Journal of Digital Health Innovation and Medical Technology
Vol 1 No 2 Februari 2025
ISSN: XXXX-XXXX (Electronic)
Open Access: https://journals.ai-mrc.com/jdhimt

Explainable AI Approach in Diabetes Disease Classification Using Gradient Boosting Algorithm

Christopher Michael Lauw [1], Nenny Sulistianingsih[2]
[1,2]Universitas Bumigora, Indonesia
[1]24010820003@universitasbumigora.ac.id*, [2]neny.sulistianingsih@universitasbumigora.ac.id

Article Info

Article history:
Received 23-01-2025
Revised 03-02-2025
Accepted 15-02-2025

Keyword:
Explainable AI, Diabetes , Disease Classification, Gradient Boosting Algorithm, Healthcare Technology, AI in Healthcare.

ABSTRACT

The classification of diabetes has become a critical focus in the healthcare domain due to the disease's rising prevalence and its severe impact on global health. While machine learning methods, such as Gradient Boosting Algorithm (GBA), have shown exceptional performance in predicting diabetes, the interpretability of these models remains a challenge for practical implementation in clinical settings. This study introduces an Explainable AI (XAI) approach to enhance the transparency and interpretability of the Gradient Boosting Algorithm for diabetes classification. Using clinical indicators such as HbA1c levels, Body Mass Index (BMI), and other risk factors, the model achieves high classification accuracy while providing insights into the feature contributions through visualization techniques. SHAP (SHapley Additive exPlanations) was utilized for detailed global and local explanations, while LIME (Local Interpretable Model-agnostic Explanations) offered localized insights into individual predictions. Both LGBoost and XGBoost were compared on the same clinical dataset, where LGBoost achieved an accuracy of 97.27% and XGBoost slightly outperformed with an accuracy of 97.36%, suggesting its marginal advantage in this dataset. The results demonstrate the potential of integrating XAI in machine learning workflows to balance performance and interpretability, thereby fostering trust among healthcare practitioners and aiding in informed decision-making. This research contributes to advancing the application of explainable models in medical diagnostics

©2025 Authors. Published by PT Mukhlisina Revolution Center.. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
(https://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

Diabetes mellitus is a chronic disease whose prevalence continues to rise globally. According to data from the World Health Organization (WHO), the number of people with diabetes is estimated to reach 422 million worldwide, with the death rate from its complications steadily increasing each year(Wang et al., 2021). In Indonesia, diabetes is one of the leading non-communicable diseases that places a significant burden on the healthcare system(Oktaria & Mahendradhata, 2022). This disease not only impacts the quality of life of individuals but also imposes a considerable economic strain on patients, their families, and society at large(Awad et al., 2022).

Clinical risk factors such as Body Mass Index (BMI), glycated hemoglobin (HbA1c), blood glucose levels, age, and family history are known to play a crucial role in detecting and predicting diabetes(Chao et al., 2021). Early identification of these factors can aid in more effective prevention and management of the disease(Lin, 2024). However, analyzing these risk factors requires an approach capable of capturing the complexity of relationships between variables with high accuracy(Afzal et al., 2021).

In this context, the application of machine learning methods, particularly the Gradient Boosting algorithm, becomes highly relevant(Konstantinov & Utkin, 2021). Gradient Boosting is an ensemble learning technique that combines the strengths of several predictive models to improve classification accuracy(Mienye & Sun, 2022). In this study, 2 models were used in Gradient Boosting, namely

.

XGBoost and LGBoost. XGBoost is an efficient, fast, and accurate gradient boosting algorithm, supported by regularization and GPU computation to prevent overfitting and handle large datasets(Bentéjac et al., 2021). LightGBM, on the other hand, is a faster and more memory-efficient boosting algorithm compared to XGBoost(Ni et al., 2024). It uses a leaf-wise approach to enhance efficiency but tends to be more prone to overfitting, particularly with smaller datasets.

. This method is well-known for handling complex datasets, including those with imbalanced data or non-linear relationships between variables(Gupta & Shukla, 2023). By leveraging this algorithm, the prediction of diabetes risk can be performed more efficiently and accurately(Oikonomou & Khera, 2023).

