A Diabetes Prediction Model with Visualized Explainable Artificial Intelligence (XAI) Technology

Yanfeng Zhao[1], Jun Kit Chaw[1(✉)], Mei Choo Ang[1], Marizuana Mat Daud[1], and Lin Liu[2]

[1] Institute of Visual Informatics, The National University of Malaysia (UKM), Bangi, Malaysia
p117631@siswa.ukm.edu.my, {chawjk,amc,marizuana.daud}@ukm.edu.my
[2] College of Information Engineering, Henan Vocational University of Science and Technology, Zhou Kou, China
1090511629@qq.com

Abstract. Diabetes is a group of non-communicable diseases (NCD) that cannot be cured by current medical technologies and can lead to various serious complications. Significantly reducing the severity of diabetes and its associated risk factors relies on accurate early prediction. Some machine learning algorithms have been developed to assist in predicting diabetes, but their predictions are not always accurate and often lack interpretability. Therefore, further efforts are required to improve these algorithms to achieve the level of clinical application. The aim of this paper is to find a high-performance and interpretable diabetes prediction model. Firstly, the dataset is subjected to necessary preprocessing, including missing value imputation using K-nearest neighbors (KNN) and data balancing using adaptive synthetic sampling (ADASYN). Then, with 10-fold cross validation, the predictive performance of six machine learning algorithms is compared in terms of accuracy, precision, recall, and F1 score. Finally, the prediction results are globally and locally explained using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). The experimental results demonstrate that the eXtreme Gradient Boosting (XGBoost) algorithm provides the best predictive performance. The visualized eXplainable Artificial Intelligence (XAI) techniques offer valuable explanatory information, helping healthcare professionals and patients better understand the risk and prediction results of diabetes.

Keywords: Diabetes Prediction · Machine learning · eXplainable AI · LIME · SHAP

1 Introduction

Diabetes is a group of non-communicable diseases (NCD) caused by chronic metabolic disorders. Its main symptom is high blood glucose levels (BGL) due to the body's inability to produce or properly use insulin, a hormone that regulates BGL. Diabetes is a major

H. Badioze Zaman et al. (Eds.): IVIC 2023, LNCS 14322, pp. 648–661, 2024.
https://doi.org/10.1007/978-981-99-7339-2_52

global public health issue, affecting hundreds of millions of people. According to the latest report from the International Diabetes Federation (IDF), the global prevalence of diabetes in adults aged 20 to 79 was estimated to be 9.8% in 2021, with approximately 537 million people affected. It is projected that by 2045, the global diabetes prevalence will rise to 11.2, with an estimated 784 million patients [1]. Unfortunately, there is currently no medical intervention that can completely cure diabetes, leading patients to often suffer from serious complications such as cardiovascular diseases, kidney failure, blindness, and amputations. Early detection and prevention of diabetes are widely recognized as crucial in slowing down the progression of diabetes and reducing the risk of complications.

Machine learning (ML) exhibits significant potential in predicting diabetes based on patient data, as it can analyze extensive datasets and uncover patterns that may elude human observation. Numerous studies have utilized ML algorithms, including random forest, decision trees, support vector machines, and neural networks [2], to forecast or diagnose diabetes using clinical and demographic data. Regrettably, these algorithms are scarcely employed in clinical practice. Some studies have reported subpar accuracy due to inappropriate model selection and inadequate data preprocessing [3, 4]. Furthermore, certain studies, despite achieving commendable predictive performance, have neglected to provide reasonable explanations for the prediction results [5]. Consequently, this lack of interpretability has engendered a distrust in these algorithms as they prove challenging to comprehend.

The objective of this study is to develop a diabetes prediction model by integrating ML algorithms with visualized explainable artificial intelligence (XAI) techniques. The model not only achieves high performance but also offers insights into the factors contributing to diabetes episodes. The study employs targeted data preprocessing techniques, such as the K-nearest neighbors (KNN) method for missing value imputation and the adaptive synthetic sampling (ADASYN) method for dataset balancing. To identify the most suitable algorithm for diabetes prediction using the PIMA Indian diabetes dataset (PIDD), six popular machine learning algorithms are compared using cross-validation. The extreme gradient boosting (XGBoost) algorithm, demonstrating the best performance, is further examined using shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) techniques to provide global and local explanations for its predictions. This research offers a solution that improves the accuracy and reliability of diabetes prediction models while providing valuable insights for disease forecasting. This can assist clinical practitioners and researchers in developing more effective prevention and treatment strategies, ultimately enhancing the quality of life for diabetes patients.

