis

Findings

Data analysis

Discussion of results

Evaluation and Reflection

Assessing project outcomes

Reflecting on methodology and design

Future work

Conclusion

References

|  |  |
| --- | --- |
| [1] | W. H. Organisation, “Diabetes,” 14 November 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed 9 December 2024]. |
| [2] | N. England, “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]. |
| [3] | E. Mahoney, Diabetes: diagnosis and management, New York: Lucent Press, 2018. |
| [4] | J. Zhang, Z. Zhang, K. Zhang, X. Ge, R. Sun and X. Zhai, “Early detection of type 2 diabetes risk: limitations of current diagnostic criteria,” vol. 14, 2023. |
| [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, S. Nahavandi and P. Pardalos, “A review of Explainable Artificial Intelligence in healthcare,” *Computers and Electrical Engineering,* vol. 118, p. 109370, 2024. |
| [7] | G. B. o. D. C. Network, “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]. |
| [8] | R. Hendawi, J. Li and S. Roy, “A Mobile App That Addresses Interpretability Challenges in Machine Learning–Based Diabetes Predictions: Survey-Based User Study,” *JMIR Form Res,* vol. 7, p. e50328, 2023. |
| [9] | G. o. t. U. Kingdom, “Health profile for England 2018: Chapter 3: Trends in morbidity and risk factors.,” 11 September 2018. [Online]. Available: https://www.gov.uk/government/publications/health-profile-for-england-2018/chapter-3-trends-in-morbidity-and-risk-factors. [Accessed 8 January 2025]. |