

Implementation of Soft Voting Ensemble Learning for Early Detection of Type 2 Diabetes Mellitus

Carles^{1)*}, Osi Draini²⁾, Sabtria Winda Sari³⁾, Nu'man⁴⁾, M. Khairul Anam⁵⁾

^{1,2,3,4)} Department of Nursing, STIKES Tengku Maharatu, Indonesia,

⁵⁾ Department of Informatics, Universitas Samudra, Indonesia

¹⁾carles.ulung1@gmail.com, ²⁾osiidrainii@gmail.com, ³⁾arienda011219@gmail.com,

⁴⁾numanjamil06@gmail.com, ⁵⁾khairulanam@unsam.ac.id

Submitted : Aug 1, 2026 | **Accepted :** Aug 15, 2026 | **Published :** Oct 5, 2026

Abstract: Type 2 Diabetes Mellitus is a chronic disease that requires early detection to minimize the risk of complications. This study aims to develop a machine learning model for the early detection of Type 2 Diabetes Mellitus using an ensemble Soft Voting approach. The proposed model combines three algorithms, namely LightGBM, Random Forest, and Support Vector Machine. The dataset used consists of patient health attributes, including pregnancies, glucose level, blood pressure, skin thickness, insulin level, body mass index, diabetes pedigree function, and age. The preprocessing stage was conducted to prepare the data before the training process, including data balancing so that each class contained the same number of records, namely 500 diabetes data and 500 non-diabetes data. The evaluation results show that the model achieved an accuracy of 80%, with precision, recall, and f1-score of 0.80 for both classes. In addition, the Area Under the Curve value of 0.89 indicates that the model has a very good capability in distinguishing between diabetes and non-diabetes classes. The system was also implemented in the form of a Graphical User Interface to facilitate patient clinical data input and display risk detection results. Based on these results, the developed Soft Voting model can be used as a practical and sufficiently accurate tool for the early detection of Type 2 Diabetes Mellitus.

Keywords: Diabetes; Machine Learning; GUI; Ensemble Learning; Soft Voting.

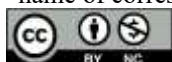
INTRODUCTION

Type 2 Diabetes Mellitus is one of the chronic diseases that has become a serious concern in the healthcare sector because the number of patients continues to increase and can cause various complications if not detected early (Iparraguirre-Villanueva et al., 2023). This disease occurs when the body experiences difficulties in using insulin effectively, causing blood glucose levels to rise above normal limits. If not properly managed, Type 2 Diabetes Mellitus can lead to complications such as cardiovascular disease, kidney disorders, nerve damage, vision impairment, and decreased quality of life. Therefore, early detection is an important step to help healthcare professionals identify diabetes risk more quickly and accurately.

The development of information technology, particularly in the field of artificial intelligence, has opened significant opportunities to support disease prediction and detection processes. Machine learning has become one of the most widely used approaches because it can learn patterns from historical data and generate prediction models based on patient characteristics (Enríquez-Ortega et al., 2025). In the context of Type 2 Diabetes Mellitus detection, machine learning can be utilized to analyze various risk factors such as age, body mass index, blood pressure, glucose level, family medical history, and other clinical indicators. The utilization of machine learning models is expected to support early screening processes so that high-risk patients can receive faster treatment.

Although various machine learning algorithms have been applied in diabetes prediction, each algorithm has its own characteristics and limitations. LightGBM is known for its strong capability in handling tabular data and efficient training processes (Kull et al., 2025). Random Forest has advantages in reducing overfitting through the combination of multiple decision trees (Hamdani et al., 2024), while Support Vector Machine is capable of forming optimal classification boundaries in high-dimensional data (Anam et al., 2025). However, the use of a single algorithm is often insufficient to produce stable performance under various data conditions (Anam et al., 2024).

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Therefore, it is necessary to develop a model capable of combining the strengths of multiple algorithms to improve prediction accuracy and stability.