The remaining sections of this paper are organized as follows. Section 2 presents a review of related work on diabetes prediction and interpretable ML methods. Section 3 elaborates on the research methodology, including the data preprocessing steps and modeling approaches. Section 4 discusses the experimental results and provides visual explanations of the prediction results using SHAP and LIME. Finally, Sect. 5 summarizes the paper and discusses future research directions.

2 Related Works

After extensive analysis of experimental data, it has been observed that implementing proper data pre-processing techniques and selecting suitable machine learning (ML) algorithms can enhance the accuracy of prediction models [6]. In a study by Chen et al. [7], mean interpolation was employed to handle missing values in PIDD, and a k-means clustering algorithm was utilized to eliminate misclassified data. However, the class imbalance issue in the dataset was not addressed. Another approach described in [8] involved balancing the dataset using SMOTE techniques, performing feature selection, and employing SVM-RBF kernels for classification, which resulted in improved prediction performance.

To enhance prediction accuracy, some studies have explored the use of hybrid algorithms. Ahmed et al. [9] combined support vector machines, artificial neural networks, and fuzzy logic to predict diabetes, achieving an accuracy rate of nearly 95% on a 30% test set. Building upon these insights, [10] employed an isolation forest for outlier removal, employed SMOTETomek for class balancing, and developed an integrated model for diabetes diagnosis with an impressive accuracy of 96.74%. However, it is worth noting that many of these studies primarily focused on improving the performance of the prediction models, often neglecting the interpretability aspect of the employed algorithms.

In order to address this challenge, researchers have explored the use of XAI techniques to provide explanations for model decision-making processes, facilitating understanding of their functioning. Common XAI methods in the literature include Fuzzy Rule-Based Systems [11, 12] and feature extraction-based models [13,14]. Additionally, more easily interpretable models such as decision trees or linear models have been employed. However, these models often lack intuitive interpretation methods, making it difficult to convey their meaning effectively to physicians and patients. Moreover, the adoption of interpretable machine learning algorithms frequently involves sacrificing predictive performance, thereby limiting their practical application in clinical settings.

In recent years, visual interpretation of black-box models has emerged as a prominent area of XAI research. Visualized XAI techniques, such as LIME and SHAP, have shown effectiveness in interpreting a wide range of machine learning and deep learning models. These methods provide visual explanations, enhancing the interpretability of the models and facilitating their comprehension by medical professionals and patients alike. In the field of predicting gestational diabetes mellitus, [15] and [16] have developed ML models using different methods. They have effectively employed the SHAP technique to visualize both global and local interpretations, significantly enhancing the credibility and acceptability of their models. Similarly, in another study by [17], a prediction model was proposed based on common factors associated with diabetes, such as blood glucose, BMI, age, and insulin. The LIME technique was utilized to identify the key attributes and the prediction rules of diabetic organization. In line with the approach presented in this paper, [18] and [19] have integrated the LIME and SHAP frameworks to develop a visualized XAI approach that explains how the model generates final prediction results. This integration allows for a comprehensive understanding of the underlying decision-making processes, thereby improving transparency and interpretability. Our work further extends their work by providing a comprehensive explanation of the predictive models from both global and local perspectives. Additionally, it delves into the implied meaning

of the discriminative features to examine how diabetes is predicted with our proposed model. This work also shows how XAI models can augment the predictive model to provide trust and confidence in their predictions.

3 Proposed Methodology

To achieve both high-performance and interpretable diabetes prediction models, this paper aims to integrate the strengths of previous studies and utilize more appropriate and effective techniques in three key stages: data preprocessing, model development and validation, and model interpretation. The overall framework of the proposed approach is illustrated in Fig. 1.

