# *Feature importance* of using *explanaible artificial intelligence (xai)* and *machine learning* for diabetes disease classification

**Muhammad Maulana[1], Neny Sulistianingsih[2], Khasnur Hidjah[3]**
[1,2,3]Master of Computer Science Study Program, Faculty of Computer Science, Bumigora University
muhammad.maulana@universitasbumigora.ac.id

Submitted: June 19, 2025 ; Revised: June 28, 2025 ; accepted:October 30, 2025

***Abstract***

*Diabetes is one of the most significant global health problems in the modern era. This disease not only has a serious impact on the quality of life of sufferers, but also poses a great economic and social burden, both for individuals and the health service system as a whole. Therefore, early detection and effective treatment are very important in an effort to reduce the prevalence and negative impact of this disease. Therefore, the purpose of this study is to design a machine learning classification model that is able to identify feature importance with the help of the Explainable Artificial Intelligence (XAI) method in the case of diabetes. This model is expected to provide a clear interpretation of the most relevant features or symptoms, making it easier to detect whether a person has diabetes or not based on the symptoms that have been selected more optimally. The results of this study in the treatment or prediction of diabetes show that the results of the selection of LIME model features are higher than the accuracy of the SHAP model, where the highest is the LIME model which is processed using classification using the XGBoost algorithm with an accuracy of 98.47%, in addition to the LIME model using the Decisien Tree and Random Forest algorithms producing an accuracy of 91.97% and 91.49%, respectively. then the SHAP model using the XGBoost algorithm produced an accuracy of 0.9094%, the Decisien Tree algorithm produced an accuracy of 0.8059% and the Random Forest produced an accuracy of 88.46%, with the amount of data used as many as 70000 data, with 80% training data and 20% test data. The findings of this study are that the LIME feature selection combined with the XGBoost classification method has the best accuracy rate of 98.47% compared to the SHAP feature selection which is the same in combination with XGBoost with an accuracy of 90.94%. These findings also show that the selection of LIME features combined with the XGBoost algorithm is able to improve the interpretability of the model as well as maintain or even improve the accuracy of the predictions. This approach allows for the identification of the most relevant features more efficiently, thus supporting more informed decision-making in the data analysis process.*

***Keywords****: diabetes, Explanaibele Artificial Intellegence, Shap and Lime, kaggle.com.*

## 1. Introduction

Diabetes is a deadly disease caused by an increase in blood sugar in the body [1] According to the World Health Organization (WHO), diabetes is characterized by increased blood glucose levels that can lead to serious complications, such as damage to the heart, blood vessels, eyes, kidneys, and nerves [2]. An increase in glucose (blood sugar) levels that are higher than normal, caused by impaired insulin secretion [3][4][5] a condition when the body does not have enough to produce or use the hormone insulin that carries glucose into the body's cells [6]. The prediction that in 2045 it is also said that there will be an increase in diabetes to 629 million people [7] The number of diabetes cases worldwide has reached 463 million people and is expected to increase to 700 million people by 2045 [3]. In fact, Diabetes Mellitus is one of the fastest-growing life-threatening chronic diseases [8]. In order to meet these challenges, the development of accurate and efficient classification methods in diagnosing diabetes is of utmost importance [9] In recent decades, The number of diabetes sufferers in the world is increasing very rapidly [10], Indonesia is no exception [3].

This condition poses a major challenge in the health care system, especially in terms of early detection and appropriate treatment [11]. One crucial aspect of addressing this issue is the ability to diagnose diabetes quickly and accurately [12]. Therefore, the development of accurate and efficient classification methods in the diagnosis process is of great importance. by leveraging modern computing technologies and data-driven approaches, such as machine learning [13], it is expected that the classification process can be carried out automatically with a high level of accuracy [14]. So, the question

that remains is how to create a machine learning classification model that can find feature importance using explanaible artificial intelligence (XAI) in health data? This research will prove it, therefore, this study aims to examine how to create a machine learning classification model that can find feature importance using explanaable artificial intelligence (XAI) in diabetes.

The implication of this study is as a reference model in determining whether an individual is indicated to have diabetes or not, based on features that have been systematically selected. Moreover It can help medical personnel in decision-making, but it also improves the overall quality of healthcare. This research contributes in the form of a classification model that can be used as a reference to detect diabetes based on the features that have been optimally selected. The research conducted is not the same as some of the latest and related studies before, including research conducted by [11], [15], [16], [17], [18], [19], [20], [21], [22]and [23]. The explanatory structure in this manuscript is that the first subsection discusses previous research on the subject and its differences from the work in this paper, the second subsection describes the methods used in this study. Meanwhile, the third subpart explains the results of the research.

The study conducted by [15] with an average AUC of 0.864 for SHAP versus 0.839 for LIME and 50 repetitions, we found that the mean difference was statistically significant with a p-value of 0.0035. Study [16] Performance metrics to predict heart failure using a variety of algorithms reveal valuable insights into its effectiveness. Among the models assessed, Logistic Regression achieved the highest accuracy of 86.89%, which indicates its strong overall performance in correctly classifying the samples. The model also shows a commendable balance between precision and memory, with an accuracy of 85.71% and a memory of 90.91%. The results show that although Logistic Regression is effective in identifying positive cases, it also maintains a low rate of false positives. LightGBM and Random Forest showed competitive results, with an accuracy of 85.33% and 85.25%, respectively. LightGBM shows 87.74% accuracy and 86.92% memory, making it a powerful option for minimizing false negatives, although slightly less effective than Logistic Regression in overall accuracy.

Study[11] pre-processing data and comparing the prediction performance of six ML algorithms using cross-validation techniques. Based on these findings, a diabetes prediction model was developed using the XGBoost algorithm, which achieved high performance and interpretability These impressive results are obtained by fine-tuning the optimal parameters using the GridSearchCV technique. To improve the interpretability of the prediction results, the proposed model uses SHAP and LIME techniques for global and local explanations. These techniques serve as a powerful tool for doctors to assess the accuracy of prediction models. Additionally, it highlights the significant potential of this model's independent interpreter in providing visual explanations for any ML model.