Previous research conducted by (Argina, 2020) using the C.45 algorithm with an accuracy of 90%, with 5 attributes. Further research was conducted by (Rahayu et al., 2023) by using the Support Vector Machine (SVM) method with an accuracy of 75% and C.45 of 75%, the study used 8 data attributes. The following study was conducted by (Diana Dewi et al., 2023) by using 2 algorithms, namely SVM and Artificial Neural Network (ANN) with 8 attributes and the accuracy obtained was 77.60%, while SVM obtained an accuracy of 65.24%. The research conducted by (Gündoğdu, 2023), by using the XGBoost Classifier, an accuracy of 99.3% was obtained, but the study did not use explainable AI. In all of these studies, no researchers used the Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive explanations (SHAP) approaches. This study uses both approaches, namely SHAP and LIME. Explainable AI (XAI) is used to ensure that decisions made by the model can be understood and trusted by users (such as doctors or medical experts). This is very important to ensure that the model is used in a transparent, fair, and accountable manner, especially in medical contexts involving life-and-death decisions.

The absence of *Explainable AI (XAI)* in these studies poses a challenge for healthcare practitioners who rely on transparent and interpretable models to make informed decisions. Without explainability, the adoption of machine learning models in clinical settings is hindered by concerns over trust, accountability, and ethical considerations. XAI approaches, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), address these challenges by providing detailed insights into model predictions. These techniques ensure that decisions made by the model can be understood and trusted by users, such as doctors or medical experts.

This research aims to bridge the gap in previous studies by developing a classification model for diabetes based on clinical risk factors using Gradient Boosting methods. By integrating SHAP and LIME, this study seeks to balance performance and interpretability, enabling transparent and accountable decision-making in clinical practice. The proposed model contributes to strengthening data-driven approaches for the prevention and management of diabetes, particularly in Indonesia.

RESEARCH METHODS

`

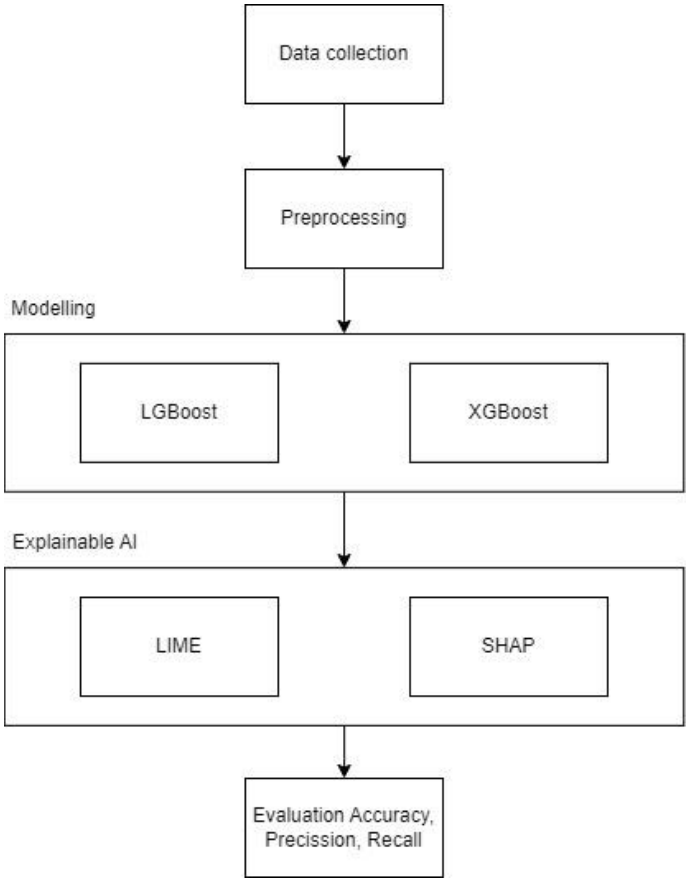

Figure 1 Research Method For Diabetes Classification

Kaggle was chosen as the primary data source due to its extensive repository of high-quality datasets, which are often well-documented and accessible for machine learning research. The platform provides a diverse range of datasets curated by an active global community, making it a reliable starting point for experimentation. However, we acknowledge potential biases, such as over-curation or the lack of real-world representativeness, which may arise from datasets predominantly created for competitions or academic purposes. To address these limitations, we performed a thorough exploratory data analysis (EDA) to validate the dataset's structure and distribution, ensuring it aligns with the objectives of our research.