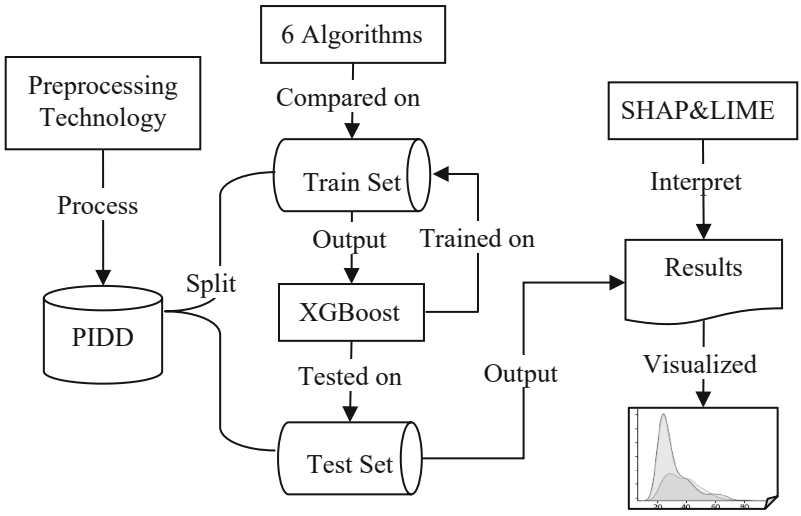

Fig. 1. Overall framework of the proposed approach

3.1 Date Preprocessing

The primary objective of this study was to explore viable solutions that strike a balance between predictive performance and interpretability of the model, rather than solely focusing on achieving high accuracy. To achieve this, we utilized the publicly available PIDD dataset, comprising only 768 samples. The dataset includes eight characteristic variables (such as the number of pregnancies, BMI, insulin levels, age, etc., of patients) and one target variable. Upon conducting exploratory data analysis, it was observed that the dataset exhibited various issues, including missing values, outliers, and data imbalances. Ignoring these issues and directly building a prediction model would inevitably yield implausible results. As a result, it was imperative to perform necessary data pre-processing steps. Here, we outline the essential pre-processing steps:

A. Missing Value Imputation. Although the PIDD dataset initially appears to have no null values, zero values in the characteristic variables 'Glucose', 'BloodPressure', 'BMI', 'SkinThickness', and 'Insulin' are unlikely to occur in surviving patients. Thus, treating these zero values as missing values and performing appropriate imputation is necessary. We chose not to use the median imputation based on target values, as described in [6], despite its higher predictive performance. This decision was influenced by two factors: the significant number of missing values in 'SkinThickness' (29.6%) and 'Insulin' (48.7%), and the fact that target value-based imputation would overestimate predictive performance, which is not feasible in real clinical practice where future outcomes are unknown. Instead, we employed the following approach:

Median imputation: For 'Glucose', 'BloodPressure', and 'BMI', which had fewer missing values, we utilized median imputation. This involved replacing the missing values with the median value of each respective characteristic variable.

KNNImputer: For 'SkinThickness' and 'Insulin', which had a higher percentage of missing values, we applied the KNNImputer method [20]. This technique identifies k samples in the dataset that are spatially or structurally similar to the missing value samples and calculates the mean of the k neighbors for interpolation. In this study, we assigned weights as the inverse of the distance to give more influence on nearby domains.

B. Data Balancing. The PIDD dataset consisted of 268 samples diagnosed with diabetes and 500 samples without diabetes, resulting in a data imbalance. Considering the small sample size of the dataset, the Synthetic Minority Oversampling Technique (SMOTE) is a commonly used algorithm for data balancing. However, SMOTE may introduce overlapping categories by interpolating minority class samples between majority class samples. ADASYN is an improvement over SMOTE because it adaptively generates different numbers of new samples for different minority class samples based on the data distribution [21]. The abbreviated algorithm works as follows: first, the dataset's imbalance is evaluated using Eq. (1) and compared to a set threshold. If the imbalance is lower than the threshold, the synthetic samples are generated using Eq. (2) on the basis of calculating the number of synthetic samples for each minority class sample x_i.

$$d = \frac{m_s}{m_l} \tag{1}$$

where m_s and m_l represent the quantities of minority class samples and majority class samples respectively, and d $\in$ (0,1].

$$s_i = x_i + \lambda(x_{zi} - x_i) \tag{2}$$

where x_i is an instance in the minority class, $(x_{zi} - x_i)$ is the difference vector in the feature space, and λ is a random number between 0 and 1.

Although tree-based classification algorithms do not necessarily require data normalization, we performed this step in order to compare the predictive performance of multiple ML algorithms. Additionally, after populating the missing values, we removed extreme outliers from the dataset, resulting in a final sample size of 722 for the development and validation of the prediction model.

3.2 Model Development and Validation

After data preprocessing, the dataset was split into a training set comprising 70% of the data and a test set comprising the remaining 30% to facilitate the development, training, and validation of the proposed model.

A. Algorithm Selection. To select the best performing prediction algorithm, we reviewed relevant literature and identified Logistic Regression (LR), Gaussian Parsimonious Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), and XGBoost as potential candidates. We evaluated these six algorithms separately by training them on the training set and using 10-fold cross-validation to obtain accuracy, precision, recall, and F1 scores. Based on these metrics, XGBoost was identified as the best performing algorithm.