Several previous studies have investigated machine learning approaches for diabetes prediction and have reported varying levels of classification performance. (Erin et al., 2022) applied Logistic Regression to the Pima Indians Diabetes Database using SMOTE and Grid Search, increasing accuracy from 77% to 82%. (Trisna & Ayu Mahessya, 2025) applied K Nearest Neighbor to the Early-Stage Diabetes Risk Prediction dataset and reported an accuracy of 97% at $K = 3$. (Ginting et al., 2022) through a scoping review of 15 studies, identified SVM, Random Forest, Decision Tree, LASSO, and several other algorithms as commonly used approaches for Type 2 Diabetes Mellitus prediction. (Haryadi et al., 2024) implemented Decision Tree using RapidMiner and reported an accuracy of 77.34%, precision of 75.08%, recall of 97.60%, and an F1 score of 0.8486.

Another study by (Pratama & Utomo, 2026) further investigated ensemble learning by combining XGBoost and LightGBM with SMOTE Tomek and K Fold cross validation. Their results showed accuracies of 97.31% for XGBoost and 97.26% for LightGBM, while the blending model achieved 97.51%. (Susanto et al., 2025) also evaluated Random Forest for diabetes prediction using Indonesian health data and considered several performance measures, including accuracy, precision, recall, F1 score, specificity, and ROC AUC. These studies demonstrate that high classification performance can be achieved using different machine learning configurations. However, the reported results cannot be directly compared solely on the basis of accuracy because the studies differ in terms of datasets, study populations, preprocessing procedures, class distributions, sampling methods, validation strategies, and model configurations.

Therefore, the objective of the present study is not to claim the highest absolute prediction accuracy compared with all previously published diabetes prediction models. Instead, the research gap addressed in this study concerns the limited empirical evaluation of a heterogeneous Soft Voting ensemble that combines LightGBM, Random Forest, and Support Vector Machine under the same experimental conditions. These algorithms represent different learning characteristics, with LightGBM and Random Forest employing tree based learning mechanisms and SVM constructing classification boundaries based on class separation. Their integration provides an opportunity to investigate whether combining complementary prediction probabilities can produce better overall classification performance than using each constituent classifier independently.

To provide stronger evidence regarding the contribution of the proposed method, this study evaluates SVM, Random Forest, LightGBM, and Soft Voting using the same dataset, preprocessing procedures, training and testing partitions, and evaluation metrics. The proposed ensemble is therefore assessed not only by its absolute performance but also by its relative improvement over its constituent classifiers under controlled experimental conditions. In addition, experiments are conducted both with and without SMOTE. This comparison is intended to distinguish the contribution of the Soft Voting ensemble mechanism from the additional effect of class balancing. Under the experimental setting used in this study, Soft Voting achieved higher accuracy, precision, recall, F1 score, and AUC than the three individual classifiers, both with and without SMOTE.

The use of SMOTE is also evaluated as a separate component of the proposed framework. SMOTE is applied exclusively to the training data after the train test split to avoid information leakage, while the original testing data remain unchanged for final evaluation. By comparing Soft Voting with and without SMOTE, the study provides additional evidence regarding whether class balancing contributes to the observed performance improvement. This experimental design allows the contribution of ensemble learning and class balancing to be examined separately rather than attributing the final performance to a single component.

This study therefore develops an early detection model for Type 2 Diabetes Mellitus using a Soft Voting ensemble consisting of LightGBM, Random Forest, and Support Vector Machine. Soft Voting combines the predicted probabilities generated by the individual classifiers and determines the final class based on the aggregated probability values (Rahayu et al., 2024). Rather than relying on a single classifier, the proposed approach integrates predictions from algorithms with different learning characteristics, allowing complementary classification patterns to contribute to the final decision.

Based on this research gap, the main objective of this study is to evaluate whether the proposed Soft Voting ensemble provides better predictive performance than its individual LightGBM, Random Forest, and SVM components under identical experimental conditions. A secondary objective is to examine the effect of SMOTE by comparing model performance with and without class balancing. Furthermore, the final model is implemented through a Graphical User Interface to demonstrate its potential use as an early screening and decision support application. The GUI is considered a practical implementation of the predictive model rather than evidence of superior classification performance. Overall, the contribution of this study lies in the controlled comparative evaluation of heterogeneous ensemble learning and class balancing for Type 2 Diabetes Mellitus prediction.