Preprocessing involved cleaning and preparing the dataset to ensure its quality and suitability for analysis. Missing values were imputed using the median for numerical variables and the mode for categorical ones, preserving the integrity of the data. Outliers were detected using the Interquartile Range (IQR) method and handled based on their potential impact on the model. Outliers were either capped at the nearest boundary or removed entirely to ensure consistency. These preprocessing steps were validated through exploratory analysis to confirm that the adjustments maintained the dataset's original distribution and patterns.

Machine learning models, such as LGBoost (Light Gradient Boosting Machine) and XGBoost (Extreme Gradient Boosting), were employed due to their proven efficiency and ability to handle large datasets. These models were trained on the prepared data to build predictive models. To enhance interpretability, LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) were applied. LIME approximates model predictions using interpretable surrogate models, while SHAP provides insights into individual feature contributions using a game-theory-based approach.

The model’s performance was assessed using metrics such as accuracy, precision, and recall. Accuracy measured the proportion of correct predictions, precision indicated the proportion of correct positive predictions, and recall evaluated the proportion of actual positive cases correctly identified. This workflow provides a comprehensive framework for building, interpreting, and evaluating machine learning models.

`

RESULTS AND DISCUSSION

The first step involves data collection, with the data obtained from the Kaggle website, consisting of 100,000 records. Once the data is acquired from Kaggle, the collected data can be seen in the image below.**Table 1 Diabetes Dataset View**

gender	age	hypertension	Heart disease	Smoking history	bmi	HbA1c level	Blood glucose level	diabetes
Female	80.0	0	1	never	25.19	6.6	140	0
Female	54.0	0	0	No Info	27.32	6.6	80	0
Male	28.0	0	0	never	27.32	5.7	158	0
Female	36.0	0	0	current	23.45	5.0	155	0
Male	76.0	1	1	current	20.14	4.8	155	0

The dataset contains 8 features, which are the criteria used for classifying diabetes, ultimately determining whether an individual is at risk of diabetes or not. The explanations for these 8 features are as follows:

Table 2 Diabetes Fiture Uses

Fitur	Deskripsi
Gender	Patient gender. Possible values: Male or Female.
Age	Patient age in years.
Hypertension	History of high blood pressure. Values: 0 (None) or 1 (Present).
Heart Disease	History of heart disease. Values: 0 (None) or 1 (Present).
Smoking History	Patient smoking history. Values: Never smoked, Former smoker, Current smoker, or Unknown.
BMI	Patient Body Mass Index, calculated as weight (kg) divided by height squared (m^2).
HbA1c Level	Average blood glucose level over the past 2-3 months, expressed as a percentage (%).
Blood Glucose Level	Patient blood glucose level at the time of examination (mg/dL).
Diabetes	Patient diabetes status. Values: 0 (None) or 1 (Present).

Before performing modeling with machine learning, preprocessing is conducted by calculating the missing values in the dataset. The next step is data preprocessing, as shown in Figure 2. There are 18 missing values in the *gender* feature and 10,451 missing values in the *smoking_history* feature, as depicted in the figure below.

```
Missing values after preprocessing:
gender                 18
age                     0
hypertension            0
heart_disease           0
smoking_history     10451
bmi                     0
HbA1c_level             0
blood_glucose_level     0
diabetes                0
dtype: int64
```

Figure 2 Missing Value Before Preprocessing

.

Next, data preprocessing is carried out by removing the missing values. After the removal of the missing data, the result shows that there are no missing values in any of the features, as seen in the image below.

```
Missing values after preprocessing:
gender                0
age                   0
hypertension          0
heart_disease         0
smoking_history       0
bmi                   0
HbA1c_level           0
blood_glucose_level   0
diabetes              0
dtype: int64
```

Figure 3 Missing Value After Preprocessing

In the data preprocessing stage, the age category feature is transformed into a numerical feature, where 0 represents female gender, 1 represents male gender, and 2 represents unknown gender. Additionally, the *Smoking_history* category is transformed, where 0 indicates individuals who have never smoked, 1 indicates no information about the patient's smoking history, 2 indicates former smokers, and 3 indicates active smokers, as shown in the table below.

