Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Performance Comparison of SelectKBest and Permutation Importance in Feature Selection for Diabetes Prediction

Nita Cahyani [1*], Rahmat Irsyada [2]
[1]Universitas Padjadjaran, Indonesia, [2]Politeknik Negeri Subang, Indonesia
[1]nita.cahyani@unpad.ac.id, [2]irsyada.rahmat@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 15-07-2025
Accepted: 20-07-2025
Published: 24-07-2025

Keywords:
Classification; diabetes; SelectKBest; Permutation Importance; Machine Learning

ABSTRACT

This study evaluates the effectiveness of two feature selection methods, namely the statistics-based SelectKBest and the model-based Permutation Importance, in improving the performance of classification algorithms for diabetes prediction. A dataset consisting of 17 clinical and demographic features was used to train 11 machine learning algorithms with two subsets of selected features. Performance evaluation used accuracy, precision, recall, F1-Score, ROC AUC, and training time. Based on the results, the SelectKBest method was able to improve the performance of Random Forest with an accuracy of 82.7%, a precision of 0.8, a recall of 0.5, and an F1-Score of 0.615. Meanwhile, the Permutation Importance method showed more consistent performance, with six models including Random Forest, K-Nearest Neighbors, and Quadratic Discriminant Analysis (QDA) achieving an accuracy of up to 86.2%. QDA stood out with the highest ROC AUC of 0.887, indicating better class detection capabilities. These findings underscore the superiority of Permutation Importance in selecting relevant and varied features, including demographic factors, thereby improving overall prediction accuracy. In practice, Random Forest with SelectKBest is recommended for applications requiring fast and interpretable models, while QDA and Gradient Boosting with Permutation Importance are recommended for those requiring high accuracy and sensitivity. This study strengthens the foundation for developing more accurate and applicable diabetes prediction models across various contexts.

INTRODUCTION

Diabetes mellitus is one of the most common metabolic conditions and has serious consequences, particularly in the elderly population, with a high risk of complications, including cerebrovascular accidents and premature death (Mehta et al., 2025), (Cantú-Brito et al., 2010). Diabetes mellitus continues to show an increasing prevalence worldwide and poses a significant public health burden (Gavin et al., 2024). Early detection and accurate classification are crucial for the management of diabetes mellitus to prevent serious complications (Gavin et al., 2024). Conventional diagnostic methods often suffer from limitations in sensitivity and the complexity of interpreting heterogeneous clinical data. Therefore, more objective methods capable of automatically processing complex data are needed. In this regard, the development of machine learning (ML) technology offers significant opportunities to improve the accuracy of early detection and predict the prognosis of chronic diseases, including diabetes (Delpino et al., 2022), (Chang et al., 2022).

Machine learning algorithms have been widely used to predict and classify diabetes due to their ability to handle complex data and numerous variables (Chang et al., 2022). Various ML classification algorithms, such as Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and Neural Network models like the Multi-layer Perceptron (MLP), have proven effective in modeling complex and non-linear medical data (Aliyu et al., 2024; Daza et al., 2024; Dharmarathne et al., 2024; C. Wang et al., 2025). The XGBoost algorithm, in particular, demonstrates superior performance due to its ability to handle high-dimensional data and robustness to feature collinearity (Cao et al., 2025) and (Ghasemieh et al., 2023).

One of the main challenges in implementing machine learning on clinical data is selecting the right features to improve model accuracy and interpretability (Khaire et al., 2022), (Zhou et al., 2024). However, the power of ML models is greatly influenced by the quality and relevance of the data features used. The presence of irrelevant or redundant features can reduce the accuracy of learning algorithms, making data dimensionality reduction through feature selection a crucial step (L. Wang et al., 2016). The SelectKBest method, which relies on univariate statistical tests such as ANOVA F-score, is an efficient technique for selecting the most informative features based on statistical correlation with the target variable (Wong et al., 2025), (Jiang et al., 2023). In contrast, Permutation Importance offers a model-agnostic approach by measuring the impact of random changes to each feature on model performance, providing interpretable insights into the contribution of each feature to a trained model (Cottin et al., 2024), (Allgaier et al., 2023).

Feature importance evaluation techniques, such as permutation feature importance, are also widely used in models to ensure that selected features truly impact classification performance (Dharmarathne et al., 2024). In several

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

medical studies, combining this feature selection method with ensemble models such as Random Forest and AdaBoost has shown improved predictive accuracy while maintaining model interpretability, which is crucial in clinical practice (Aliyu et al., 2024), (Daza et al., 2024), (Cottin et al., 2024), (Allgaier et al., 2023). Model readability is a critical concern, especially in medical contexts, making interpretable machine learning approaches indispensable for understanding risk factors for diabetes and its complications (C. Wang et al., 2025). This advantage underscores the importance of integrating appropriate feature selection techniques with algorithm choices to achieve reliable and informative predictions.

In general, machine learning has shown great potential in the early diagnosis and prediction of diabetes, as evidenced by numerous studies analyzing the performance of various machine learning methods in clinical contexts (S et al., 2024) and (Katiyar et al., 2024). Feature selection and ensemble classification techniques have also improved the accuracy and reliability of predictions for chronic diseases, including diabetes (Kumari et al., 2021). To increase user confidence in prediction results, feature selection methods based on variable contribution importance, such as Permutation Importance, have become significant in medical applications (Baniecki et al., 2025) and (Mienye et al., 2024). The permutation method for measuring feature importance provides more accurate corrections for features contributing to the model (Molnar et al., 2024). Ultimately, understanding the important variables and features in a dataset is fundamental to developing effective machine learning models for medical diagnosis.

Considering this background, this study aims to analyze and compare the performance of eleven classification algorithms frequently applied in the healthcare sector. These algorithms include Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Multi-layer Perceptron (MLP) (Cahyani, Irsyada, & Kartini, 2025; Cahyani, Irsyada, & Mahmuda, 2025; Cahyani, Putri, et al., 2025; Cahyani et al., 2022). The evaluation process was carried out using two different feature selection approaches, namely SelectKBest and Permutation Importance, to find the combination of models and features that can produce the best accuracy while remaining easy to interpret in predicting diabetes mellitus.

LITERATURE REVIEW

In recent years, the use of machine learning (ML) algorithms in healthcare has experienced significant development, particularly in predicting chronic diseases such as diabetes mellitus. Various algorithms, ranging from classical methods like Logistic Regression, Support Vector Machine (SVM), and Decision Trees to modern ensemble approaches like Random Forest and XGBoost, have proven capable of effectively handling the complexity of medical data (Chang et al., 2022; Daza et al., 2024; S et al., 2024). The application of this technology is becoming increasingly crucial given the increasing prevalence of diabetes globally and the importance of early detection, which can facilitate faster and more appropriate intervention (Gavin et al., 2024; Mehta et al., 2025).

The success of ML-based predictive systems depends heavily on the quality of the features used in model training. Redundant or irrelevant features can degrade classification performance and increase computational complexity (Khaire & Dhanalakshmi, 2022; Baniecki et al., 2025). Therefore, the feature selection process is a crucial part of the analysis pipeline, aiming to filter the most informative attributes. One widely used method is SelectKBest, which selects features based on univariate statistical scores, such as the F-test or chi-square, to identify the attributes that contribute most to the target variable (Jiang et al., 2023; Wong et al., 2025). This method is widely appreciated for its ease of implementation and speed of selection.