XGBoost [22] is an improved version of the gradient boosting algorithm that is known for its simplicity, speed, and excellent results. As show in Eq. (3), its loss function comprises two components: the loss of the gradient boosting algorithm and a regularization term.

$$L(\Phi) = \sum_{i=1}^{n} l(y_i\prime, y_i) + \sum_{k} \Omega(f_k) \tag{3}$$

where n represents the number of training samples, l represents the loss of a single sample, $y_i\prime$ and y_i respectively represent the predicted value and true label value of the model for the training sample. The regularization term is used to describe the complexity of the model as follows:

$$\Omega(f_k) = \Upsilon T + \frac{1}{2}\lambda ||w||^2 \tag{4}$$

where γ and λ are hyperparameters, w is the vector formed by the values of all leaf nodes of the decision tree, and T is the number of leaf nodes.

B. Model Training and Testing. For the XGBoost algorithm selected, the hyperparameters need to be fine-tuned to achieve optimal prediction performance. We empirically set candidate values for the hyperparameters "learning_rate", "min_samples_split", "max_depth", "subsample", and "n_estimators", respectively. The Grid Search with Cross Validation method is then utilized to search for the best combination of these parameters to construct the final prediction model. The resulting model is trained on the training set, and the final model with optimal parameters is validated on the test set. In addition to calculating accuracy, precision, recall, and F1 scores, the ROC curve of the model is plotted, and the AUC value is calculated to provide a more objective evaluation of the model's performance.

3.3 Model Interpretation

As mentioned earlier, we employed a combination of SHAP and LIME techniques to provide both global and local explanations for the final model. These techniques allowed us to visualize and interpret the model's predictions comprehensively.

A. SHAP Technique: SHAP is an additive explanation model based on cooperative game theory that can provide explanations for the output of any machine learning model [23]. Given that x_i represents the ith sample, x_i^j represents the jth feature of the ith sample, y_i represents the predicted value of the model for that sample, and y_{base} represents the baseline of the entire model (typically the mean of the target variables for all samples), the SHAP value follows the following equation:

$$y_i = y_{base} + f\left(x_i^1\right) + f\left(x_i^2\right) + \cdots + f\left(x_i^k\right) \tag{5}$$

where $f\left(x_i^j\right)$ represents the SHAP value of x_i^j, which indicates the contribution of the jth feature in the ith sample to the final prediction value y_i. A positive $f\left(x_i^j\right)$ value implies that the feature has a positive effect on the predicted value, while a negative value suggests a negative effect. This provides a more detailed understanding of how each feature influences the prediction outcome compared to conventional feature importance measures that only indicate the importance of a feature.

B. LIME Technique. Unlike delving deep into the inner workings of the model, LIME focuses on understanding how the output of a black box model changes by applying slight perturbations to the input data. It achieves this by training an interpretable model specifically around the point of interest, which is the original input [17]. The representation of data in LIME is defined as follows:

$$\exp(x) = argmin_{g \epsilon G} L(f, g, \pi_x) + \Omega(g) \tag{6}$$

Here, g represents the explanatory model for a given instance x. The approximation of the original model f and the model g is compared by minimizing the loss function. $\Omega(g)$ represents the model complexity of g, and G encompasses all potential explanatory models. The neighborhood of x is defined by π_x. By determining the model g, the neighborhood range size π_x, and the model complexity $\Omega(g)$, the model f can be made interpretable by minimizing the loss L.

4 Results and Discussions

The proposed method was implemented and tested in the Python 3.9 integrated environment. This section aims to present the experimental results of algorithm selection, model validation, and visual model interpretation, along with a concise discussion.

4.1 Performance Comparison of Different ML Algorithms

During the algorithm selection stage, the classification results of six machine learning (ML) algorithms were analyzed to determine their prediction performance. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) results were extracted and used to calculate and compare the performance of the algorithms. Metrics such as accuracy, precision, recall, and F1 score were utilized to evaluate their effectiveness. Higher values of these metrics indicate superior prediction performance.

Table 1 compares these performance metrics, using the default parameter settings for each algorithm for fairness. The XGBoost algorithm achieves accuracy of 0.90/0.90, precision of 0.84/0.92, recall of 0.95/0.80 and F1 score of 0.89/0.84 on the training and test sets respectively. Although the KNN algorithm achieves equal or better results in some metrics, overall the XGBoost algorithm still has a clear advantage. On average, compared to the KNN algorithm, which has the second highest overall performance, the XGBoost algorithm improved by 2.9% in accuracy, 6.7% in precision, 3.6%% in F1 score, and only slightly decreased in recall by 1.1%. These metrics highlight the advantages of using the XGBoost algorithm for the diabetes classification problem. Based on these findings, the XGBoost algorithm was selected as the preferred model for further analysis and modeling.