Table 3 Result Feature Transformed

Gender	age	hypertension	Heart disease	Smoking history	bmi	HbA1c level	Blood Glucose level	diabetes
0.0	80.0	0	1	0.0	25.19	6.6	140	0
0.0	54.0	0	0	1.0	27.32	6.6	80	0
1.0	28.0	0	0	0.0	27.32	5.7	158	0
0.0	36.0	0	0	3.0	23.45	5.0	155	0
1.0	76.0	1	1	3.0	20.14	4.8	155	0

After preprocessing, the next step is modeling using XGBoost and LGBoost. The evaluation results for XGBoost, based on precision, recall, and accuracy, are as follows

Table 4 XGBoost Evaluation Result

Precision	Recall	Accuracy
97%	100%	97.27%

The evaluation results for the LGBM classifier, based on precision, recall, and accuracy, are as follows

Table 5 LGBoost Evaluation Result

Precision	Recall	Accuracy
97%	100%	97.36%

Both the LGBM and LGBoost classifiers demonstrated strong performance in terms of precision, recall, and accuracy. The LGBM classifier achieved an accuracy of 97.27%, with a precision of 97% and a perfect recall of 100%. The LGBoost classifier performed slightly better, with an accuracy of 97.36%, a precision of 97%, and a recall of 100%. Both models exhibit high precision and recall, making them effective in correctly identifying positive cases, with minimal false positives and false negatives. Overall, both models are highly effective for the task at hand.

The SHAP results for the LightGBM (LGBoost) model reveal the relative impact of each feature on the model output. HbA1c_level exhibits a strong positive impact, where higher levels are associated with increased model outputs. Similarly, blood_glucose_level and age have moderate

`

positive contributions, indicating that elevated blood glucose levels and older age are linked to higher model predictions. Conversely, bmi shows a moderate negative impact, with lower BMI values being associated with higher model outputs. Smoking_history has a weak negative influence, suggesting that non-smokers are slightly associated with higher outputs. Both gender and hypertension demonstrate very weak impacts on the model output, with gender contributing negatively and hypertension positively. Notably, heart_disease has a strong negative impact, indicating that individuals with heart disease are significantly associated with lower model predictions. These results highlight the varying degrees of influence that each feature has on the model's decision-making process.The results of the SHAP graph for LGBoost can be seen in the image below.