As a complement or alternative, the Permutation Importance technique offers a more flexible and contextual approach. Unlike traditional statistical methods, this approach measures feature contribution by randomizing feature values and calculating the resulting model performance degradation (Cottin et al., 2024). This method allows for assessing feature importance within the context of the trained model and captures the effects of interactions between features that univariate methods might miss (Molnar et al., 2024). Within the explainable AI (XAI) framework, Permutation Importance is often chosen because it bridges the gap between predictive accuracy and interpretability in medical contexts (Allgaier et al., 2023; Mienye et al., 2024).

In clinical contexts, improving accuracy is insufficient if the model is not well-interpretable. A study by Wang et al. (2025) showed that understanding important features not only increases user trust but also unlocks new insights into risk factors for diseases such as diabetes. Therefore, many studies integrate the feature selection process with model optimization techniques, including metaheuristic methods designed to automatically tune model parameters while addressing data imbalance, as demonstrated by Aliyu et al. (2024). As a result, the model is not only more stable but also capable of providing more precise predictions.

Furthermore, several studies have shown that the use of ensemble techniques such as stacking, bagging, and boosting provides additional benefits in medical classification, especially when combined with appropriate feature selection strategies (Ghasemieh et al., 2023; Kumari et al., 2021). The combination of an effective feature selection method and an appropriate classification algorithm has been shown to improve overall model performance, in terms of accuracy, sensitivity, and computational efficiency.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Based on these considerations, this study aims to evaluate and compare two feature selection approaches, SelectKBest and Permutation Importance, in improving the performance of eleven popular classification algorithms. The analysis focuses not only on predictive aspects such as accuracy and sensitivity, but also on time efficiency and ease of interpretation of results, to support real-world applications in diabetes mellitus prediction.

METHOD

Research Design

This research was designed using a quantitative experimental approach with a machine learning-based computational approach. The primary objective of this design was to evaluate and compare the performance of several classification algorithms in predicting diabetes, utilizing two feature selection techniques: SelectKBest and Permutation Importance. This design was chosen to test the effectiveness of the combination of feature selection methods and classification models in the context of binary classification of diabetes diagnoses.

Materials/Research Subjects

The research subjects were secondary data in .xlsx format, titled Data Diabetes.xlsx. This dataset consists of 18 attributes, including 17 input variables and one output variable, Diabetes_Y, which is a binary label (0 = no diabetes, 1 = diabetes). Input attributes included demographic variables (age, gender), physical indicators (BMI, obesity), and clinical symptoms (e.g., frequent thirst, blurred vision, itchy skin). This data was further processed to meet the requirements for application in the classification model.

Research Procedures

The research was conducted through the following stages:

1. Data Preprocessing
2. Data was read using the pandas library.
3. Categorical values were encoded into numeric form using LabelEncoder.
4. All features were scaled to a standard distribution using StandardScaler.
5. The data was divided into training data (70%) and testing data (30%) using a stratified split.
6. Feature Selection
7. Two approaches were used: SelectKBest with an ANOVA F-test score to select the 8 best features. Permutation Importance based on a Random Forest model with 10 iterations to determine the 8 most influential features.
8. Model Development and Training
9. Eleven classification algorithms were applied:
10. Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and Multi-layer Perceptron (MLP).
11. Each model was trained twice, using the features from SelectKBest and the features from Permutation Importance.
12. Model Evaluation
13. The model was tested on test data and assessed using performance metrics: accuracy, precision, recall, F1-score, and ROC AUC. Additionally, a confusion matrix and ROC curve were visualized to support model performance analysis.

Research Instruments

The primary instrument used in this study was a computer running the Python programming language. Several Python libraries were used, including pandas and numpy for data manipulation and processing, matplotlib and seaborn for graphic visualization and data analysis, and scikit-learn, which played a key role in preprocessing, feature selection, classification model training, and model performance evaluation. Furthermore, the time library was used to record the training time of each model. The entire experiment and analysis process was conducted in the interactive Jupyter Notebook environment, which facilitates efficient documentation and replication of the experiments.

Data Analysis Techniques

The analysis was conducted quantitatively based on the evaluation of classification metrics (Jiang et al., 2023) and (Cahyani et al., 2022):

1. Accuracy

   The proportion of correct classifications to the total test sample.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

The proportion of correct positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall (Sensitivity): The proportion of positive cases successfully detected.

$$Recall = \frac{TP}{TP + FN}$$

4. F1 Score
   The harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5. ROC AUC
   The area under the ROC (Receiver Operating Characteristic) curve, an indicator of the overall ability of a binary classification. A numerical representation that illustrates the trade-off between the True Positive Rate (Recall) and False Positive Rate at various thresholds. There is no single, simple formula; it is usually calculated numerically from the data.

The results of each model are compared to identify the best combination of model and feature selection method based on the highest F1 Score and ROC AUC values. Additional visualizations such as confusion matrix and ROC Curve are used to provide a visual overview of model performance.

RESULT

Dataset Description

In the initial stage, a description of the data used as the basis for modeling was carried out. Data were taken from Sosodoro Djatikusumo Bojonegoro Regional Hospital through the distribution of questionnaires to patients diagnosed with diabetes or non-diabetes with certain symptoms. The dataset consists of 17 independent variables (X) and 1 target variable (Y). The variables include demographic factors (age, gender), physical condition (BMI, obesity), and clinical symptoms relevant to diabetes. The data was then pre-processed in the form of encoding and normalization to prepare for the next stage of analysis.

Table 1. Summary of Descriptive Statistics and Distribution of Research Variables

Attribute	Data Type	Description
Age_X1	Numeric	Mean: 39.77; Std Dev: 14.2; Min: 19.0; Max: 87.0
BMI_X16	Numeric	Mean: 26.2; Std Dev: 12.7; Min: 16.65; Max: 104.08
Gender_X2	Categorical	Female: 59.38%; Male: 40.62%
Frequent excessive urination_X3	Categorical	No: 71.88%; Yes: 28.12%
Frequent thirst or urge to drink_X4	Categorical	No: 69.79%; Yes: 30.21%
Sudden weight loss_X5	Categorical	No: 87.5%; Yes: 12.5%
Frequent fatigue or lack of energy_X6	Categorical	No: 73.96%; Yes: 26.04%
Increased appetite beyond normal_X7	Categorical	No: 83.33%; Yes: 16.67%
Genital itching or irritation_X8	Categorical	No: 95.83%; Yes: 4.17%
Blurred vision (loss of visual sharpness)_X9	Categorical	No: 84.38%; Yes: 15.62%
Persistent skin itching_X10	Categorical	No: 90.62%; Yes: 9.38%
Frequent mood swings and irritability_X11	Categorical	No: 72.92%; Yes: 27.08%
Slow wound healing_X12	Categorical	No: 93.75%; Yes: 6.25%
Weakness in certain body parts_X13	Categorical	No: 88.54%; Yes: 11.46%
Muscle pain, joint stiffness, or tingling in arms or legs_X14	Categorical	No: 76.04%; Yes: 23.96%
Hair loss on the scalp_X15	Categorical	No: 77.08%; Yes: 22.92%
Obesity_X17	Categorical	Normal weight: 51.04%; Overweight: 30.21%; Obese: 10.42%; Underweight: 8.33%
Diabetes_Y	Categorical	No: 73.96%; Yes: 26.04%

Table 1 presents an overview of the demographic characteristics and clinical conditions of the respondents in this study. Numerical variables such as age and BMI exhibit considerable variability: respondents ranged in age from 19 to 87 years with a mean age approaching 40 years, while BMI values ranged widely from normal to extreme obesity,

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

indicating the presence of severe obesity in the sample data. Gender distribution showed a higher proportion of women (59.38%) than men (40.62%) in this sample. All categorical variables representing common clinical symptoms associated with diabetes exhibited varying distributions of Yes and No responses. For example, symptoms such as frequent urination with large volumes (28.12%), excessive thirst (30.21%), and sudden weight loss (12.5%) were experienced by a small proportion of respondents, reflecting the relative prevalence of these symptoms in the population.