LITERATURE REVIEW

To provide a clearer overview of previous studies and to identify the research gap addressed in this study, a concept-centric literature review was conducted. Rather than organizing the literature solely based on individual

authors, this approach compares previous studies according to the main concepts, methods, and techniques relevant to Type 2 Diabetes Mellitus detection. The comparison focuses on the use of data balancing, LightGBM, Random Forest, Support Vector Machine (SVM), ensemble learning, Soft Voting, and Graphical User Interface (GUI) implementation. Previous studies have applied various individual machine learning algorithms and ensemble approaches for diabetes prediction, while Soft Voting has also been explored in several other classification domains. The concept-centric comparison presented in Table X highlights the similarities and differences among existing studies and demonstrates the position of the proposed study, which integrates LightGBM, Random Forest, and SVM within a Soft Voting framework for the early detection of Type 2 Diabetes Mellitus.

Table 1 Concept Centric Literature Review

No.	Study	Diabetes Detection	Data Balancing	LightGBM	Random Forest	SVM	Ensemble	Soft Voting	GUI
1	(Erlin et al., 2022)	✓	✓	–	–	–	–	–	–
2	(Trisna & Ayu Mahessya, 2025)	✓	–	–	–	–	–	–	–
3	(Ginting et al., 2022)	✓	–	–	✓	✓	–	–	–
4	(Pratama & Utomo, 2026)	✓	✓	✓	–	–	✓	–	–
5	(Susanto et al., 2025)	✓	–	–	✓	–	–	–	–
6	(Rahayu et al., 2024)	–	–	–	–	–	✓	✓	–
7	(Karlos et al., 2020)	–	–	–	–	–	✓	✓	–
8	(Rai et al., 2022)	–	–	–	–	–	✓	✓	–
9	(Suandi et al., 2024)	–	✓	–	–	–	✓	–	–
10	(Atiyah, 2024)	–	–	–	–	–	✓	✓	–
11	(Jabbar, 2024)	–	–	–	–	–	✓	✓	–
12	Proposed Study	✓	✓	✓	✓	✓	✓	✓	✓

As shown in Table 1, previous studies have applied various machine learning approaches for the prediction and early detection of Type 2 Diabetes Mellitus (T2DM). Study (Erlin et al., 2022) employed Logistic Regression combined with SMOTE and Grid Search, while (Trisna & Ayu Mahessya, 2025) applied the K-Nearest Neighbor algorithm. Study (Ginting et al., 2022), through a scoping review, identified several machine learning algorithms used for T2DM prediction, including Random Forest and Support Vector Machine (SVM). Study (Pratama & Utomo, 2026) explored an ensemble approach combining XGBoost and LightGBM with SMOTE-Tomek, whereas (Susanto et al., 2025) evaluated Random Forest using a healthcare dataset from Indonesia.

Several studies have also investigated voting-based ensemble approaches and Soft Voting in other classification domains. Study (Rahayu et al., 2024) applied a voting technique in sentiment analysis, (Karlos et al., 2020) investigated a Soft Voting ensemble-based approach for binary classification, and (Rai et al., 2022) applied Soft Voting for COVID-19 mortality prediction. Voting-based ensemble methods were also explored in sentiment analysis, earthquake prediction, and advanced threat detection (Suandi et al., 2024; Atiyah, 2024; Jabbar, 2024).

Based on the reviewed literature, the simultaneous integration of LightGBM, Random Forest, and SVM within a Soft Voting framework for the early detection of T2DM remains limited. Therefore, the proposed study addresses this gap by combining these three heterogeneous classifiers using Soft Voting, supported by data balancing and comprehensive model evaluation. Furthermore, the resulting model is implemented in a Graphical User Interface (GUI) to facilitate patient clinical data input and diabetes risk prediction.

METHOD

This section explains the research stages carried out in the development of an early detection model for Type 2 Diabetes Mellitus using a Soft Voting-based machine learning approach. The research methodology was systematically designed, starting from dataset collection, data preprocessing, class balancing, classification model construction, model performance evaluation, and implementation of the model into a Graphical User Interface (GUI).

The Soft Voting approach was used to combine three machine learning algorithms, namely LightGBM, Random Forest, and Support Vector Machine. All methodological stages were designed to produce a model capable of classifying the risk of Type 2 Diabetes Mellitus accurately, stably, and practically as an early screening tool based on patient clinical data.