Table 1. Performance comparison of different ML algorithms

Algorithms	Accuracy		Precision		Recall		F1 Score	
Dataset	Train	Test	Train	Test	Train	Test	Train	Test
LR	0.74	0.77	0.73	0.63	0.76	0.64	0.72	0.62
NB	0.75	0.81	0.76	0.70	0.78	0.73	0.75	0.70
KNN	0.85	**0.90**	0.77	0.88	**0.95**	**0.82**	0.83	**0.84**
DT	0.83	0.85	0.80	0.80	0.83	0.79	0.81	0.78
SVM	0.78	0.84	0.81	0.72	0.79	0.80	0.77	0.75
XGBoost	**0.90**	**0.90**	**0.84**	**0.92**	**0.95**	0.80	**0.89**	**0.84**

4.2 Test Results of the Final Model

First, 1080 candidate combinations of hyperparameters (433310) were generated for the selected XGBoost algorithm. By applying a 10-fold cross-validation technique, 10800 fits were obtained. These fits were then used as input for automatic parameter tuning using the GridSearchCV tool. The optimal parameter combination was found to be: learning_rate = 0.2, max_depth = 5, min_samples_split = 0.2777777777778, n_estimators = 100, sub-sample = 1.0. Subsequently, the XGBoost model, employing this optimal parameter combination, was validated on the test set. The obtained results were as follows: accuracy of 0.89, precision of 0.84, recall of 0.86, and F1 score of 0.85. Although these scores are slightly lower than those achieved on the training set, we consider this difference reasonable. Overall, the prediction performance remains excellent.

To further validate the usefulness and accuracy of the proposed algorithm, we employed the area under the curve (AUC) and receiver operating characteristic (ROC) curve. Figure 2 displays the ROC curve for the final model, demonstrating that the parameter-tuned XGBoost algorithm achieved an AUC value close to 0.97, indicating a significantly high area under the ROC curve. Additionally, we evaluated the risk differentiation ability of the model by calculating the Kolmogorov-Smirnov (KS) value.

The KS value represents the maximum difference between the cumulative distribution of good and bad samples. A larger cumulative difference signifies a stronger risk differentiation ability, resulting in a higher KS metric. In our final model, the KS value is close to 0.81, indicating that the proposed algorithm excels in risk differentiation.

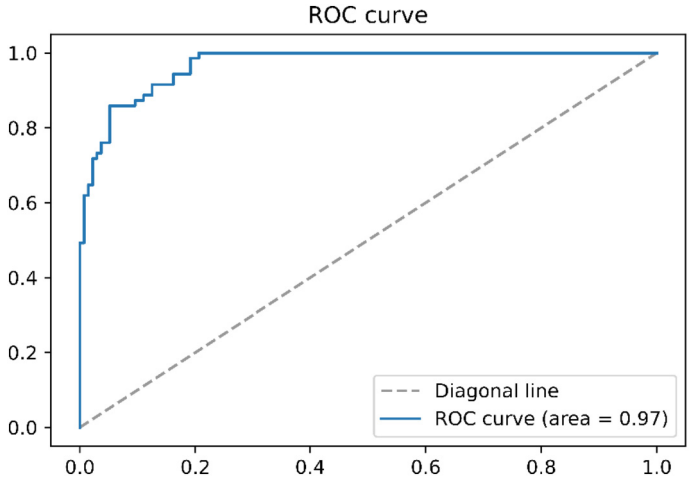

Fig. 2. ROC curve of the final model

It is worth mentioning that the PIDD dataset is characterized by a small sample size and comprises only eight predictor variables. This limitation hinders the potential for further enhancements in prediction performance. It is also possible that better prediction results could have been achieved by employing the ensemble models [24]. it is crucial to emphasize that our primary objective did not revolve around maximizing performance metrics alone. Therefore, we did not deliberately select a better dataset or an ensemble method in this particular context.

4.3 Visualized Model Explanation

In this study, we aimed to determine the importance of each feature and provide an explanation for the role of these features in influencing specific decisions. Both global and local representations of the feature impact are presented and discussed in detail.

A. Global Explanation.
The global interpretation aims to demonstrate how features contribute to a given dataset within a specific algorithm. In this study, the permutation importance of the features is utilized to fulfill this purpose. As depicted in the Fig. 3, the XGBoost algorithm highlights Glucose as the most influential feature, followed by BMI, SkinThickness, Age, PedigreeFunction, and Insulin, in descending order. Conversely, BloodPressure and Pregnancies are deemed the least significant features. The summary plot provides additional valuable information, with the y-axis representing feature importance rankings

and colors denoting low to high feature values. The x-axis showcases Shapley values, and the jittering of the overlapping points upward on the y-axis allows for visualizing the distribution of Shapley values for each feature. Figure 4 reaffirms the ranking of feature influence seen in Fig. 3, revealing a corresponding increase in diabetes risk with higher levels of blood glucose, BMI, and age. Such a conclusion aligns perfectly with existing medical knowledge.