Additionally, some subjects also experienced other symptoms such as weakness (26.04%), blurred vision (15.62%), mood swings (27.08%), and slow wound healing (6.25%). These conditions are clinically relevant as indications of complications that may be associated with diabetes. Obesity data show that more than 40% of respondents are overweight or obese, which is an important risk factor and is widely recognized as a trigger for diabetes. The target variable Diabetes_Y indicates that approximately 26.04% of the sample had been diagnosed with diabetes, marking the patient group that is the primary focus for analysis and development of the predictive model. Overall, this table not only provides an overview of the basic demographic and clinical characteristics of the study subjects but also highlights the heterogeneity of health conditions within the study population. This heterogeneity provides an important basis for interpreting the results of the statistical analysis and developing a diabetes classification model, which will be carried out in the next stage.

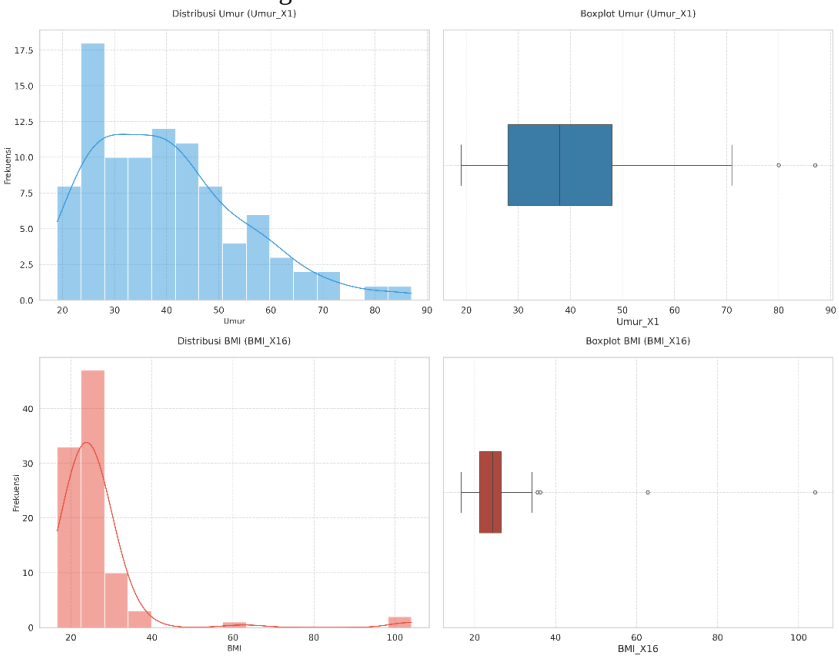

Fig. 1 Distribution and Summary Statistics of Respondents' Age and Body Mass Index (BMI)

Figure 1 shows a visualization of the data distribution for the two main numeric variables in the study: age (Age_X1) and body mass index (BMI_X16). On the top and bottom left, there are histograms complemented by density curves, depicting the frequency of occurrence of these variable values. The age distribution shows a slightly right-skewed pattern, with the majority of respondents concentrated in the 20-40 age range, indicating a majority in the young to middle-aged age group, while there are also a few older individuals, approaching 90 years of age. The BMI histogram displays the highest concentration of values in the normal to overweight categories, while also showing a number of very high extreme values, indicating cases of severe obesity in the sample. On the top and bottom right, boxplots provide summary statistics in the form of medians, interquartile ranges, and the presence of outliers. The boxplot for age shows a median of approximately 40 years, with a relatively tight interquartile range and few outliers in older ages. Meanwhile, the BMI boxplot shows a lower median with several extreme BMI outliers, which can be of particular clinical concern in the context of disease risk such as diabetes. Overall, this visualization provides a clear picture of the variation and distribution of demographic data and the key health conditions analyzed in this study.

Feature Selection

Feature selection was performed to identify the independent variables that most contribute to predicting diabetes status. Two methods were compared: SelectKBest (based on ANOVA F-Score) and Permutation Importance (based on Random Forest).

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Selected features with SelectKBest: 8 dominant features consisting of clinical symptoms such as frequent urination, excessive thirst, weight loss, limb weakness, and tingling sensations. Selected features with Permutation Importance: 8 dominant features including demographic variables (age, gender), BMI, obesity, and several similar clinical symptoms.

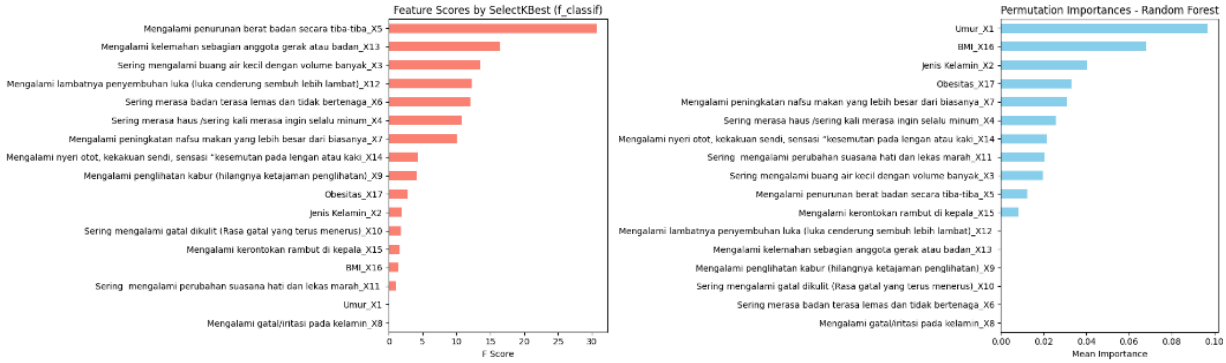

Fig. 2 Feature Scores from Feature Selection using SelectKBest and Permutation Importance

Figure 2 shows the results of feature selection using two different methods to determine the independent variables most contributing to predicting diabetes status. On the left, the bar graph displays feature scores based on the SelectKBest method using ANOVA F-Score, which identifies eight dominant features, primarily clinical symptoms, such as frequent urination with large volumes, excessive thirst, weight loss, limb weakness, and tingling sensations in the hands and feet. Meanwhile, the right panel shows the results of the Random Forest-based Permutation Importance method, which highlights eight key features that include a combination of demographic variables such as age and gender, body mass index (BMI), obesity status, and several similar clinical symptoms also identified in the SelectKBest method. Comparing these two graphs provides a comprehensive overview of the most relevant perspective variables in predicting diabetes, while also demonstrating the strength of various features from a statistical and machine learning. This information is crucial for selecting variables for subsequent prediction models in the study.

Model Evaluation

After feature selection, the model was trained using 11 different classification algorithms. Performance assessment was based on the following metrics: Accuracy, Precision, Recall, F1-Score, and ROC AUC.