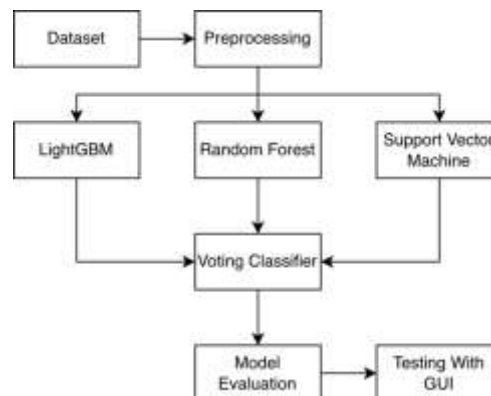


Fig. 1 Methodology Flow

Dataset

The dataset used in this study is the Pima Indians Diabetes Database, accessed through the Kaggle platform at <https://www.kaggle.com/code/alaabdelstar/diabetes/input>. The dataset originates from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and is widely used as a benchmark dataset for diabetes classification. The dataset contains 768 patient records with eight predictor variables and one binary target variable.

The study population represented in the dataset consists of female patients of Pima Indian heritage who were at least 21 years old. Historical descriptions of the dataset indicate that the population was drawn from the Pima Indian community in the United States. The dataset is a publicly available secondary dataset and contains no direct personal identifiers used in the present study. Therefore, this research did not involve direct patient recruitment, clinical intervention, or new data collection.

The eight predictor variables are Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age. Pregnancies represents the number of previous pregnancies; Glucose represents plasma glucose concentration measured during an oral glucose tolerance test; BloodPressure represents diastolic blood pressure; SkinThickness represents triceps skinfold thickness; Insulin represents the 2-hour serum insulin level; BMI represents body mass index; DiabetesPedigreeFunction represents diabetes-related family history; and Age represents patient age in years. The target variable, Outcome, is binary, where 0 indicates non-diabetes and 1 indicates diabetes.

The original dataset contains 500 non-diabetes records and 268 diabetes records, resulting in a total of 768 observations. Thus, the dataset exhibits an imbalanced class distribution, with the non-diabetes class representing approximately 65.1% of the observations and the diabetes class representing approximately 34.9%. This class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE), which was applied exclusively to the training data after the train-test split to prevent information leakage.

The exact year of clinical data collection is not explicitly documented in the Kaggle copy used in this study. Therefore, this study does not make additional assumptions regarding the period of data collection. As the analysis uses an existing publicly accessible secondary dataset and does not involve direct interaction with participants, no new participant recruitment or informed consent procedure was conducted as part of this study.

Data Splitting and Class Balancing Using SMOTE

To prevent data leakage, the original dataset was divided into training and testing sets before the class balancing procedure was performed. A stratified 80:20 train-test split was used to preserve the original class proportions in both subsets. From the original 768 records, 614 observations were allocated to the training set and 154 observations were retained as the independent testing set.

*name of corresponding author



Before SMOTE, the training set consisted of 400 non-diabetes samples (Class 0) and 214 diabetes samples (Class 1). SMOTE was then applied exclusively to the training set to increase the representation of the minority diabetes class. After oversampling, the training set contained 400 non-diabetes samples and 400 diabetes samples, resulting in a balanced training set of 800 observations.

Importantly, the testing set was not included in the SMOTE procedure and retained its original distribution of 100 non-diabetes samples and 54 diabetes samples. Keeping the testing data completely separate from synthetic sample generation ensures that the final evaluation is performed on previously unseen original observations and reduces the risk of information leakage between the training and testing stages.

Figure 2 presents the class distribution of the training set after the SMOTE balancing procedure. The balanced training distribution provides equal representation of diabetes and non-diabetes observations during model learning, while the unchanged testing set provides a more realistic assessment of model generalization performance.

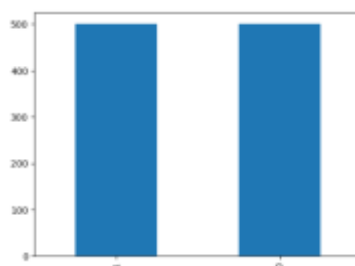


Fig. 2 Graph After Class Balancing Using SMOTE

This balancing process is important to ensure that the machine learning model does not become biased toward the majority class during training (Nugroho H et al., 2025). With a balanced data distribution, the model is expected to learn patterns from each class more effectively, thereby improving classification performance in detecting the risk of Type 2 Diabetes Mellitus.