Research conducted by [21] This study explores the development of a robust clinical decision support system for detecting gestational diabetes by leveraging different machine learning architectures. The process involves applying a combination of five data balancing techniques to improve detection performance. The best results were obtained from an ensemble model trained using the Synthetic Minority Oversampling Technique (SMOTE) method combined with Edited Nearest Neighbors (ENN), resulting in 96% accuracy, 95% sensitivity, and 99% precision. To improve model transparency, explainable AI (XAI) approaches were applied to the highest performing model, using libraries such as SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), Quantum lattice, Explain Like I'm 5 algorithm, Anchor, and Feature Importance. The study also analyzes the role of factors such as visceral fat deposition in influencing gestational diabetes risk prediction. Study [22] The models were trained using the Diabetes Health Indicators dataset, which has inherent class imbalance problems and produces biased predictions. This imbalance is overcome by using the technique of oversampling the minority with the majority weight. The experimental findings showed that LeDNet achieved an F1 score of 85%, recall of 84%, accuracy of 85%, and precision of 86%. Similarly, HiDenNet, achieved accuracy, F1 scores, recall, and precision of 85%, 86%, 86%, and 86%, respectively.

Study[23] Statistical tests such as Friedman and ANOVA were used to identify significant differences between FusionNet and other sub-networks. To improve interpretability, FusionNet was integrated with three XAI algorithms: SHAP, LIME, and Grad-CAM. The model showed high performance with 99.05% accuracy, 98.18% recall, 100% precision, 99.09% AUC, and 99.08% F1 score. These results indicate that FusionNet has the potential to be an effective diagnostic tool in differentiating DFU from healthy skin.

Study [17] The analysis that has been carried out obtained the XGBoost model with a combination of hyperparameters, namely N estimator = 394, max depth = 5, learning rate = 0.0174, subsample = 0.7075, and colsample bytree = 0.8855 and using only 12 variables is the best model in classifying suitable drinking water sources in West Java with accuracy values and F1-scores of 77.43% and 80.17%

so that clustering can be carried out based on SHAP values. Study [18] It found that the Gradient Boosting Classifier algorithm had an accuracy of 81% in predicting diabetes, higher than Random Forest (which had an accuracy of 79%) The results showed that the Gradient Boosting algorithm had an accuracy above 90% in most cases, making it a great choice for the classification of diabetic disease diseases. Study [19] with test results showing quite good accuracy, namely 81.11%, Precision 77.55% and Recall 66.67%. The Neural Network algorithm has an AUC value of 0.78, which means a good classification, which indicates that using it for the classification of diabetics has good accuracy. This research is in accordance with the achievement of the accuracy value target of > 80%, suggestions for further research by trying to increase the percentage of testing value to 0.25 (25%), with the hope of increasing the accuracy of > 81.11%, and then adding more hiden layers than before, namely as many as 3 hiden layers.

Study [20] From the results of the research that has been carried out, the Sup- port Vector Machine algorithm has a higher accuracy value compared to using the Naive Bayes algorithm. The accuracy value for the Support Vector Machine algorithm model was 78.04 percent and the accuracy value of the Naive Bayes algorithm was 76.98 percent. Based on this value, an accuracy difference of 1.06 percent was obtained. So it can be concluded that the application of the Support Vector Machine algorithm is able to produce a better level of accuracy in diagnosing diabetes disease compared to using the algorithm, namely the Naive Bayes algorithm.Meanwhile, Random Forest showed the highest recall of 96.97%, which highlights its ability to identify true positive cases, although its accuracy of 80% indicates a higher incidence of false positives compared to LightGBM. Both models also produce strong AUC scores, with Random Forest achieving an AUC of 0.9372, indicating excellent discriminative ability. The Support Vector Machine (SVM) achieved an accuracy of 83.61% but excelled in recalls with a score of 93.94%, highlighting its power in capturing actual cases of heart failure. However, it noted a lower AUC of 0.9026 compared to Random Forest, which suggests that while SVM is effective in identifying true positives, its overall discriminatory performance may be less robust. F1 scores for all models are relatively close, with Logistic Regression leading at 0.8824, followed by Random Forest at 0.8767.

Although various methods have been used in previous studies to address this problem, there are still a number of limitations, both in terms of methodology, the quality of the data used, and the validity of the results obtained. One of the main problems is the lack of attention to the application of the feature selection process, which has an impact on the low relevance and efficiency of the features in the classification process.

In this study, after feature selection using SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) methods, model retesting was performed using Random Forest (RF), Decision Tree (DT), and XGBoost algorithms. The aim was to evaluate the extent to which feature selection affects the performance of the classification model, especially in terms of accuracy. Comparison of accuracy before and after the feature selection process provides an overview of the effectiveness of the feature selection method in improving model performance.The evaluation methods used include accuracy, precision and recall. The results of this study in the treatment or prediction of diabetes show that the results of the selection of LIME model features are higher than the accuracy of the SHAP model where the highest is the LIME model which is processed using classification using the XGBoost algorithm with an accuracy of 98.47%, in addition to the LIME model using the Decisien Tree and Random Forest algorithms resulting in an accuracy of 91.97% and 91.49%, respectively. then the SHAP model using the XGBoost algorithm produced an accuracy of 90.94%, the Decisien Tree algorithm produced an accuracy of 80.59% and Random Forest produced an accuracy of 88.46%, with the amount of data used as many as 70000 data, with 80% training data and 20% test data.

## 2. Method

The sequence of research methodology processes in this article consists of five main processes, namely the dataset collection process, data preprocessing, the explanaable artificial intelligence (XAI) process, model classification, and finally model evaluation as shown in figure 1.