Table 2. Evaluation of SelectKBest on Diabetes Prediction

No	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Training Time (s)
1	Random Forest	0.827	0.800	0.500	0.615	0.786	0.187
2	Decision Tree	0.793	0.667	0.500	0.571	0.667	0.002
3	AdaBoost	0.793	0.667	0.500	0.571	0.818	0.100
4	Naive Bayes	0.793	0.667	0.500	0.571	0.818	0.001
5	Gradient Boosting	0.793	0.667	0.500	0.571	0.738	0.129
6	Logistic Regression	0.759	0.571	0.500	0.533	0.732	0.016
7	Linear Discriminant Analysis	0.759	0.571	0.500	0.533	0.726	0.002
8	K-Nearest Neighbors	0.793	0.750	0.375	0.500	0.673	0.001
9	Multi-layer Perceptron	0.724	0.500	0.250	0.333	0.631	0.283
10	Support Vector Machine	0.759	1.000	0.125	0.222	0.762	0.004
11	Quadratic Discriminant Analysis	0.724	0.500	0.125	0.200	0.682	0.002

Table 2 displays the evaluation results of various classification models after feature selection using the SelectKBest method for diabetes prediction. Overall, Random Forest performed best compared to other models. This is evident from its highest accuracy value of 0.827, followed by several other models, including Decision Tree, AdaBoost, Naive Bayes, Gradient Boosting, and K-Nearest Neighbors, which all achieved uniform accuracy of 0.793. In terms of precision, the Support Vector Machine (SVM) achieved a perfect score of 1.000, but its weakness was seen in its recall, which was only 0.125, resulting in a low F1-score (0.222). This means that the SVM was too strict in classifying data as positive, so despite its accurate positive predictions, only a small number of positive cases were detected. Conversely, Random Forest demonstrated a much better balance between precision (0.800) and recall (0.500), resulting in a higher F1-score (0.615) than the other models.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Based on Recall values, most models successfully recognized at least half of the positive cases. Random Forest, Decision Tree, AdaBoost, Naive Bayes, Gradient Boosting, Logistic Regression, and Linear Discriminant Analysis all recorded a Recall of 0.500. Meanwhile, K-Nearest Neighbors performed slightly lower, with a Recall of 0.375. Multi-Layer Perceptron, SVM, and Quadratic Discriminant Analysis showed the lowest Recall results. Looking at the F1-Score, which balances Precision and Recall, Random Forest remains superior with a score of 0.615, followed by Decision Tree, AdaBoost, Naive Bayes, and Gradient Boosting at 0.571, and Logistic Regression and Linear Discriminant Analysis at 0.533.

In the ROC AUC metric, which assesses a model's ability to distinguish between positive and negative classes at various thresholds, the highest scores were recorded by AdaBoost and Naive Bayes (0.818). Random Forest performed slightly lower with an ROC AUC of 0.786, but was still considered more stable overall, producing a more balanced combination of accuracy and F1-score metrics. In terms of training time, Random Forest did require longer (0.187 seconds) than simpler models like Naive Bayes (0.001 seconds) or Decision Tree (0.002 seconds). This time difference is still relatively small and is worth compensating for in order to achieve more stable and accurate model performance.

Table 3. Evaluation of Permutation Importance in Diabetes Prediction

No	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Training Time (s)
1	**Random Forest**	**0.862**	**0.750**	**0.750**	**0.750**	**0.783**	**0.176**
2	K-Nearest Neighbors	0.862	0.833	0.625	0.714	0.851	0.002
3	Logistic Regression	0.862	0.833	0.625	0.714	0.768	0.004
4	Gradient Boosting	0.862	0.833	0.625	0.714	0.798	0.140
5	Decision Tree	0.862	0.833	0.625	0.714	0.789	0.003
6	Quadratic Discriminant Analysis	0.862	0.833	0.625	0.714	0.887	0.002
7	Linear Discriminant Analysis	0.828	0.714	0.625	0.667	0.768	0.003
8	Naive Bayes	0.759	0.545	0.750	0.632	0.798	0.002
9	Support Vector Machine	0.828	0.800	0.500	0.615	0.804	0.005
10	Multi-layer Perceptron	0.828	0.800	0.500	0.615	0.786	0.535
11	AdaBoost	0.759	0.571	0.500	0.533	0.833	0.287

Table 3 shows the evaluation results of the classification models using feature selection using the Permutation Importance method. This table shows that several models demonstrated very competitive and relatively stable performance with high accuracy values. The highest accuracy values were achieved by six models: Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis, all achieving an accuracy of 0.862. This indicates that features selected using Permutation Importance generally improve the model's predictive power.

In the Precision metric, the best performance was achieved by K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis, with a precision value of 0.833, slightly higher than Random Forest's 0.750. This figure indicates that the majority of positive predictions are truly relevant or match the original class. The recall of the best models (Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis) was consistently between 0.625 and 0.750, indicating that most positive cases were detected. Interestingly, Naive Bayes actually had a higher recall value (0.750) than Random Forest, but its precision was relatively lower (0.545), so its performance balance was still better than the other models.

The F1-Score metric, which balances precision and recall, confirmed that Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis were the most stable with F1-Scores of 0.714 and 0.750, outperforming the other models. This F1-Score reflects the model's ability to maintain a balance of positive case detection without many false predictions. The highest ROC AUC value was recorded by Quadratic Discriminant Analysis (0.887), indicating excellent potential for separating positive and negative classes at various prediction thresholds. Random Forest also performed solidly with an ROC AUC of 0.783, followed by Gradient Boosting (0.798) and AdaBoost (0.833). In terms of training time, Random Forest did require slightly longer (0.176 seconds) than linear models like Logistic Regression (0.004 seconds) or K-Nearest Neighbors (0.002 seconds). However, the training time was generally efficient and proportional to the quality of the resulting predictions.

Based on these results, Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis can be recommended as the best-performing models for diabetes prediction using features selected using the Permutation Importance method. However, when considering the balance of accuracy, F1-Score, and ROC AUC, Quadratic Discriminant Analysis (QDA) emerged as the strongest candidate due to the combination of the highest ROC AUC value (0.887) and balanced accuracy and F1-Score performance.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Results Visualization

This section presents visualizations to support the numerical interpretation of the classification model evaluation results for diabetes prediction. The main visualizations are the Receiver Operating Characteristic (ROC) Curve and confusion matrix heatmap, which illustrate the model's ability to distinguish classes and the pattern of correct and incorrectpredictions in the test data.