Preprocessing

The preprocessing stage was conducted to prepare the data so that it could be properly processed by the machine learning model (Ikhwan Saputra et al., 2025). This process includes checking for missing values, removing duplicate data, data transformation, feature normalization, and splitting the dataset into training and testing data. The preprocessing stage is important to improve data quality, which can ultimately enhance the performance of the classification model (Van Fc et al., 2024).

Voting Classifier

This study employed an ensemble learning method based on Soft Voting by combining three machine learning algorithms: LightGBM, Random Forest, and Support Vector Machine (SVM). These algorithms were trained using the preprocessed data to generate prediction probabilities for each class. The Voting Classifier then combined the probabilities from each model and determined the final prediction result based on the highest average probability value (Karlos et al., 2020; Rai et al., 2022; Suandi et al., 2024).

The Soft Voting approach was selected because it can utilize the strengths of each algorithm, resulting in more stable and accurate predictions compared to using a single model (Atiyah, 2024; Jabbar, 2024). LightGBM was chosen for its efficiency in handling tabular data, Random Forest for its ability to reduce overfitting through multiple decision trees, and SVM for its capability in forming optimal classification boundaries on complex data. The Soft Voting formula used in this study is expressed as follows:

$$\hat{y} = \arg \max_c \left(\frac{1}{N} \sum_{i=1}^N P_i(c) \right) \quad (1)$$

Description:

$\hat{y}$ = final prediction result

$P_i(c)$ = probability of class c from the i -th model

N = number of models used

c = prediction class

Model Evaluation

The model evaluation stage was conducted to measure the performance of the developed model. Evaluation was performed using several metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. The evaluation results were used to determine the model's capability in detecting the risk of Type 2 Diabetes Mellitus accurately and consistently.

*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Testing With GUI

The testing with GUI stage was conducted to implement the model into a user interface so that it could be used more easily. Users can input patient health data through the GUI, after which the system processes the data using the Soft Voting model and displays the diabetes risk prediction results. This stage aims to ensure that the model can be practically applied as an early detection tool for Type 2 Diabetes Mellitus.

RESULT

This section presents the results of the development and evaluation of a machine learning model for the early detection of Type 2 Diabetes Mellitus using a Soft Voting approach that combines the LightGBM, Random Forest, and Support Vector Machine (SVM) algorithms. The testing process was conducted to determine the model's capability in classifying patient data based on diabetes risk factors. In addition, this study also evaluates the contribution of the ensemble method in improving prediction performance compared to the use of a single model.

The research results were analyzed using several evaluation metrics, such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC to measure the effectiveness of the model in detecting the risk of Type 2 Diabetes Mellitus. This section also discusses the implementation of the model through a Graphical User Interface (GUI) as a practical application of the developed early detection system. All obtained results were then compared and analyzed to identify the strengths, limitations, and potential future developments of the model.

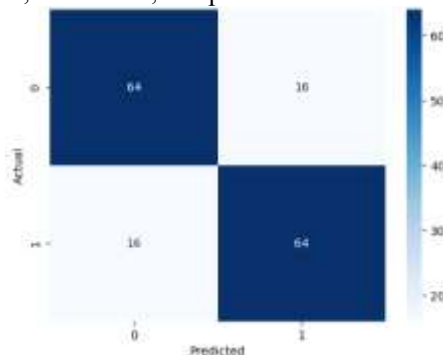


Fig. 3 Soft Voting Test Confusion Matrix Results

The Fig. 3 shows the Confusion Matrix from the testing results of the Type 2 Diabetes Mellitus classification model. The confusion matrix is used to evaluate model performance by comparing prediction results with actual data. The matrix consists of two classes, namely class 0 (non-diabetes) and class 1 (diabetes).

Based on the testing results, the model correctly predicted 64 data instances of class 0, referred to as True Negative (TN). In addition, the model also correctly predicted 64 data instances of class 1, referred to as True Positive (TP). Meanwhile, there were 16 class 0 data instances incorrectly predicted as class 1 (False Positive/FP) and 16 class 1 data instances incorrectly predicted as class 0 (False Negative/FN).

