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

Student: Tayyeba Sadaq – 100611584 Supervisor: Dr. Oluwarotimi Samuel

Abstract

Diabetes is a chronic disease that affects insulin production and utilization, leading to severe health risks [1]. In 2023, the National Diabetes Audit reported 3.6 million pre-diabetes cases in the UK, an 18% increase from 2022 [2]. This research aims to enhance trust in AI-driven healthcare by integrating Explainable AI (XAI) with machine learning (ML) for diabetes prediction. It evaluates different ML models—Random Forest, Logistic Regression, and Gradient Boosting—using LIME to determine which model provides the best predictive performance and interpretability.

Methodology

Random forest, logistic regression and gradient boosting models were trained using the Pima Indian Diabetes Database. LIME was then used to explain the predicted diagnosis by showcasing the feature importance (Figure 1). Helping the user understand which features specifically contributed the most to the diagnosis they received. Models were also evaluated, and final models are deployed on a full stack application (Figure 2) to provide accessibility to patients and professionals.

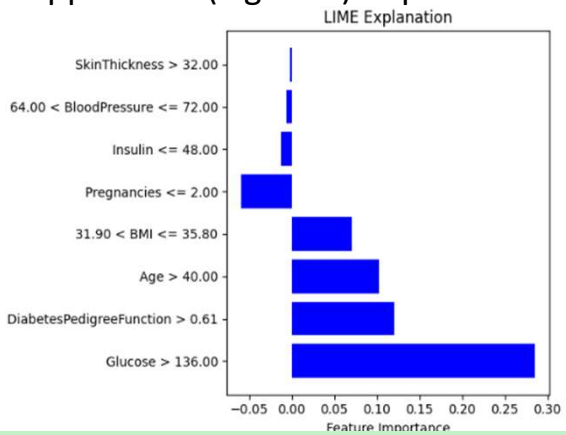

Figure 1: Lime Explanation for sample diagnosis

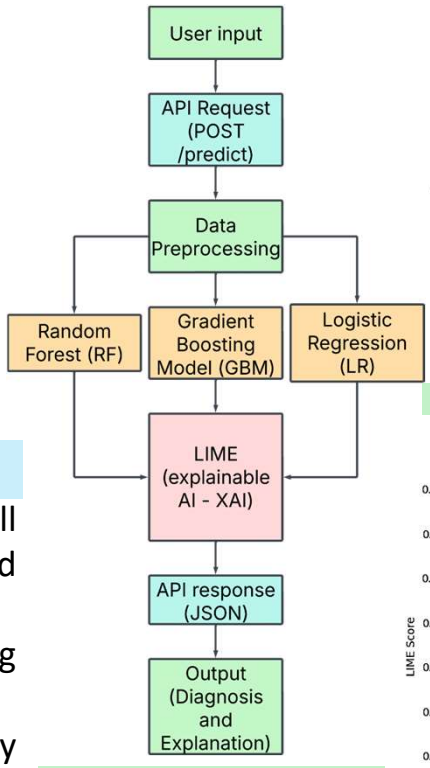

Figure 2: Methodology

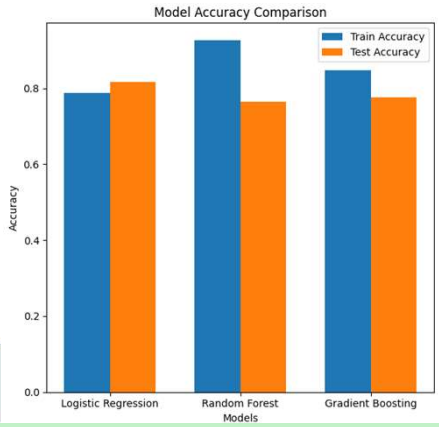

Figure 3: Model accuracy for diagnosis

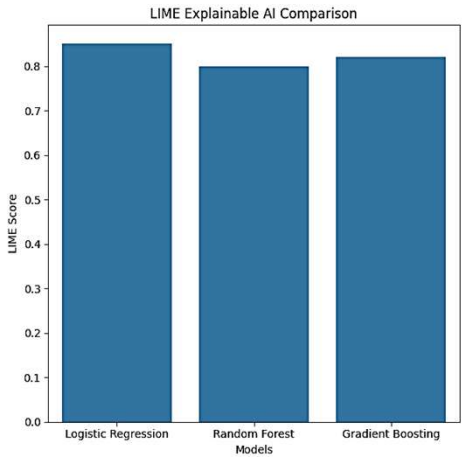

Figure 4: Comparison of LIME explainability

Results

Gradient Boosting (GBM): Best overall performance, balancing accuracy and generalisation.

Random Forest (RF): Highest training accuracy but exhibited overfitting,

Logistic Regression (LR): Lowest accuracy but the highest explainability(Figure 3 and 4)

Conclusion

The trade-off between accuracy and explainability suggests that model choice should depend on the application—LG for transparency and GBM/RF for higher predictive performance.

Future work: explore hybrid models, expand into other medical conditions

Recommendations

Random forest: overfitting should be addressed.

Gradient Boosting: interpretability with additional XAI's.

Data: could include more/larger datasets

Frontend/app: improve accessibility for patients and offer patient centric features[3]

References

[1] 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.

[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/.

[3] 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, 2023.