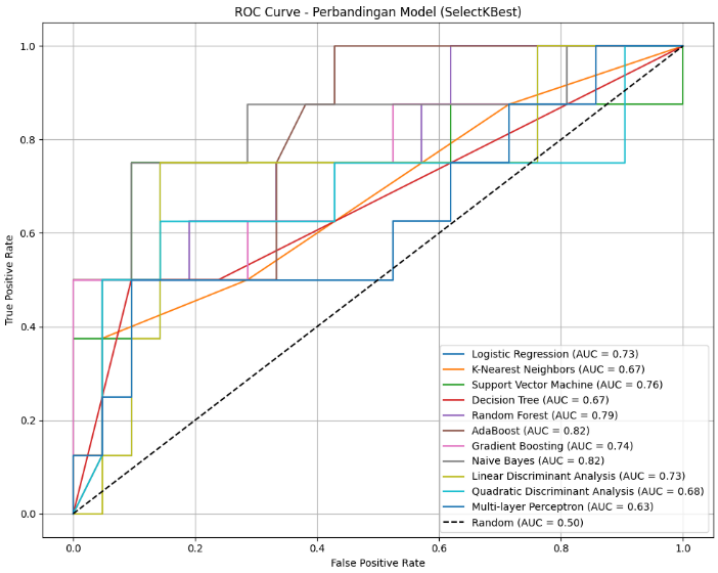

Fig. 3 SelectKBest ROC curve graph

Figure 3 displays the Receiver Operating Characteristic (ROC) Curve for each classification algorithm based on the features selected with SelectKBest. In this graph, the AdaBoost and Naive Bayes models obtained the highest AUC value of 0.82, followed by Random Forest with an AUC of 0.79, Support Vector Machine (SVM) with an AUC of 0.76, and Gradient Boosting with 0.74. The Logistic Regression and Linear Discriminant Analysis models each had an AUC of 0.73, while K-Nearest Neighbors (KNN) was recorded at 0.67, Decision Tree at 0.67, Quadratic Discriminant Analysis (QDA) at 0.68, and Multi-layer Perceptron (MLP) with an AUC of 0.63. The black diagonal line represents random predictions (AUC = 0.50). This pattern supports the evaluation results in Table 2, that the combination of SelectKBest tends to make AdaBoost, Naive Bayes, and Random Forest the models with the highest discriminatory power.

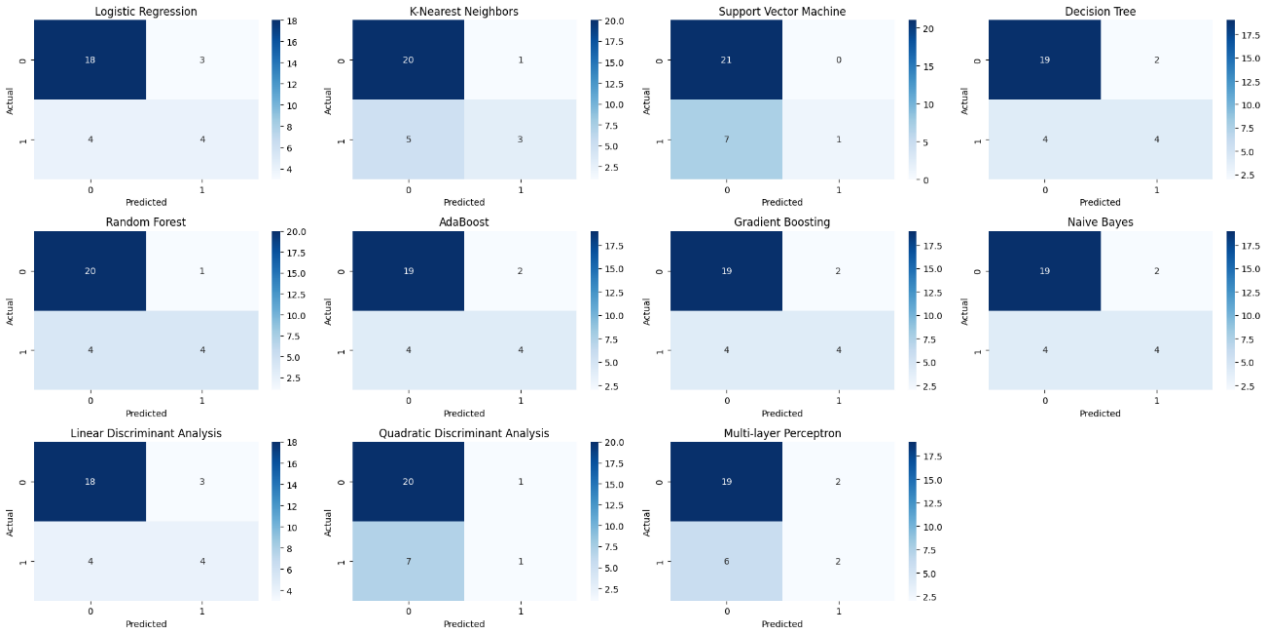

Fig. 4 heatmap confusion matrix of each SelectKBest model in Diabetes Prediction

Figure 4 shows a confusion matrix heatmap from the results of feature selection using the SelectKBest method, which illustrates the number of correct and incorrect predictions in each model. For example, Random Forest

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

successfully predicted 20 negative data correctly (True Negative/TN), only incorrectly classified 1 negative data as positive (False Positive/FP), was able to correctly recognize 4 positive data (True Positive/TP), but failed to detect 4 other positive data and therefore counted as incorrect (False Negative/FN). This pattern supports the highest accuracy and F1-Score performance of Random Forest in Table 2. The AdaBoost, Decision Tree, and Gradient Boosting models also have a similar pattern, on average predicting 19 negative data correctly, 2 negative data were misread as positive, 4 positive data were correctly detected, and 4 positive data were missed. The Support Vector Machine (SVM) looks different because it is able to predict all negative data without error (21 correct predictions) but only detected 1 positive data, while the remaining 7 were not detected, so the Precision of this model is high but its Recall is low. This heatmap visualization is useful for assessing the balance between correct and incorrect predictions in each model. Generally, models with high AUC values tend to predict positive data more evenly and have fewer errors. Conversely, models with lower AUCs, such as the Multi-Layer Perceptron (MLP) with an AUC of 0.63, have a higher number of incorrect positive data detections (6 positive data missed), and only 2 positive data were correctly predicted.

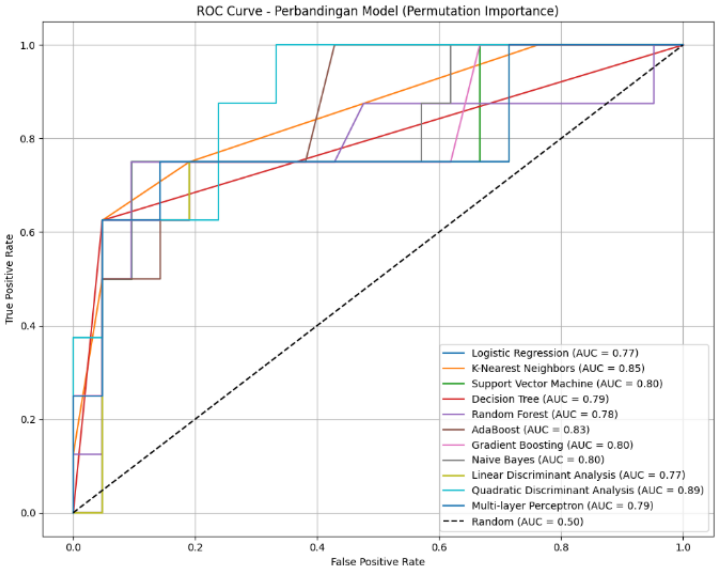

Fig. 5 Permutation Importance ROC curve graph

Figure 5 shows the same ROC curves but for models with features selected using the Permutation Importance method. It can be seen that several models experience an increase in AUC compared to SelectKBest. For example, Quadratic Discriminant Analysis (QDA) stands out with an AUC of 0.89, followed by Naive Bayes (AUC = 0.80), Gradient Boosting (AUC = 0.80), and K-Nearest Neighbors (KNN) (AUC = 0.85). The curve lines moving further away from the black diagonal indicate that the model's ability to distinguish between classes improves. These results demonstrate that feature selection using Permutation Importance can improve the discriminatory power of several algorithms, supporting the numerical evaluation in Table 3.