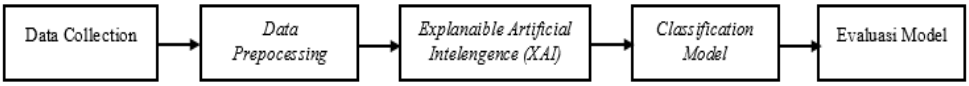


Figure 1 Research Stages

2.1. Data Collection

The primary data of this study is sourced from the public diabetes dataset available on the Kaggle.com platform and a total of 70,000 data entries, has 33 attributes and 13 classes. The attributes of this research dataset are shown in Table 1. The data used as machine learning training data in this study is data taken from Kaggle.com. Data mining learning is very helpful in systematically predicting whether a person will be diagnosed with diabetes based on existing attributes. Data from Kaggle.com site can be seen in Table 2. Machine learning that implements data mining methods has intelligence that can reveal hidden patterns in big data. Data mining is a method of determining certain patterns from a large dataset. To view the dataset, you can view it via the following Kaggle link (https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset). Data mining has many techniques, one of which is classification techniques. Classification is widely used to predict a class on a given label, by classifying data (building a model) based on a set of training and tests along with a value (class label) in the classification attribute.

Table 1 Attribute Dataset

| No. | Attributes/Variables | Description | No. | Attributes/Variables | Description |
|---|---|---|---|---|---|
| 1 | *Genetic Markers* | Genetic markers that indicate the risk of a specific disease (diabetes). | 18 | *Glucose Tolerance Test* | Tests to see the body's response to sugar. |
| 2 | *Autoantibodies* | Antibodies that attack the body itself, a sign of an autoimmune disease. | 19 | *History of PCOS* | History of polycystic ovary syndrome in women. |
| 3 | *Family History* | A family history of diabetes. | 20 | *Previous Gestational Diabetes* | History of diabetes during pregnancy. |
| 4 | *Environmental Factors* | Environmental influences, such as pollution or lifestyle. | 21 | *Pregnancy History* | History of previous pregnancies |
| 5 | *Insulin Levels* | The amount of insulin in the blood, is important for controlling blood sugar. | 22 | *Weight Gain During Pregnancy* | Weight gain during pregnancy |
| 6 | *Age* | A person's age. | 23 | *Pancreatic Health* | The health condition of the pancreas. |
| 7 | *BMI* (Body Mass Index) | Weight and height comparison to measure body health. | 24 | *Pulmonary Function* | Lung function capacity and efficiency. |
| 8 | *Physical Activity* | The level of exercise or physical activity. that have an impact on health. | 25 | *Cystic Fibrosis Diagnosis* | Diagnosis of cystic fibrosis disease. |
| 9 | *Dietary Habits* | Diet and the type of food consumed, including nutritional balance, that affect health. | 26 | *Steroid Use History* | History of steroid use. |
| 10 | *Blood Pressure* | Blood pressure, which can be an indicator of cardiovascular health. | 27 | *Genetic Testing* | Genetic testing for diabetes risk |
| 11 | *Cholesterol Levels* | The amount of cholesterol in the | 28 | *Neurological Assessments* | Examination of nerve |

| | | | | | |
|---|---|---|---|---|---|
| | | blood. | | | function. |
| 12 | *Waist Circumference* | Waist circumference, an indicator of body fat. | 29 | *Liver Function Tests* | Tests to check liver health. |
| 13 | *Blood Glucose Levels* | Glucose levels in the blood, an important indicator for diabetes. | 30 | *Digestive Enzyme Levels* | The amount of digestive enzymes in the body. |
| 14 | *Ethnicity* | A person's race or ethnicity | 31 | *Urine Test* | Urine tests to look for signs of diabetes. |
| 15 | *Socioeconomic Factors* | Socioeconomic status, such as income and education. | 32 | *Birth Weight* | Birth weight |
| 16 | *Smoking Status* | A person's smoking habit. | 33 | *Early Onset Symptoms* | Early symptoms of diabetes appear. |
| 17 | *Alcohol Consumption* | How often and how much alcohol is consumed. | | | |

Table 2 Class Datasets

| No | Types of Diseases | Description | No | Types of Diseases | Description |
|---|---|---|---|---|---|
| 1 | *Cystic Fibrosis-Related Diabetes (CFRD)* | Diabetes that occurs in people with cystic fibrosis, due to impaired pancreatic function. | 8 | *Steroid-Induced Diabetes* | Diabetes that occurs as a result of long-term steroid use. |
| 2 | *Gestational Diabetes* | Diabetes that appears during pregnancy. | 9 | *Type 1 Diabetes* | Autoimmune diabetes in which the body does not produce insulin at all. |
| 3 | *LADA (Latent Autoimmune Diabetes in Adults)* | Type 1 autoimmune diabetes that develops slowly in adults. | 10 | *Type 2 Diabetes* | Diabetes is a common occurrence, where the body does not use insulin properly. |
| 4 | *MODY (Maturity-Onset Diabetes of the Young)* | Genetic diabetes that usually appears at a young age. | 11 | *Type 3c Diabetes (Pancreatogenic Diabetes)* | Diabetes caused by damage to the pancreas, for example as a result of surgery or injury. |
| 5 | *Neonatal Diabetes Mellitus (NDM)* | Diabetes is rare that occurs in newborns or less than 6 months of age. | 12 | *Wolcott-Rallison Syndrome* | Genetic diabetes is a rare disease that usually occurs in infants and is accompanied by other health problems, such as bones or liver. |
| 6 | *Prediabetic* | The condition of blood sugar is higher than the normal limit, but not high enough to be | 13 | *Wolfram Syndrome* | A rare genetic disorder involving diabetes and nervous, vision, or |

| | | | |
|---|---|---|---|
| | | diagnosed with diabetes. | hearing problems. |
| 7 | *Secondary Diabetes* | Diabetes caused by other diseases or conditions, such as pancreatitis or hormonal disorders. | |

2.2. **Data Preprocessing**

Before testing the algorithm model, improvements are first made to the data to be processed, at this stage, checks are carried out on the value of the missing data because the dataset may contain incomplete data. The missing data values are replaced with the median values of each variable, so that each data on the dataset variable has a complete value[3].