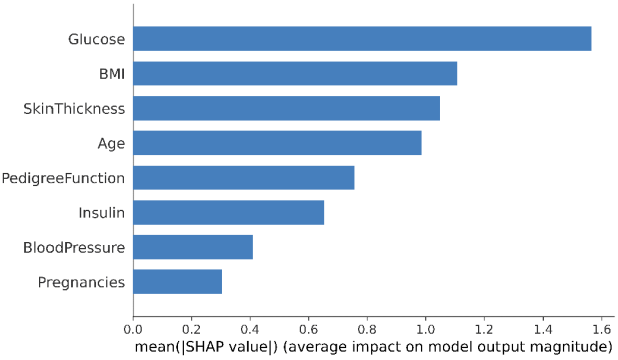

Fig. 3. The permutation importance of the features for the final model

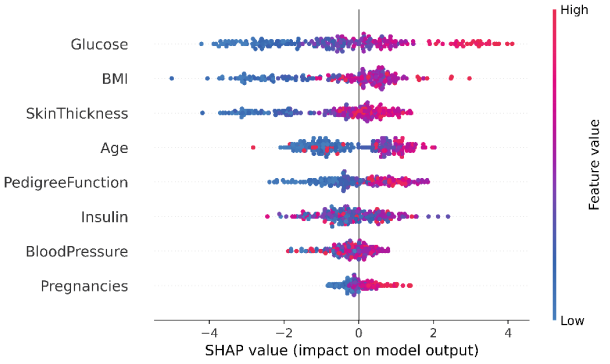

Fig. 4. The summary plot for the final model

The dependence plot illustrates the impact of an individual feature on the variability in the predicted outcome of a model. It provides insights into how a single feature influences the model's output by comparing the SHAP value of that feature with the feature values across the entire dataset. Figure 5 visualizes the changes in predictions as the Glucose feature varies. It is evident that the Glucose value exhibits a near-linear relationship with its corresponding SHAP value, indicating that higher Glucose levels are associated with an increased risk of being predicted to have diabetes. Additionally, the color-coded dots represent the distribution of another feature, Insulin, throughout the range of Glucose values. Notably, there is a clear tendency for the Insulin value to rise as the blood glucose levels increase.

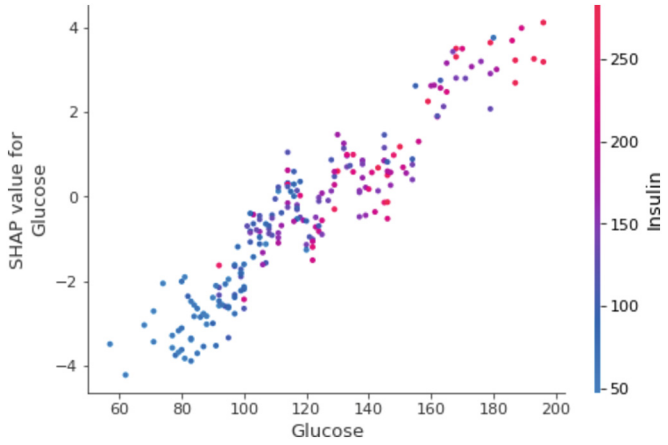

Fig. 5. The dependence plot of Glucose and Insulin for the final model

B. Local explanation.
Local explanation is employed to elucidate the contribution of each feature to the predicted values in a specific test dataset. Figure 6 showcases the SHAP force plot for a sample from the test dataset. The width of the color bar in the plot represents the magnitude of a feature's contribution to the predicted values. Red indicates a positive contribution, while blue signifies a negative contribution. The summation of contributions from all features results in the transformation of the target variable from its baseline value (the average value of the target variable across all records) to the final predicted value.

In this case, the factors that escalate the risk of developing diabetes, ranked in order of importance, are higher Glucose, BMI, Insulin, and SkinThickness values. Conversely, being 26 years old and having 0 pregnancies diminish the risk of developing diabetes. However, the negative contributions fail to counterbalance the positive contributions adequately. As a result, the predicted risk of diabetes for this sample is higher (with a predicted value of 2.58), signifying an increase of approximately 2.5 units from the baseline value.