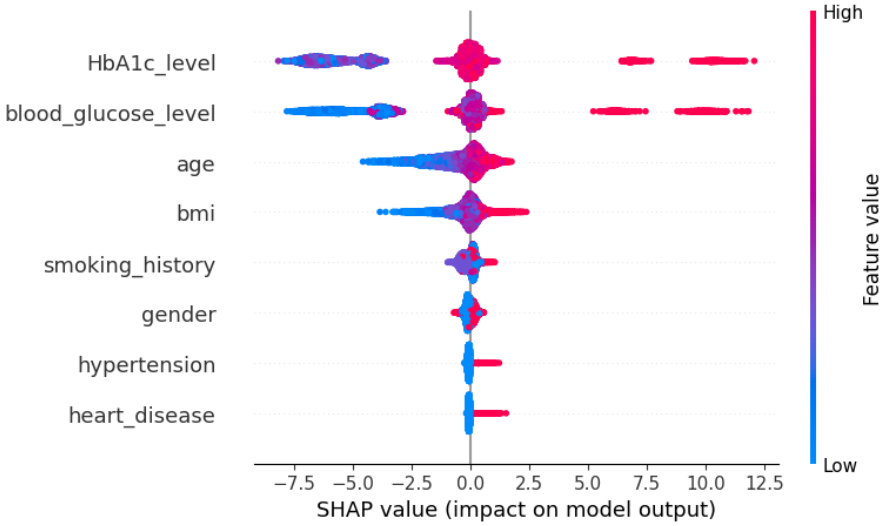

Figure 4 SHAP results for the LightGBM (LGBoost) model

The SHAP analysis results show that several factors significantly influence the output of both the LGBoost and XGBoost models. *HbA1c_level* has a strong positive impact on the model's predictions, where higher HbA1c levels are associated with higher predicted values. *Blood_glucose_level* also has a moderate positive impact on the model output, indicating that higher blood glucose levels are linked to higher predictions. The factor of age (*age*) shows a moderate positive effect, where older individuals tend to have higher predicted values. The SHAP results for LGBoost indicate that HbA1c_level and blood_glucose_level are the most influential features in predicting diabetes risk. These insights enable healthcare practitioners to focus on patients with elevated HbA1c levels, ensuring timely interventions such as dietary adjustments or medication. Similarly, the identification of age as a contributing factor allows for prioritizing screening for older individuals who may otherwise be overlooked. By providing transparency into the model's decision-making, SHAP and LIME foster trust and aid clinicians in making informed, data-driven decisions. In contrast, *bmi* has a moderate negative impact, with lower BMI values associated with higher predicted values. *Smoking_history* also exhibits a moderate negative effect, where non-smokers are slightly more likely to have higher predicted values compared to smokers. *Gender* and *hypertension* have a very weak impact on the model output, both positively and negatively. *Heart_disease* has a strong negative impact, with individuals having heart disease associated with significantly lower predicted values. Overall, SHAP values indicate that both the LGBoost and XGBoost models are influenced by similar factors, with *HbA1c_level* and *blood_glucose_level* having the strongest positive impact on both model outputs, while *heart_disease* has a strong negative effect on both models. Age also plays a moderate role in influencing the model output, with older individuals generally having higher risk predictions.

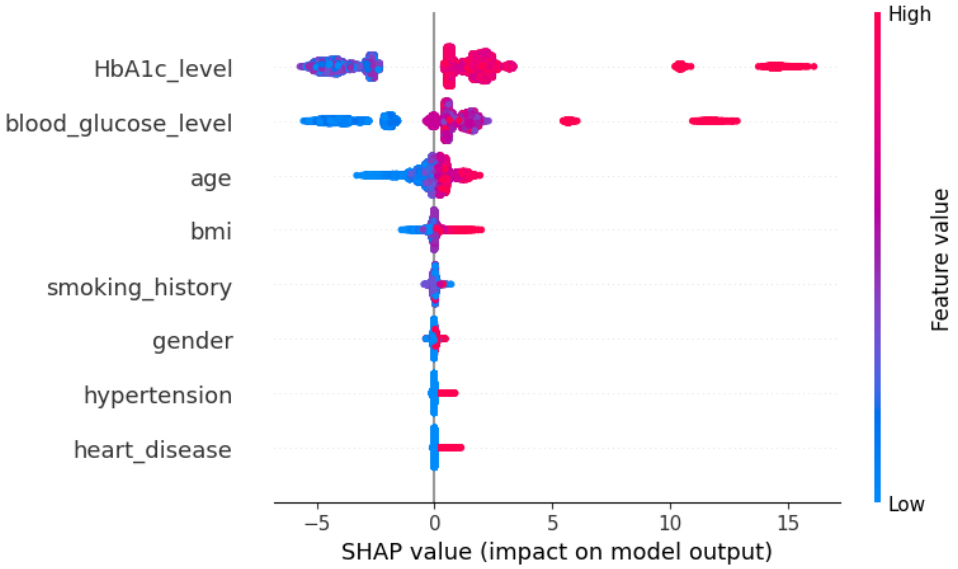

Figure 5 SHAP results for the XGboost model

Both images represent Explainable AI (XAI) visualizations used to interpret the prediction decisions of two different models, in a diabetes classification task. the LightGBM model predicts the patient as non-diabetic (*No Diabetes*) with a probability of 1.00. The most influential feature driving this prediction is HbA1c_level, which strongly supports the *No Diabetes* class. Additional features such as heart_disease, blood_glucose_level, and hypertension also contribute positively to this prediction. However, some features, including bmi and smoking_history, show minor contributions toward the *Diabetes* class, though their overall impact remains negligible.