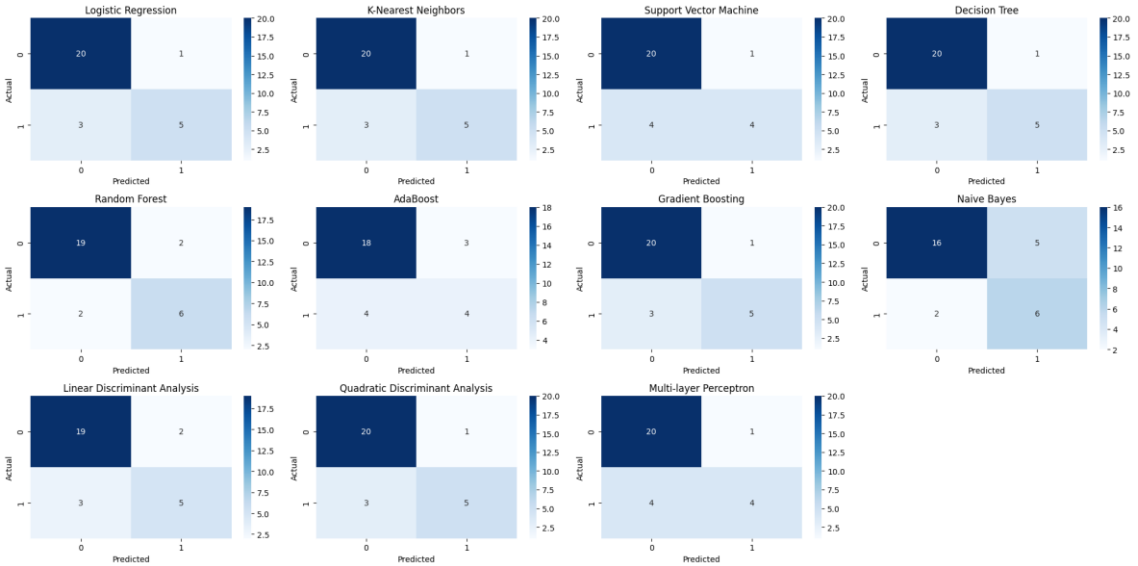

Fig. 6 Heatmap confusion matrix of each model Permutation Importance in Diabetes Prediction

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Figure 6 displays the confusion matrix heatmap from the evaluation using Permutation Importance. For example, Random Forest correctly identified 19 negative data (TN), made 2 false positive predictions (FP), correctly detected 6 positive data (TP), and missed 2 positive data (FN). In K-Nearest Neighbors, a similar pattern was seen with 20 TN, 1 FP, 3 TP, and 5 FN. Logistic Regression also recorded 20 TN, 1 FP, 3 TP, and 5 FN. Meanwhile, Decision Tree and Gradient Boosting showed a similar pattern, with an average of 20 negative data correctly classified and True Positives in the range of 3 to 5. Quadratic Discriminant Analysis (QDA) also showed 20 TN, 1 FP, 5 TP, and 3 FN, indicating that this model is quite good at detecting positive data in this scenario.

DISCUSSION

This section compares the performance of two feature selection techniques, SelectKBest and Permutation Importance, and their practical implications. The results show that SelectKBest tends to select variables directly related to clinical symptoms of diabetes, such as frequent urination, excessive thirst, or sudden weight loss. Conversely, Permutation Importance successfully captures the influence of demographic variables such as age, gender, body mass index (BMI), and obesity, thus providing a broader range of information for the model. In terms of performance, Random Forest on SelectKBest achieved the highest accuracy of 82.7% with an F1-score of 0.615. Meanwhile, Gradient Boosting on Permutation Importance excelled with an accuracy of 86.2% and an F1-score of 0.714. Several other models, such as Quadratic Discriminant Analysis, Logistic Regression, and K-Nearest Neighbors, also demonstrated consistent results on Permutation Importance. Relatively high ROC, AUC, and Precision values support the superiority of these models for diabetes prediction. Based on these findings, the method selection can be tailored to user needs. If training speed and simple interpretation are prioritized, then Random Forest or Naive Bayes with SelectKBest are still suitable. However, if the primary goal is maximum prediction accuracy, then Gradient Boosting with the Permutation Importance feature is recommended. These results emphasize the importance of adapting methods and algorithms to the data context and analysis objectives.

CONCLUSION

This study explored the comparison of two feature selection techniques, namely SelectKBest and Permutation Importance, to improve diabetes prediction performance by involving 11 classification models. Based on the analysis results, the SelectKBest method showed that Random Forest emerged as the most reliable algorithm with an accuracy of 82.7%, a precision of 0.800, a recall of 0.500, and an F1-Score of 0.615. Several other models such as Decision Tree, AdaBoost, Naive Bayes, and Gradient Boosting also recorded quite good accuracy in the range of 79.3%, although their F1-Scores were slightly below Random Forest. Meanwhile, the highest ROC AUC in SelectKBest was achieved by AdaBoost and Naive Bayes (0.818), indicating good class separation ability, although the overall balance still favors Random Forest. On the other hand, in the Permutation Importance method, model performance appears more evenly distributed, with six models (Random Forest, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, Decision Tree, and Quadratic Discriminant Analysis/QDA) all achieving the highest accuracy of 86.2%. Of this group, Random Forest demonstrated stable evaluation metrics (precision 0.750, recall 0.750, F1-Score 0.750, ROC AUC 0.783), while QDA attracted attention because it recorded the highest ROC AUC of 0.887 with precision 0.833, recall 0.625, and F1-Score 0.714, demonstrating its strength in separating positive and negative classes. The overall results indicate that Permutation Importance successfully boosted prediction consistency by utilizing a wider variety of features, including demographic aspects. Thus, Random Forest with SelectKBest is suitable for use when a fast and easy-to-interpret model is needed. Conversely, if the focus is on maximum accuracy and optimal detection capability, then QDA, or Gradient Boosting with Permutation Importance, is more recommended. These findings are expected to support the development of more precise, effective, and practical diabetes early detection models for application in various contexts.

REFERENCES

Aliyu, H. A., Muritala, I. O., Bello-Salau, H., Mohammed, S., Onumanyi, A. J., & Ajayi, O.-O. (2024). Optimizing machine learning algorithms for diabetes data: A metaheuristic approach to balancing and tuning classifiers parameters. Franklin Open, 8, 100153. doi: https://doi.org/10.1016/j.fraope.2024.100153

Allgaier, J., Mulansky, L., Draelos, R. L., & Pryss, R. (2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, 143, 102616. doi: https://doi.org/10.1016/j.artmed.2023.102616

Baniecki, H., Sobieski, B., Szatkowski, P., Bombinski, P., & Biecek, P. (2025). Interpretable machine learning for time-to-event prediction in medicine and healthcare. Artificial Intelligence in Medicine, 159, 103026. doi: https://doi.org/10.1016/j.artmed.2024.103026

Cahyani, N., Irsyada, R., & Kartini, A. Y. (2025). Implementasi Machine Learning Model sebagai Sistem Prediksi Penyakit Breast Cancer. Digital Transformation Technology, 4(2), 1112–1120. doi: 10.47709/digitech.v4i2.5209

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Cahyani, N., Irsyada, R., & Mahmuda, R. (2025). Penerapan Algoritma Neural Network untuk Klasifikasi Diabetes Mellitus: Perbandingan Backpropagation dan Resillient Backpropagation. Digital Transformation Technology, 4(2), 1067–1074. doi: 10.47709/digitech.v4i2.5208

Cahyani, N., & Kartini, A. Y. (2022). DLNN dan BPNN-GA Pada Prediksi Penyakit Diabetes di Bojonegoro. Journal of Mathematics Education and Science, 6(1), 1–9. doi: 10.32665/james.v6i1.416

Cahyani, N., Putri, W. A., & Irsyada, R. (2025). Improving Multiclass Rainfall Prediction with Multilayer Perceptron and SMOTE: Addressing Class Imbalance Challenges. Brilliance: Research of Artificial Intelligence, 4(2), 901–908. doi: 10.47709/brilliance.v4i2.5203

Cantú-Brito, C., Mimenza-Alvarado, A., & Sánchez-Hernández, J. J. (2010). Diabetes mellitus and aging as a risk factor for cerebral vascular disease: epidemiology, pathophysiology and prevention. Revista de Investigacion Clinica, 62(4), 333–342.