2.3. **SHapley Additive exPlanations (SHAP)**

SHAP is based on cooperative game theory, which uses Shapley's values to quantitatively assess the contribution of each feature to the model's predictions. This method ensures a fair distribution of each feature's influence, allowing for a deeper understanding of how individual variables drive the model's decision-making process[24]. By providing insight into which features are most significant, SHAP allows researchers to gain a deeper understanding of the fundamental factors that influence predictions, especially in complex models. SHAP values are often represented in color-coded plots, where the intensity of color typically reflects the direction and magnitude of the impact of each feature[25]. For example, red can show a positive contribution to prediction (increasing the likelihood of a heart attack), while blue can show a negative impact (lowering the likelihood). These visual representations not only aid in understanding but also facilitate faster identification of important features, increase transparency, and build trust in model outputs[16]

2.4. **Local Interpretable Model-agnostic Explanations (LIME)**

Locally Interpretable Model-Agnostic Explanation is a post-hoc model-agnostic explanation technique that aims to estimate any black-box machine learning model with a locally interpretable model to explain each individual prediction[26]. Basically, LIME (Local Interpretable Model-agnostic Explanations) works locally, which means that it provides a specific explanation of each observation. Similar to SHAP, LIME focuses on individual interpretations of predicted outcomes[13]. This method works by building a simple local model around the observation you want to describ, using a sample of data that is similar to or in the vicinity of the observation point. Thus, LIME produces an interpretation that is local and easy to understand, even though the original model is complex and non-transparent[27]

2.5. **Classification Method**

The realization of data mining classification using the data mining method or machine learning algorithm involves two sets of data: the first is a dataset for training, and the second is a dataset for testing. Each set of items involves attributes and categories of each training attribute with a specific target value. This study used 80% training data and 20% testing data. This study uses the Random Forest, Decision Tree, and XGBoost algorithms using diabetes data sourced from Kaggle.com this algorithm combines several tree predictors or decision trees, where each tree depends on the value of a random vector that is sampled freely and evenly across all trees in the forest. The predicted results of the Random Forest are determined through the aggregation method: using voting for classification and average for regression. If the RF consists of N trees, then the prediction results can be formulated as seen in equation 1.

$$l(y) = \arg max_e \sum_{n=1}^{N} I(h_n(y) = c) \qquad (1)$$

$l(y)$ The final prediction label for input y, obtained from the majority voting result in the Random Forest Operator looking for $\arg max_e$ (class) that maximizes the number of occurrences in the prediction Number of decision trees (h$\sum_{n=1}^{N} I(h_n(y) = c)$n) in the forest (Random Forest) that classifies input y into class c, $h_n(y)$ : An nth decision tree model that generates a class prediction for y, An indicator function that is valued at 1 if the nth tree predicts class c, and 0 if not and N: The total number of decision trees in the Random Forest.$I(h_n(y) = c)$

The Decision Tree data mining method is a guided learning classification method that aims to find the optimal hyperfield by maximizing the distance or margin between data classes calculating the value of information gain for each attribute, it is necessary to calculate entropy first. The entropy value is used to determine how informative an attribute is. The entropy formula can be seen in the following equation [28]

$$Entropy(S) = \sum_{i=1}^{n} - pi\ log_2\ pi \qquad (2)$$

In the formula is the set of data being analyzed, the number of classes in the data and is the probability of the occurrence of the element of the class to- the way to calculate it is to multiply each probability by the base logarithm of 2 , then add up all the results of the multiplication and multiply it by -1 to ensure that the final value is positive. The entropy value indicates how random or diverse the data is; The lower the entropy (close to 0), the purer the data is (dominant in one class), while the higher entropy indicates a more even distribution of the data among several classes. After determining the entropy of each attribute, then determining $S\ npiipi\ piroot$ of the tree structure by counting *information gain* of each attribute. Attributes that have value *information gain* The largest *root* on the structure of the tree. The gain formula can be seen in the following equation [28]

$$Gain(S, A) = \sum_{f=1}^{n} \frac{|Si|}{|S|} + Entropy\ (Si) \qquad (3)$$

$S$ = case set = number of partitions attribute = number of cases on partition to = number of cases in = attributes$nA|Si|i|S|\ SA$

Extreme Gradient Boosting or XGBoost is one of the supervised learning methods used for classification and regression XGBoost works by combining various weak classifiers into a stronger model, through sequential training using the results of previous classification, called residuals or errors. The XGBoost formula introduces regularization in objective functions to prevent overfitting, with objective functions defined in the Equation [28]

$$0 \sum_{i=1}^{n} L(y_i, Fx_i) + \sum_{k=1}^{t} R(f_k) + C \qquad (4)$$

$L(y_i, Fx_i)$ The term regularization that functions to prevent $R(f_k)$*overfitting,* is formulated as stating the level of complexity indicating the number of leaves in the model indicating the penalty parameter refers to the output produced by each node.$\alpha H + \frac{1}{2}n + \sum_{j=1}^{H} wj^2\ aHnwj^2$

### 2.6. **Model Evaluation**

Confusion matrix is one of the methods used to evaluate the performance of a classification model. This matrix shows the number of correct and false predictions by comparing the model's predictions with the actual labels. These matrices typically have two main classes: Positive (P) and Negative (N). True Positive (TP) refers to the number of positive data points that are correctly predicted as positive, while False Negative (FN) indicates the number of positive data points that are incorrectly predicted as negative. False Positive (FP) is the number of negative data points that are incorrectly predicted as positive, and True Negative (TN) indicates the number of negative data points that are correctly predicted as negative. Accuracy reflects the percentage of correct predictions, Recall or Sensitivity measures the proportion of positive data correctly identified as positive, F1-Score represents the harmonic mean of Precision and Recall measures the proportion of negative data correctly identified as negative.

Table 3 Confusion Matrik

| | **Positive Predicted** | **Negative Predicted** |
|---|---|---|
| Positive Predicted | TP | FP |
| Negative Predicted | FN | TN |

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (5)$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (6)$$

$$F1 - Score = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \qquad (7)$$

## 3. Result and Discussion

This section presents the results of this research starting from the stages of data collection, data preprocessing, explanaible artificial intelligence (XAI), and model evaluation. The dataset used in this study is diabetes data taken from Kaggle.com. As presented in Table 3. The implementation process of diabetes classification uses Explanaible Artificial Intelligence (XAI) and uses 3 data mining algorithms implemented using python.