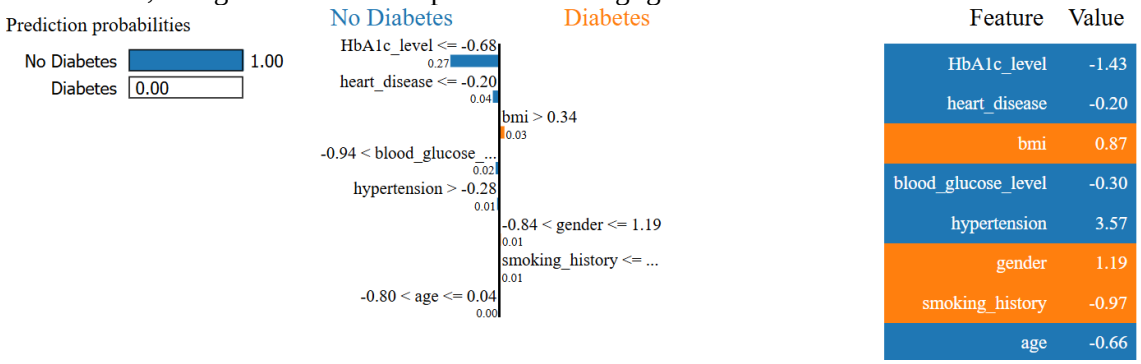

Figure 6 LIME results for the LightGBM model

the XGBoost model also predicts the patient as non-diabetic (*No Diabetes*) with a probability of 1.00. Similarly, **HbA1c_level** emerges as the dominant feature influencing this prediction. Other features, such as **hypertension** and **bmi**, exhibit slight contributions toward the *Diabetes* class but are outweighed by the stronger influence of features supporting *No Diabetes*. Compared to LightGBM, XGBoost places a more significant emphasis on the primary feature, **HbA1c_level**, with less distributed contributions from the remaining features. While both models produce identical predictions, the differences in feature attribution reflect the distinct mechanisms employed by LightGBM and XGBoost in processing and prioritizing input data.

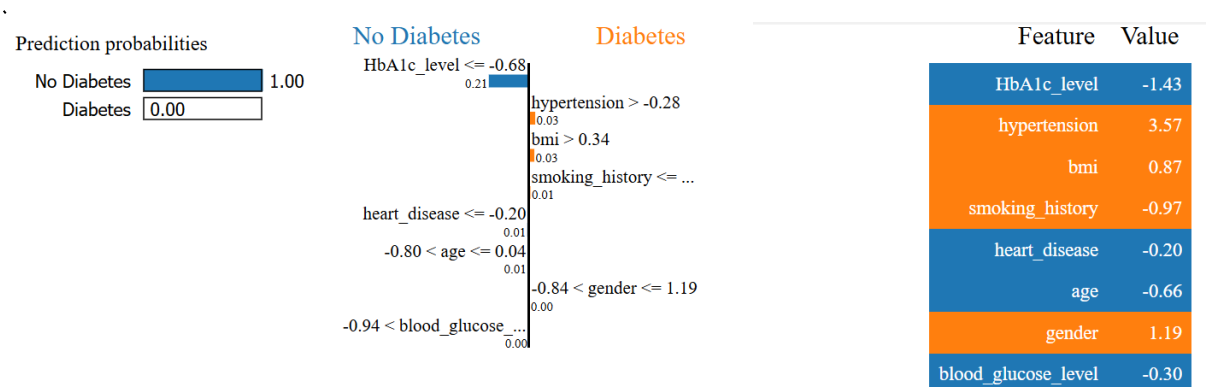

Figure 7 LIME results for the XGboost model

While SHAP provides detailed feature contributions, its computational complexity makes it less feasible for real-time decision-making in clinical settings. Additionally, the explanations may require further simplification or visualization for practitioners without technical expertise. LIME, on the other hand, offers a lightweight alternative but may produce less reliable explanations for highly complex models, as it relies on local approximations. Addressing these limitations will require developing more efficient and user-friendly tools tailored to clinical workflows.

CONCLUSION

This study demonstrates that LightGBM (Light Gradient Boosting Machine) achieves the highest accuracy in diabetes classification, outperforming XGBoost (Extreme Gradient Boosting) by a small margin. The SHAP analysis reveals that key clinical features, such as HbA1c levels and blood glucose levels, significantly influence the model's predictions, providing valuable insights into the factors contributing to diabetes risk. These findings highlight the potential for AI models to not only deliver high predictive performance but also enhance transparency and trust through explainable AI techniques like SHAP and LIME.