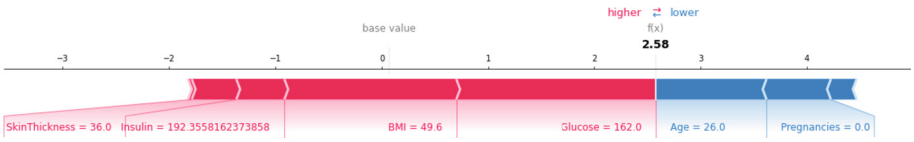

Fig. 6. The SHAP force plot of a specific sample

LIME is another model-agnostic method for local explanation. It can display the positive and negative impact of each feature on decision-making, helping doctors understand why the model makes certain predictions. Figure 7 presents the LIME plot for a specific sample from the test dataset. In the plot, blue represents features that decrease

the risk of diabetes, while orange represents features that increase the risk. It is evident that the probability of having diabetes for this sample is 0.82. The features that elevate the risk of diabetes, in order of importance, are Glucose, BMI, SkinThickness, and Age. On the other hand, the features PredigreeFunction and BloodPressure decrease the risk of diabetes. Pregnancies and Insulin, however, do not significantly affect the predicted outcome.

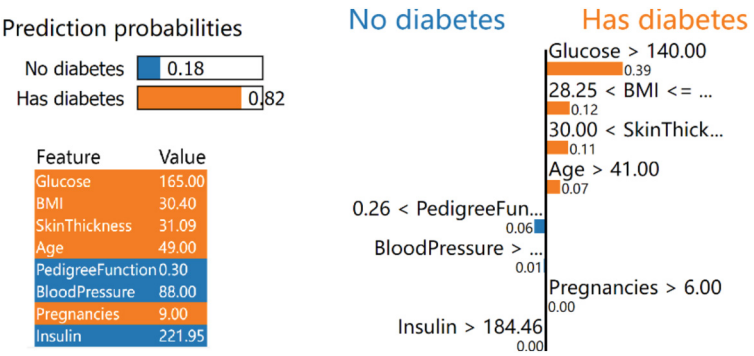

Fig. 7. The LIME tabular explainer of a specific sample

Interestingly, based on the visualized explanation results, we drew the inference that the majority of diabetic patients in the PIMA dataset were type 2 diabetic. This inference was drawn from the positive correlations observed between BMI, skin thickness, age, and insulin with the diabetes prediction results in both global and local interpretations, which aligns with the characteristics of type 2 diabetes. Conversely, indicators such as BMI, skin thickness, age, and insulin are likely to exhibit negative associations with predictive outcomes for type 1 diabetes. Although we couldn't find specific evidence supporting this conclusion in the existing literature, it has been endorsed by several physicians.

5 Conclusion

This study conducted data pre-processing and compared the prediction performance of six ML algorithms using cross-validation techniques. Based on the findings, a diabetes prediction model was developed using the XGBoost algorithm, which achieved high performance and interpretability. The results demonstrate that the final model achieved high performance with an accuracy, precision, recall, F1 score, and AUC value of 0.89, 0.84, 0.86, 0.85, and 0.97, respectively. These impressive results were obtained by fine-tuning the optimal parameters using the GridSearchCV technique. To enhance the interpretability of the prediction results, the proposed model employed SHAP and LIME techniques for global and local explanation. These techniques serve as powerful tools for physicians to assess the accuracy of prediction models. Moreover, it highlights the significant potential of these model-independent interpreters in providing visualized explanations for any ML model.

Acknowledgement. The authors would like to thank the Universiti Kebangsaan Malaysia for support-ing this work through Geran Galakan Penyelidik Muda (GGPM-2022-063) and re-search incentives TAP-K024478.