Table 3 Dataset

| No | Target | Autoantibodies | Environmental Factors | Insulin Levels | Age | BMI | | Birth Weight | Early Onset Symptoms |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 0 | 1 | 40 | 44 | 38 | …. | 2629 | 0 |
| 2 | 4 | 0 | 1 | 13 | 1 | 17 | …. | 1881 | 1 |
| 3 | 5 | 1 | 1 | 27 | 36 | 24 | …. | 3622 | 1 |
| 4 | 8 | 1 | 1 | 8 | 7 | 16 | …. | 3542 | 0 |
| 5 | 12 | 0 | 1 | 17 | 10 | 17 | …. | 1770 | 0 |
| 6 | 2 | 0 | 1 | 17 | 41 | 26 | …. | 3835 | 1 |
| 7 | 9 | 0 | 0 | 29 | 30 | 31 | …. | 4426 | 0 |
| … | …. | … | …. | … | … | … | … | … | … |
| … | …. | … | …. | … | … | … | … | … | … |
| 69999 | 10 | 0 | 1 | 30 | 60 | 32 | …. | 4335 | 0 |
| 70000 | 8 | 1 | 1 | 19 | 16 | 18 | …. | 2940 | 0 |

```
missing_per_column = df.Is null ().Sum ()
total_missing = missing percolumn.sum ()
print ("Jumlah missing value per kolom:")
print(missing_per_column)
print (f"\nTotal missing value di seluruh dataset: {total_missing}")
```

Figure 2 missing value check

```
Jumlah missing value per kolom:
Target                          0
Genetic Markers                 0
Autoantibodies                  0
Family History                  0
Environmental Factors           0
Insulin Levels                  0
Age                             0
BMI                             0
Physical Activity               0
Dietary Habits                  0
Blood Pressure                  0
Cholesterol Levels              0
Waist Circumference             0
Blood Glucose Levels            0
Ethnicity                       0
Socioeconomic Factors           0
Smoking Status                  0
Alcohol Consumption             0
Glucose Tolerance Test          0
History of PCOS                 0
Previous Gestational Diabetes   0
Pregnancy History               0
Weight Gain During Pregnancy    0
Pancreatic Health               0
Pulmonary Function              0
Cystic Fibrosis Diagnosis       0
Steroid Use History             0
Genetic Testing                 0
Neurological Assessments        0
Liver Function Tests            0
Digestive Enzyme Levels         0
Urine Test                      0
Birth Weight                    0
Early Onset Symptoms            0
dtype: int64

Total missing value di seluruh dataset: 0
```

Figure 3 Missing value check results

```
total_duplicates_before = df. duplicated (). sum ()
print (f"Total data duplikat sebelum
penghapusan: {total_duplicates_before}")
data = df. Drop duplicates ()
total_duplicates_after = data. duplicated ().sum ()
print (f"Total data duplikat setelah penghapusan:
{total_duplicates_after}")
```

Figure 4 Delete duplicate data

```
Total data duplikat sebelum penghapusan: 0
Total data duplikat setelah penghapusan: 0
```

Figure 5 of the results of deleting duplicate data

Figure 2 is the *source code* to check missing values and empty data before the process goes to the feature selection stage, then in figure 3 shows the results of checking the missing value. Figure 4 is the source code to delete duplicate data and figure 5 shows the results of deleting duplicate data.
In Figure 6 shown is a *SHAP* diagram that shows the contribution of each feature to the model's prediction with an output of f(x)=4. This diagram shows how much each feature adds or subtracts the final predictive value compared to the model's average value E[f(x)] = 6.21.


Figure 6 features of the importance of the random forest method

SHAP Force Plot or Summary Plot, which shows the contribution of each feature to the model's prediction. Each blue bar signifies a negative influence on the output value, while a red bar signifies a positive influence. The longer the bar, the greater the impact of the feature on the prediction results. As shown in figure 7.


Figure 7 fitur importance decision tree

Figure 8 refers to the SHAP Waterfall Plot, which shows the contribution of each feature to the model's prediction in one particular observation. This graph helps understand how each feature pushes the model's prediction up or down from the baseline.