These confusion matrix results indicate that the model has a relatively balanced classification capability in detecting both diabetes and non-diabetes patients. The high number of True Positive and True Negative values indicates that the model can recognize data patterns effectively. However, there are still some prediction errors, as shown by the False Positive and False Negative values of 16 instances each.

Classification Report:				
	precision	recall	f1-score	support
0	0.80	0.80	0.80	80
1	0.80	0.80	0.80	80
accuracy			0.80	160
macro avg	0.80	0.80	0.80	160
weighted avg	0.80	0.80	0.80	160

Fig. 4 Soft Voting Test Classification Report Results

The figure 4 also shows the Classification Report from the evaluation results of the machine learning model for the early detection of Type 2 Diabetes Mellitus. The classification report is used to measure model performance based on several evaluation metrics, including precision, recall, f1-score, and accuracy for each class.

For class 0 (non-diabetes), the model achieved a precision value of 0.80, recall of 0.80, and f1-score of 0.80 with a support value of 80 data instances. A precision value of 0.80 indicates that 80% of the data predicted as non-diabetes actually belong to the non-diabetes class. A recall value of 0.80 indicates that the model correctly identified 80% of the non-diabetes data. Meanwhile, the f1-score value of 0.80 indicates a good balance between

precision and recall.

For class 1 (diabetes), the model also achieved precision, recall, and f1-score values of 0.80 with a total support of 80 data instances. These results indicate that the model has consistent capability in detecting both diabetes and non-diabetes patients. The high recall value for the diabetes class indicates that the model performs well in identifying patients at risk of Type 2 Diabetes Mellitus.

Furthermore, the model achieved an accuracy value of 0.80 or 80% from a total of 160 testing data instances. The macro average and weighted average values also showed results of 0.80 for precision, recall, and f1-score. This indicates that the model performance is relatively stable and balanced across both classes without significant performance differences between diabetes and non-diabetes classes.

Based on the classification report results, it can be concluded that the developed Soft Voting model is capable of classifying Type 2 Diabetes Mellitus with good and stable performance across all evaluation metrics.

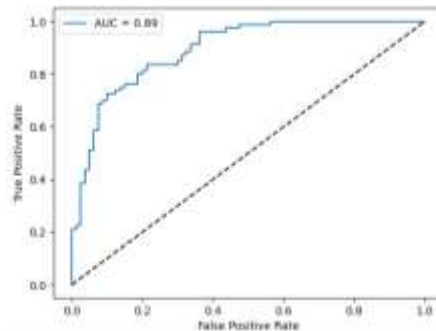


Fig. 5 Soft Voting Test ROC Results

The fig. 5 presents the ROC Curve (Receiver Operating Characteristic Curve) of the machine learning model used for the early detection of Type 2 Diabetes Mellitus. The ROC Curve is used to evaluate the model's capability in distinguishing between the positive class (diabetes) and the negative class (non-diabetes). The ROC Curve illustrates the relationship between the True Positive Rate (TPR) or sensitivity on the vertical axis and the False Positive Rate (FPR) on the horizontal axis.

The diagonal dashed line in the graph represents the performance of a random classifier. The farther the ROC curve is above the diagonal line, the better the model's classification capability. In the figure, the ROC curve is positioned considerably above the diagonal line, indicating that the model has a good capability in distinguishing diabetes and non-diabetes patients.

The obtained AUC (Area Under Curve) value is 0.89. The AUC value is used to measure the overall quality of the classification model. The closer the value is to 1, the better the model performance. An AUC value of 0.89 indicates that the model has a high capability in recognizing data patterns and making accurate predictions. In general, the AUC value can be interpreted as follows:

- 0.50 = random classification
- 0.60–0.70 = fair
- 0.70–0.80 = good
- 0.80–0.90 = very good
- 0.90–1.00 = excellent

Based on the ROC Curve results, the developed Soft Voting model demonstrates strong performance in detecting the risk of Type 2 Diabetes Mellitus and effectively distinguishes between diabetic and non-diabetic classes. To further evaluate the contribution of the ensemble approach, its performance was compared with the individual SVM, Random Forest, and LightGBM models, as presented in Table 2. The comparison demonstrates that the Soft Voting model achieved the highest performance across all evaluated metrics.