Cao, Z., Wu, X., Wu, B., Zhang, Z., & Sun, J. (2025). Combining UDT with XGBoost to identify the geographical origin of black beans by near-infrared spectroscopy. Current Research in Food Science, 11, 101131. doi: https://doi.org/10.1016/j.crfs.2025.101131

Chang, V., Ganatra, M. A., Hall, K., Golightly, L., & Xu, Q. A. (2022). An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. Healthcare Analytics, 2, 100118. doi: https://doi.org/10.1016/j.health.2022.100118

Cottin, A., Zulian, M., Pécuchet, N., Guilloux, A., & Katsahian, S. (2024). MS-CPFI: A model-agnostic Counterfactual Perturbation Feature Importance algorithm for interpreting black-box Multi-State models. Artificial Intelligence in Medicine, 147, 102741. doi: https://doi.org/10.1016/j.artmed.2023.102741

Daza, A., Ponce Sánchez, C. F., Apaza-Perez, G., Pinto, J., & Zavaleta Ramos, K. (2024). Stacking ensemble approach to diagnosing the disease of diabetes. Informatics in Medicine Unlocked, 44, 101427. doi: https://doi.org/10.1016/j.imu.2023.101427

Delpino, F. M., Costa, Â. K., Farias, S. R., Chiavegatto Filho, A. D. P., Arcêncio, R. A., & Nunes, B. P. (2022). Machine learning for predicting chronic diseases: a systematic review. Public Health, 205, 14–25. doi: https://doi.org/10.1016/j.puhe.2022.01.007

Dharmarathne, G., Jayasinghe, T. N., Bogahawaththa, M., Meddage, D. P. P., & Rathnayake, U. (2024). A novel machine learning approach for diagnosing diabetes with a self-explainable interface. Healthcare Analytics, 5, 100301. doi: https://doi.org/10.1016/j.health.2024.100301

Gavin, J. R., Rodbard, H. W., Battelino, T., Brosius, F., Ceriello, A., Cosentino, F., Giorgino, F., Green, J., Ji, L., Kellerer, M., Koob, S., Kosiborod, M., Lalic, N., Marx, N., Prashant Nedungadi, T., Parkin, C. G., Topsever, P., Rydén, L., Huey-Herng Sheu, W., … Schnell, O. (2024). Disparities in prevalence and treatment of diabetes, cardiovascular and chronic kidney diseases – Recommendations from the taskforce of the guideline workshop. Diabetes Research and Clinical Practice, 211, 111666. doi: https://doi.org/10.1016/j.diabres.2024.111666

Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. Decision Analytics Journal, 7, 100242. doi: https://doi.org/10.1016/j.dajour.2023.100242

Jiang, L., Xia, Z., Zhu, R., Gong, H., Wang, J., Li, J., & Wang, L. (2023). Diabetes risk prediction model based on community follow-up data using machine learning. Preventive Medicine Reports, 35, 102358. doi: https://doi.org/10.1016/j.pmedr.2023.102358

Katiyar, N., Thakur, H. K., & Ghatak, A. (2024). Recent advancements using machine learning & deep learning approaches for diabetes detection: a systematic review. E-Prime - Advances in Electrical Engineering, Electronics and Energy, 9, 100661. doi: https://doi.org/10.1016/j.prime.2024.100661

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. Journal of King Saud University - Computer and Information Sciences, 34(4), 1060–1073. doi: https://doi.org/10.1016/j.jksuci.2019.06.012

Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. International Journal of Cognitive Computing in Engineering, 2, 40–46. doi: https://doi.org/10.1016/j.ijcce.2021.01.001

Mehta, R. I., Capuano, A. W., Biswas, R., Bennett, D. A., & Arvanitakis, Z. (2025). Permutations of cerebrovascular pathologies in older adults with and without diabetes. Cerebral Circulation - Cognition and Behavior, 8(March), 100381. doi: 10.1016/j.cccb.2025.100381

Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. Informatics in Medicine Unlocked, 51, 101587. doi: https://doi.org/10.1016/j.imu.2024.101587

Molnar, C., König, G., Bischl, B., & Casalicchio, G. (2024). Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach. Data Mining and Knowledge Discovery, 38(5), 2903–2941. doi: 10.1007/s10618-022-00901-9

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

S, G., Venkata Siva Reddy, R., & Ahmed, M. R. (2024). Exploring the effectiveness of machine learning algorithms for early detection of Type-2 Diabetes Mellitus. Measurement: Sensors, 31, 100983. doi: https://doi.org/10.1016/j.measen.2023.100983

Wang, C., Xu, X., Luo, S., Luo, M., Li, S., & Si, J. (2025). Interpretable machine learning insights into the association between PFAS exposure and diabetes mellitus. Ecotoxicology and Environmental Safety, 302, 118569. doi: https://doi.org/10.1016/j.ecoenv.2025.118569

Wang, L., Wang, Y., & Chang, Q. (2016). Feature selection methods for big data bioinformatics: A survey from the search perspective. Methods, 111, 21–31. doi: https://doi.org/10.1016/j.ymeth.2016.08.014

Wong, P.-Y., Zeng, Y.-T., Su, H.-J., Lung, S.-C. C., Chen, Y.-C., Chen, P.-C., Hsiao, T.-C., Adamkiewicz, G., & Wu, C.-D. (2025). Effects of feature selection methods in estimating SO2 concentration variations using machine learning and stacking ensemble approach. Environmental Technology & Innovation, 37, 103996. doi: https://doi.org/10.1016/j.eti.2024.103996

Zhou, P., Liang, J., Yan, Y., Zhao, S., & Wu, X. (2024). Explainable feature selection and ensemble classification via feature polarity. Information Sciences, 676, 120818. doi: https://doi.org/10.1016/j.ins.2024.120818

Aliyu, H. A., Muritala, I. O., Bello-Salau, H., Mohammed, S., Onumanyi, A. J., & Ajayi, O.-O. (2024). Optimizing machine learning algorithms for diabetes data: A metaheuristic approach to balancing and tuning classifiers parameters. Franklin Open, 8, 100153. doi: https://doi.org/10.1016/j.fraope.2024.100153

Allgaier, J., Mulansky, L., Draelos, R. L., & Pryss, R. (2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, 143, 102616. doi: https://doi.org/10.1016/j.artmed.2023.102616

Baniecki, H., Sobieski, B., Szatkowski, P., Bombinski, P., & Biecek, P. (2025). Interpretable machine learning for time-to-event prediction in medicine and healthcare. Artificial Intelligence in Medicine, 159, 103026. doi: https://doi.org/10.1016/j.artmed.2024.103026