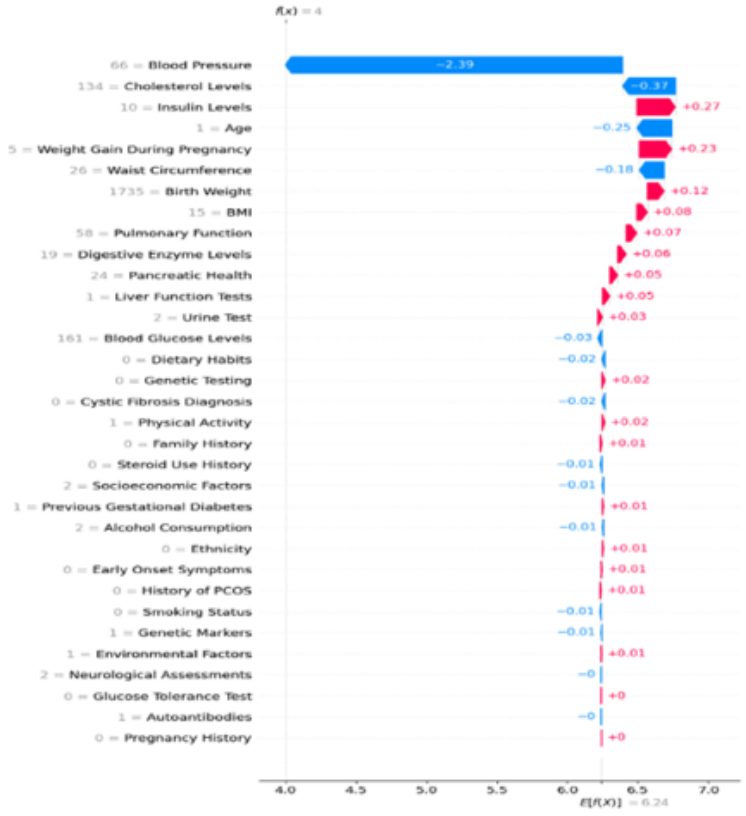


Figure 8 of the importance features of the XGBoost method

Figure 9 is the result of the interpretation of the prediction model using the LIME method in the case of diabetes prediction. At the top left is Prediction Probabilities, which are possible model predictions for various types of diabetes. The results showed that these patients were most likely to be diagnosed with Secondary Diabetes (62%) and Type 2 Diabetes (37%), while the chances for other types such as Steroid-Induced or Cystic Fibrosis-Related were almost zero. On the LIME Explanation chart, which shows the factors that contribute the most to the model's predictions. Each bar indicates how much the feature pushes the prediction in a certain direction.

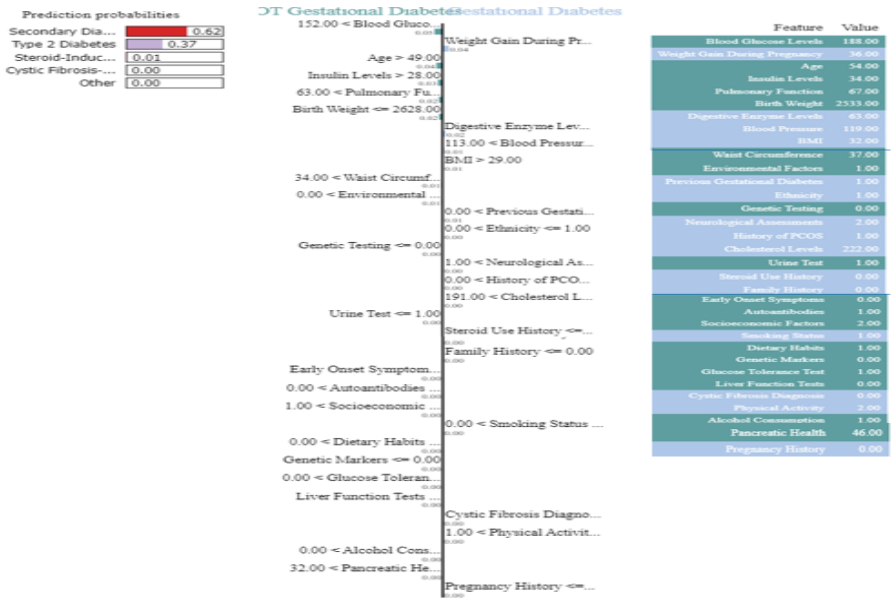


Figure 9 *of the importance of LIME Random Forest*

Figure 10 shows the results of the model's prediction interpretation using LIME for the classification of diabetes types. At the top left there is a Prediction Probabilities which shows that the model gives a 100% probability for predicting Type 2 Diabetes, while the probability for other types of diabetes such as Cystic Fibrosis, Gestational Diabetes, LADA, and Other is 0%.

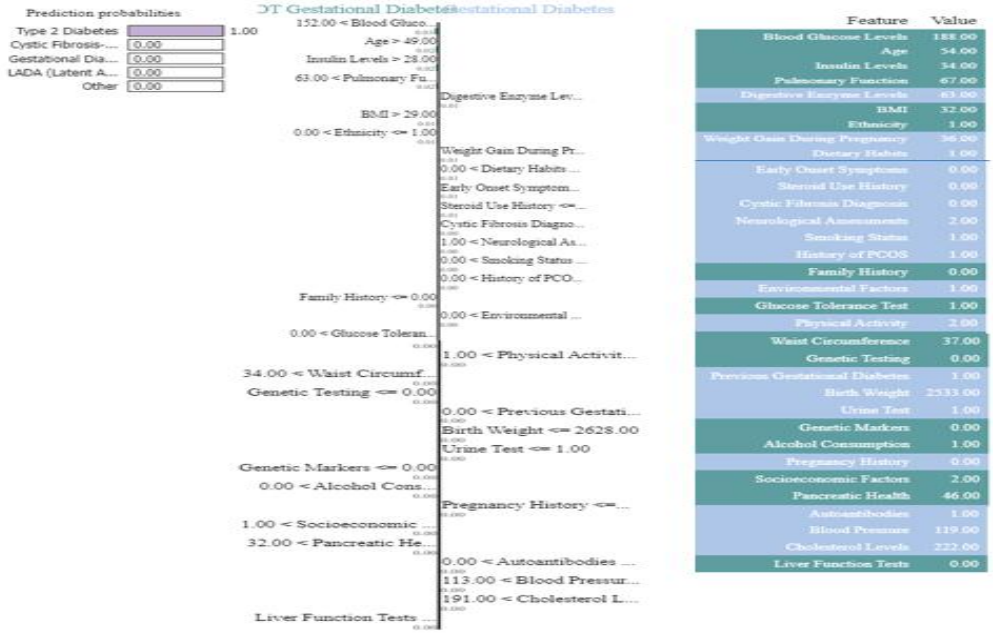


Figure 10 *of the importance of LIME Decision Tree features*

The results of the interpretation of the prediction of the diabetes classification model use the LIME (Local Interpretable Model-agnostic Explanations) method. At the top left, there is a Prediction Probabilities which shows that the model predicts a probability of Type 2 Diabetes of 61%, while the prediction for other types of diabetes such as Secondary Diabetes, Cystic Fibrosis, Gestational Diabetes, Steroid-Induced Diabetes, LADA, and Other is 0% as seen in figure 11