From a healthcare perspective, the ability of SHAP and LIME to provide interpretable insights into model decisions can empower clinicians to make data-driven, patient-specific decisions. For example, identifying high HbA1c levels as a primary predictor allows practitioners to prioritize early interventions and monitor at-risk patients more effectively. Furthermore, by fostering trust in AI-driven predictions, these models pave the way for broader adoption of machine learning solutions in clinical workflows, promoting better patient outcomes and resource optimization.

For future research, addressing the limitations of this study is essential. Instead of removing missing data, imputation techniques should be employed to retain valuable information, particularly when working with large datasets. Additionally, future work should explore more robust methods for handling outliers, as their presence can adversely affect model performance. Expanding on these areas will further refine the application of explainable AI in clinical settings, ensuring that models remain accurate, interpretable, and clinically relevant.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Afzal, F., Yunfei, S., Nazir, M., & Bhatti, S. M. (2021). A review of artificial intelligence based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects. *International Journal of Managing Projects in Business*, *14*(2), 300–328.

Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, *1*(2), 29–33. https://doi.org/10.33096/ijodas.v1i2.11

.

Awad, S. F., Critchley, J. A., & Abu-Raddad, L. J. (2022). Impact of diabetes mellitus on tuberculosis epidemiology in Indonesia: A mathematical modeling analysis. *Tuberculosis*, *134*, 102164.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937–1967.

Chao, G., Zhu, Y., & Chen, L. (2021). Role and risk factors of glycosylated hemoglobin levels in early disease screening. *Journal of Diabetes Research*, *2021*(1), 6626587.

Diana Dewi, D., Qisthi, N., Lestari, S. S. S., & Putri, Z. H. S. (2023). Perbandingan Metode Neural Network Dan Support Vector Machine Dalam Klasifikasi Diagnosa Penyakit Diabetes. *Cerdika: Jurnal Ilmiah Indonesia*, *3*(09), 828–839. https://doi.org/10.59141/cerdika.v3i09.662

Gündoğdu, S. (2023). Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique. *Multimedia Tools and Applications*, *82*(22), 34163–34181. https://doi.org/10.1007/s11042-023-15165-8

Gupta, S. K., & Shukla, D. P. (2023). Handling data imbalance in machine learning based landslide susceptibility mapping: a case study of Mandakini River Basin, North-Western Himalayas. *Landslides*, *20*(5), 933–949.

Konstantinov, A. V, & Utkin, L. V. (2021). Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, *222*, 106993.

Lin, W. (2024). The Association between Body Mass Index and Glycohemoglobin (HbA1c) in the US Population's Diabetes Status. *International Journal of Environmental Research and Public Health*, *21*(5), 517.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, *10*, 99129–99149.

Ni, C., Huang, H., Cui, P., Ke, Q., Tan, S., Ooi, K. T., & Liu, Z. (2024). Light Gradient Boosting Machine (LightGBM) to forecasting data and assisting the defrosting strategy design of refrigerators. *International Journal of Refrigeration*, *160*, 182–196.

Oikonomou, E. K., & Khera, R. (2023). Machine learning in precision diabetes care and cardiovascular risk prediction. *Cardiovascular Diabetology*, *22*(1), 259.

Oktaria, V., & Mahendradhata, Y. (2022). The health status of Indonesia's provinces: the double burden of diseases and inequality gap. *The Lancet Global Health*, *10*(11), e1547–e1548.

Rahayu, D. S., Afifah, J., & Intan, S. (2023). Classification of Diabetes Mellitus Using C4 . 5 Algorithm , Support Vector Machine (SVM) and Linear Regression Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma C4 . 5 , Support Vector Machine (SVM) dan Regresi Linear. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, *1*(1 SE-), 56–63. https://journal.irpi.or.id/index.php/sentimas/article/view/550

Wang, L., Li, X., Wang, Z., Bancks, M. P., Carnethon, M. R., Greenland, P., Feng, Y.-Q., Wang, H., & Zhong, V. W. (2021). Trends in prevalence of diabetes and control of risk factors in diabetes among US adults, 1999-2018. *Jama*, *326*(8), 704–716.