Cahyani, N., Irsyada, R., & Kartini, A. Y. (2025). Implementasi Machine Learning Model sebagai Sistem Prediksi Penyakit Breast Cancer. Digital Transformation Technology, 4(2), 1112–1120. doi: 10.47709/digitech.v4i2.5209

Cahyani, N., Irsyada, R., & Mahmuda, R. (2025). Penerapan Algoritma Neural Network untuk Klasifikasi Diabetes Mellitus: Perbandingan Backpropagation dan Resillient Backpropagation. Digital Transformation Technology, 4(2), 1067–1074. doi: 10.47709/digitech.v4i2.5208

Cahyani, N., & Kartini, A. Y. (2022). DLNN dan BPNN-GA Pada Prediksi Penyakit Diabetes di Bojonegoro. Journal of Mathematics Education and Science, 6(1), 1–9. doi: 10.32665/james.v6i1.416

Cahyani, N., Putri, W. A., & Irsyada, R. (2025). Improving Multiclass Rainfall Prediction with Multilayer Perceptron and SMOTE: Addressing Class Imbalance Challenges. Brilliance: Research of Artificial Intelligence, 4(2), 901–908. doi: 10.47709/brilliance.v4i2.5203

Cantú-Brito, C., Mimenza-Alvarado, A., & Sánchez-Hernández, J. J. (2010). Diabetes mellitus and aging as a risk factor for cerebral vascular disease: epidemiology, pathophysiology and prevention. Revista de Investigacion Clinica, 62(4), 333–342.

Cao, Z., Wu, X., Wu, B., Zhang, Z., & Sun, J. (2025). Combining UDT with XGBoost to identify the geographical origin of black beans by near-infrared spectroscopy. Current Research in Food Science, 11, 101131. doi: https://doi.org/10.1016/j.crfs.2025.101131

Chang, V., Ganatra, M. A., Hall, K., Golightly, L., & Xu, Q. A. (2022). An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. Healthcare Analytics, 2, 100118. doi: https://doi.org/10.1016/j.health.2022.100118

Cottin, A., Zulian, M., Pécuchet, N., Guilloux, A., & Katsahian, S. (2024). MS-CPFI: A model-agnostic Counterfactual Perturbation Feature Importance algorithm for interpreting black-box Multi-State models. Artificial Intelligence in Medicine, 147, 102741. doi: https://doi.org/10.1016/j.artmed.2023.102741

Daza, A., Ponce Sánchez, C. F., Apaza-Perez, G., Pinto, J., & Zavaleta Ramos, K. (2024). Stacking ensemble approach to diagnosing the disease of diabetes. Informatics in Medicine Unlocked, 44, 101427. doi: https://doi.org/10.1016/j.imu.2023.101427

Delpino, F. M., Costa, Â. K., Farias, S. R., Chiavegatto Filho, A. D. P., Arcêncio, R. A., & Nunes, B. P. (2022). Machine learning for predicting chronic diseases: a systematic review. Public Health, 205, 14–25. doi: https://doi.org/10.1016/j.puhe.2022.01.007

Dharmarathne, G., Jayasinghe, T. N., Bogahawaththa, M., Meddage, D. P. P., & Rathnayake, U. (2024). A novel machine learning approach for diagnosing diabetes with a self-explainable interface. Healthcare Analytics, 5, 100301. doi: https://doi.org/10.1016/j.health.2024.100301

Gavin, J. R., Rodbard, H. W., Battelino, T., Brosius, F., Ceriello, A., Cosentino, F., Giorgino, F., Green, J., Ji, L., Kellerer, M., Koob, S., Kosiborod, M., Lalic, N., Marx, N., Prashant Nedungadi, T., Parkin, C. G., Topsever, P., Rydén, L., Huey-Herng Sheu, W., … Schnell, O. (2024). Disparities in prevalence and treatment of diabetes, cardiovascular and chronic kidney diseases – Recommendations from the taskforce of the guideline workshop. Diabetes Research and Clinical Practice, 211, 111666. doi: https://doi.org/10.1016/j.diabres.2024.111666

E-ISSN : 2807-9035
Volume 5, **Number** 1, May 2025
https://doi.org/10.47709/brilliance.v5i1.6507

Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. Decision Analytics Journal, 7, 100242. doi: https://doi.org/10.1016/j.dajour.2023.100242

Jiang, L., Xia, Z., Zhu, R., Gong, H., Wang, J., Li, J., & Wang, L. (2023). Diabetes risk prediction model based on community follow-up data using machine learning. Preventive Medicine Reports, 35, 102358. doi: https://doi.org/10.1016/j.pmedr.2023.102358

Katiyar, N., Thakur, H. K., & Ghatak, A. (2024). Recent advancements using machine learning & deep learning approaches for diabetes detection: a systematic review. E-Prime - Advances in Electrical Engineering, Electronics and Energy, 9, 100661. doi: https://doi.org/10.1016/j.prime.2024.100661

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. Journal of King Saud University - Computer and Information Sciences, 34(4), 1060–1073. doi: https://doi.org/10.1016/j.jksuci.2019.06.012

Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. International Journal of Cognitive Computing in Engineering, 2, 40–46. doi: https://doi.org/10.1016/j.ijcce.2021.01.001

Mehta, R. I., Capuano, A. W., Biswas, R., Bennett, D. A., & Arvanitakis, Z. (2025). Permutations of cerebrovascular pathologies in older adults with and without diabetes. Cerebral Circulation - Cognition and Behavior, 8(March), 100381. doi: 10.1016/j.cccb.2025.100381

Mienye, I. D., Obaido, G., Jere, N., Mienye, E., Aruleba, K., Emmanuel, I. D., & Ogbuokiri, B. (2024). A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges. Informatics in Medicine Unlocked, 51, 101587. doi: https://doi.org/10.1016/j.imu.2024.101587

Molnar, C., König, G., Bischl, B., & Casalicchio, G. (2024). Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach. Data Mining and Knowledge Discovery, 38(5), 2903–2941. doi: 10.1007/s10618-022-00901-9

S, G., Venkata Siva Reddy, R., & Ahmed, M. R. (2024). Exploring the effectiveness of machine learning algorithms for early detection of Type-2 Diabetes Mellitus. Measurement: Sensors, 31, 100983. doi: https://doi.org/10.1016/j.measen.2023.100983

Wang, C., Xu, X., Luo, S., Luo, M., Li, S., & Si, J. (2025). Interpretable machine learning insights into the association between PFAS exposure and diabetes mellitus. Ecotoxicology and Environmental Safety, 302, 118569. doi: https://doi.org/10.1016/j.ecoenv.2025.118569

Wang, L., Wang, Y., & Chang, Q. (2016). Feature selection methods for big data bioinformatics: A survey from the search perspective. Methods, 111, 21–31. doi: https://doi.org/10.1016/j.ymeth.2016.08.014

Wong, P.-Y., Zeng, Y.-T., Su, H.-J., Lung, S.-C. C., Chen, Y.-C., Chen, P.-C., Hsiao, T.-C., Adamkiewicz, G., & Wu, C.-D. (2025). Effects of feature selection methods in estimating SO2 concentration variations using machine learning and stacking ensemble approach. Environmental Technology & Innovation, 37, 103996. doi: https://doi.org/10.1016/j.eti.2024.103996

Zhou, P., Liang, J., Yan, Y., Zhao, S., & Wu, X. (2024). Explainable feature selection and ensemble classification via feature polarity. Information Sciences, 676, 120818. doi: https://doi.org/10.1016/j.ins.2024.120818