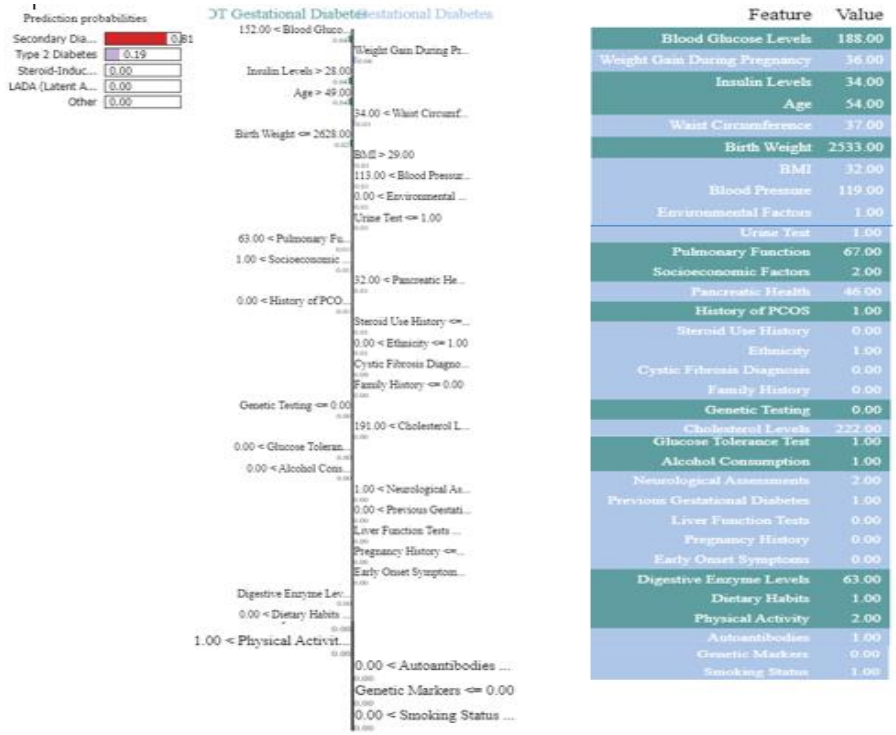


Figure 11 *of the importance of LIME XGBoost*

Table 4 SHAP and LIME Accuracy Results

| *Metode* | ***Original*** | ***Performance SHAP*** | | | ***Performance LIME*** | | |
|---|---|---|---|---|---|---|---|
| | | *Akurasi* | *Recall* | *F1-Score* | *Akurasi* | *Recall* | *F1-Score* |
| *Random Forest* | 90,00 | 88,46 | 82,11 | 88,38 | 91,49 | 84,57 | 91,40 |
| *Decisien Tree* | 86,00 | 80,59 | 63,69 | 77,82 | 91,97 | 97,53 | 92,85 |
| *XGBoost* | 90,00 | 90,94 | 89,86 | 91,38 | 98,47 | 99,20 | 98,58 |

Table 4 shows the results of the comparison of SHAP and LIME accuracy after feature selection, where the XGBoost method with LIME feature selection results in an accuracy of 98.47%, then the decision tree produces an accuracy of 91.97% and Random Forest produces an accuracy of 91.49%. Finally, the XGBoost method with SHAP feature selection produces an accuracy of 90.94%, then Random Forest produces an accuracy of 88.46% and Decisien Tree produces an accuracy of 80.59%.

## 4. Conclusion

The results of the study in the context of diabetes disease management and prediction showed that the model with feature selection using LIME produced higher accuracy compared to the model that used SHAP. The best model was obtained from the combination of LIME with the XGBoost algorithm, which achieved an accuracy of 98.47%. In addition, the LIME model combined with the Decision Tree and Random Forest algorithms yielded an accuracy of 91.97% and 91.49%, respectively. Meanwhile, the SHAP model using the XGBoost algorithm showed an accuracy of 90.94%, Decision Tree of 80.59%, and Random Forest of 88.46%. The novelty of this study's findings lies in the use of a combination of Decision Tree, Random Forest, and XGBoost algorithms with the LIME feature selection method, which is proven to provide higher accuracy than other methods. In addition, this research also has novelty in terms of the topics, objectives, and results studied, which until now have not been much or even discussed in depth by previous researchers. These findings have implications for increasing the effectiveness of early diagnosis by medical personnel, as well as opening up opportunities for the development of an Explainable AI system that can be used in digital health applications for independent and reliable early

detection of diseases. The weakness of this study is still limited to the use of narrow datasets so that the results are not necessarily representative of a wider population. In addition, only three algorithms were tested, and interpretation methods such as LIME and SHAP have limitations in consistency on complex data. Validation in a real clinical environment has also not been carried out.

**Reference**


[1] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 258–264, 2023, doi: 10.30630/joiv.7.1.1069.

[2] A. Brahmandjati, A. M. A. Rahim, and F. Asharudin, "Optimization of Diabetes Prediction with XGBoost Algorithm and Data Preprocessing Techniques," vol. 3, no. 1, pp. 116–125, 2024.

[3] Y. N. Marlim, L. Suryati, and N. Agustina, "Early Detection of Diabetes Using Machine Learning with Logistic Regression Algorithm," vol. 11, no. 2, pp. 88–96, 2022.

[4] Q. R. Cahyani, M. J. Finandi, J. Rianti, D. L. Arianti, and A. D. Pratama, "Diabetes Risk Risk Prediction using Logistic Regression Algorithm," vol. 1, no. 2, pp. 107–114, 2022, doi: 10.55123/jomlai.v1i2.598.

[5] D. C. P. Buani, "Early Detection of Diabetes Using the Random Forest Algorithm," *EVOLUTION J. Science and Management.*, vol. 12, no. 1, pp. 1–8, 2024, doi: 10.31294/evolution.v12i1.21005.

[6] A. W. Mucholladin, F. A. Bachtiar, and M. T. Furqon, "Classification of Diabetic Diseases using the Support Vector Machine Method," *J. Pengemb. Technology. Inf. and Computing Science.*, vol. 5, no. 2, pp. 622–633, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[7] A. Oktaviana, D. P. Wijaya, A. Pramuntadi, and D. Heksaputra, "Prediction of Type 2 Diabetes Mellitus Using the K-Nearest Neighbor (K-NN) Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 812–818, 2024, doi: 10.57152/malcom.v4i3.1268.