References

1. International Diabetes Federation., IDF Diabetes Atlas, 10th edn., International Diabetes Federation (2021). https://diabetesatlas.org/data/en/world/ (Accessed 06 May 2023)
2. Al Sadi, K., Balachandran, W.: Prediction model of type 2 diabetes mellitus for Oman prediabetes patients using artificial neural network and six machine learning Classifiers. Appl. Sci., **13**(4) (2023). https://doi.org/10.3390/app13042344
3. Sisodia, D., Sisodia, D.S.: Prediction of diabetes using classification algorithms. Procedia Comput. Sci., **132**(Iccids), 1578–1585 (2018). https://doi.org/10.1016/j.procs.2018.05.122
4. Mahboob Alam, T., et al.: A model for early prediction of diabetes. Informatics Med. Unlocked **16**, 100204 (2019). https://doi.org/10.1016/j.imu.2019.100204
5. Tiwari, P., Singh, V.: Diabetes disease prediction using significant attribute selection and classification approach. J. Phys. Conf. Ser., **1714**(1) (2021). https://doi.org/10.1088/1742-6596/1714/1/012013
6. Kibria, H.B., Nahiduzzaman, M., Goni, M.O.F., Ahsan, M., Haider, J.: An Ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. Sensors 22(19) (2022). https://doi.org/10.3390/s22197268
7. Chen, W., Chen, S., Zhang, H., Wu, T.: A hybrid prediction model for type 2 diabetes using K-means and decision tree. In: Proceedings of IEEE International Conference on Software Engineering and Service Sciences ICSESS, vol. 2017(61272399), pp. 386–390 (2018). https://doi.org/10.1109/ICSESS.2017.8342938
8. Ramesh, J., Aburukba, R., Sagahyroon, A.: A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthc. Technol. Lett. **8**(3), 45–57 (2021). https://doi.org/10.1049/htl2.12010
9. Ahmed, U., et al.: Prediction of diabetes empowered with fused machine learning. IEEE Access **10**, 8529–8538 (2022). https://doi.org/10.1109/ACCESS.2022.3142097
10. Fitriyani, N.L., Syafrudin, M., Alfian, G., Rhee, J.: Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. IEEE Access **7**, 144777–144789 (2019). https://doi.org/10.1109/ACCESS.2019.2945129
11. Aamir, K.M., Sarfraz, L., Ramzan, M., Bilal, M., Shafi, J., Attique, M.: A fuzzy rule-based system for classification of diabetes. Sensors **21**(23) (2021). https://doi.org/10.3390/s21238095
12. El-Sappagh, S., Alonso, J.M., Ali, F., Ali, A., Jang, J.H., Kwak, K.S.: An ontology-based interpretable fuzzy decision support system for diabetes diagnosis. IEEE Access **6**, 37371–37394 (2018). https://doi.org/10.1109/ACCESS.2018.2852004
13. Kocbek, S., Kocbek, P., Gosak, L., Fijačko, N., Štiglic, G.: Extracting new temporal features to improve the interpretability of undiagnosed Type 2 diabetes Mellitus Prediction models. J. Pers. Med. **12**(3) (2022). https://doi.org/10.3390/jpm12030368
14. Hao, J., Luo, S., Pan, L.: Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes. Sci. Rep. **12**(1), 1–12 (2022). https://doi.org/10.1038/s41598-022-14143-8
15. Du, Y., Rafferty, A.R., McAuliffe, F.M., Wei, L., Mooney, C.: An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Sci. Rep. 12(1), 1–14 (2022). https://doi.org/10.1038/s41598-022-05112-2

16. El-Rashidy, N., ElSayed, N.E., El-Ghamry, A., Talaat, F.M.: Prediction of.gestational diabetes based on explainable deep learning and fog computing. Soft. Comput.Comput. **26**(21), 11435–11450 (2022). https://doi.org/10.1007/s00500-022-07420-1
17. Nagaraj, P., Muneeswaran, V., Dharanidharan, A., Balananthanan, K., Arunkumar, M., Rajkumar, C.: A prediction and recommendation system for diabetes mellitus using XAI-based lime explainer," International Conference on Sustainable Computing and Data Communication Systems ICSCDS 2022 - Proc.eedings, pp. 1472–1478 (2022). https://doi.org/10.1109/ICSCDS53736.2022.9760847
18. Tasin, I., Nabil, T.U., Islam, S., Khan, R.: Diabetes prediction using machine learning and explainable AI techniques. Healthc. Technol. Lett., 1–10 (2022). https://doi.org/10.1049/htl2.12039
19. Assegie, T.A., Karpagam, T., Mothukuri, R., Tulasi, R.L., Engidaye, M.F.: Extraction of human understandable insight from machine learning model for diabetes prediction. Bull. Electr. Eng. Informatics **11**(2), 1126–1133 (2022). https://doi.org/10.11591/eei.v11i2.3391
20. Zhao, X., Jiang, C.: The prediction of distant metastasis risk for male breast cancer patients based on an interpretable machine learning model. BMC Med. Inform. Decis. Mak.Decis. Mak. **23**(1), 74 (2023). https://doi.org/10.1186/s12911-023-02166-8
21. Technique, A.O., et al.: DAD-Net : Classification of Alzheimer ' s Disease Using Neural Network, pp. 1–21 (2022)
22. Noorunnahar, M., Chowdhury, F.A., Mila, A.H.: A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh, pp. 1–15 (2023). https://doi.org/10.1371/journal.pone.0283452
23. Nohara, Y., Matsumoto, K., Soejima, H., Nakashima, N.: Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput. Methods Programs Biomed. 214 (2022). http://10.0.3.248/j.cmpb.2021.106584
24. Kumari, S., Kumar, D., Mittal, M.: An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int. J. Cogn. Comput. Eng. **2**, 40–46 (2021). https://doi.org/10.1016/j.ijcce.2021.01.001