[8] J. Ginting, R. Ginting, and H. Hartono, "Detection and Prediction of Type 2 Diabetes Mellitus Using Machine Learning (Scooping Review)," *J. Prior Nursing.*, vol. 5, no. 2, pp. 93–105, 2022, doi:10.34012/jukep.v5i2.2671.

[9] M. Salsabil, N. Lutvi, and A. Eviyanti, "Implementation of Data Mining in Predicting Diabetes Using Random Forest and Xgboost Methods," *J. Ilm. Computing*, vol. 23, no. 1, pp. 51–58, 2024, doi: 10.32409/jikstik.23.1.3507.

[10] G. Quellec, H. Al Hajj, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," *Med. Image Anal.*, vol. 72, no. 2016, 2021, doi: 10.1016/j.media.2021.102118.

[11] Y. Zhao, J. K. Chaw, M. C. Ang, M. M. Daud, and L. Liu, "A Diabetes Prediction Model with Visualized Explainable Artificial Intelligence (XAI) Technology," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14322 LNCS, no. February 2025, pp. 648–661, 2024, doi: 10.1007/978-981-99-7339-2_52.

[12] A. Car *et al.*, "Covariance Structure Analysis on Health-Related Indicators in Elderly People at Home with a Focus on Subjective Health PerceptionsTitle," *Int. J. Technol.*, vol. 47, no. 1, p. 100950, 2023, [Online]. Available: https://doi.org/10.1016/j.tranpol.2019.01.002%0Ahttps://doi.org/10.1016/j.cstp.2023.100950%0Ahttps://doi.org/10.1016/j.geoforum.2021.04.007%0Ahttps://doi.org/10.1016/j.trd.2021.102816%0Ahttps://doi.org/10.1016/j.tra.2020.03.015%0Ahttps://doi.org/10.1016/j

[13] R. Ganguly and D. Singh, "Explainable Artificial Intelligence (XAI) for the Prediction of Diabetes Management: An Ensemble Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 7, pp. 158–163, 2023, doi: 10.14569/IJACSA.2023.0140717.

[14] I. Uysal, "Interpretable Diabetes Prediction using XAI in Healthcare Application," *J. Multidiscip. Dev.*, vol. 8, no. 1, pp. 20–38, 2023, [Online]. Available: https://www.researchgate.net/publication/376208551

[15] A. Gramegna and P. Giudici, "SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk," *Forehead. Artif. Intell.*, vol. 4, no. September, pp. 1–6, 2021, doi: 10.3389/frai.2021.752558.

[16] T. Öznacar and Z. T. Sertkaya, "Heart Failure Prediction: A Comparative Study of SHAP, LIME, and ICE in Machine Learning Models," *Int. J. Comput. Exp. Sci. Eng.*, vol. 10, no. 4, pp. 1885–1892, 2024, doi: 10.22399/ijcesen.589.

[17] K. Eligibility *et al.*, "Classification of Drinking Water Source Suitability in West Java Using XGBoost and Cluster Analysis Based on SHAP Values *," vol. 8, no. 2, pp. 202–214, 2024.

[18] A. T. Pratiwi, A. Barizi, M. I. Maulana, and P. Rosyani, "Systematic Literature Review of the Application of Gradient Boosting for the Classification of Type 2 Diabetes Disease," vol. 2, no. 3, pp. 454–458, 2024.

[19] S. Sutrisno and Jupron, "Analysis of Diabetes Classification with Neural Network Algorithm," *bit-Tech*, vol. 6, no. 3, pp. 303–310, 2024, doi:10.32877/bt.v6i3.1161.

[20] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediction of Diabetes Mellitus Using Support Vector Machine and Naive Bayes Methods," *Indones. J. Softw. Eng.*, vol. 7, no. 1, pp. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.

[21] V. Vivek Khanna *et al.*, "Explainable artificial intelligence-driven gestational diabetes mellitus prediction using clinical and laboratory markers," *Cogent Eng.*, vol. 11, no. 1, p., 2024, doi: 10.1080/23311916.2024.2330266.

[22] I. Shaheen, N. Javaid, N. Alrajeh, Y. Asim, and S. M. A. Akber, "New AI explained and validated deep learning approaches to accurately predict diabetes," *Med. Biol. Eng. Comput.*, 2025, doi: 10.1007/s11517-025-03338-6.

[23] S. Biswas, R. Mostafiz, M. S. Uddin, and B. K. Paul, "XAI-FusionNet: Diabetic foot ulcer detection based on multi-scale feature fusion with explainable artificial intelligence," *Heliyon*, vol. 10, no. 10, p. e31228, 2024, doi: 10.1016/j.heliyon.2024.e31228.

[24] H. K. Vasireddi, K. S. Devi, and G. N. V. R. Reddy, "DR-XAI: Explainable Deep Learning Model for Accurate Diabetic Retinopathy Severity Assessment," *Arab. J. Sci. Eng.*, vol. 49, no. 9, pp. 12899–12917, 2024, doi:10.1007/s13369-024-08836-7.

[25] Y. Du, A. R. Rafferty, F. M. McAuliffe, L. Wei, and C. Mooney, "An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi:10.1038/s41598-022-05112-2.

[26] M. Azad, M. F. K. Khan, and S. A. El-Ghany, "XAI-Enhanced Machine Learning for Obesity Risk Classification: A Stacking Approach with LIME Explanations," *IEEE Access*, vol. 13, no. December 2024, pp. 13847–13865, 2025, doi: 10.1109/ACCESS.2025.3530840.

[27] M. M. Hasan, "Understanding Model Predictions: A Comparative Analysis of SHAP and LIME on Various ML Algorithms," *J. Sci. Technol. Res.*, vol. 5, no. 1, pp. 17–26, 2024, doi: 10.59738/jstr.v5i1.23(17–26).eaqr5800.

[28] D. M. Pratiwi and L. Mufidah, "Comparison of Decision Tree Classifier and XGBoost Classifier Methods in Predicting Heart Disease," pp. 991–1000, 2024.