Table 2
Performance Comparison Between Individual Models and the Soft Voting Ensemble

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC
SVM	71%	72%	71%	70%	79%
Random Forest	65%	65%	65%	65%	69%
LightGBM	74%	75%	74%	74%	82%

Table 2 presents a performance comparison between the individual machine learning algorithms, namely SVM, Random Forest, and LightGBM, and the proposed Soft Voting ensemble model. The comparison was conducted

using accuracy, precision, recall, F1-score, and AUC as evaluation metrics. Among the individual models, LightGBM achieved the best overall performance, with an accuracy of 74%, precision of 75%, recall of 74%, F1-score of 74%, and AUC of 82%. SVM achieved an accuracy of 71% and an AUC of 79%, while Random Forest showed the lowest individual performance, with an accuracy of 65% and an AUC of 69%.

The Soft Voting model, which combines the prediction outputs of SVM, Random Forest, and LightGBM, achieved the highest performance across all evaluation metrics. The ensemble model obtained an accuracy, precision, recall, and F1-score of 80%, with an AUC of 89%. Compared with the best-performing individual model, LightGBM, Soft Voting increased accuracy from 74% to 80%, recall from 74% to 80%, F1-score from 74% to 80%, and AUC from 82% to 89%. Precision also increased from 75% to 80%.

Compared with SVM, the Soft Voting model improved accuracy by 9 percentage points, precision by 8 percentage points, recall by 9 percentage points, F1-score by 10 percentage points, and AUC by 10 percentage points. A larger difference was observed compared with Random Forest, where Soft Voting improved accuracy, precision, recall, and F1-score by 15 percentage points and AUC by 20 percentage points. These results indicate that combining the three classifiers through the Soft Voting mechanism provides better overall predictive performance than using each classifier independently.

The improvement obtained by Soft Voting can be attributed to the complementary prediction patterns of the individual classifiers. Rather than relying on the decision of a single algorithm, Soft Voting combines the predicted probabilities produced by SVM, Random Forest, and LightGBM to determine the final class. Consequently, prediction errors produced by one classifier can potentially be compensated for by the other classifiers, contributing to a more balanced performance in distinguishing between diabetes and non-diabetes cases.

To further evaluate whether this improvement was primarily associated with the ensemble mechanism rather than the class balancing process, an additional experiment was conducted without applying SMOTE. In this scenario, SVM, Random Forest, LightGBM, and Soft Voting were trained using the original class distribution of the training data. The purpose of this experiment was to determine whether the proposed Soft Voting approach could still provide better classification performance than the individual classifiers in the absence of synthetic oversampling. The performance comparison is presented in Table 3 using accuracy, precision, recall, F1-score, and AUC as evaluation metrics.

Table 3
Performance Comparison Between Individual Models and Soft Voting Without SMOTE

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC
SVM	70%	71%	71%	70%	79%
Random Forest	63%	64%	64%	64%	67%
LightGBM	70%	71%	73%	73%	78%
Soft Voting	77%	78%	77%	77%	85%

Table 3 presents the performance of SVM, Random Forest, LightGBM, and the proposed Soft Voting ensemble when the models were trained without applying SMOTE. This experiment was conducted to evaluate the classification performance under the original class distribution and to examine whether the Soft Voting ensemble could maintain its advantage over the individual classifiers without synthetic oversampling.

Among the individual classifiers, SVM achieved an accuracy of 70%, precision of 71%, recall of 71%, F1-score of 70%, and AUC of 79%. Random Forest obtained the lowest performance, with an accuracy of 63%, precision, recall, and F1-score of 64%, and an AUC of 67%. LightGBM achieved an accuracy of 70%, precision of 71%, recall of 73%, F1-score of 73%, and AUC of 78%. Although SVM and LightGBM obtained the same accuracy, LightGBM produced higher recall and F1-score, whereas SVM achieved a slightly higher AUC.

The Soft Voting ensemble achieved the best overall performance even without SMOTE, with an accuracy of 77%, precision of 78%, recall of 77%, F1-score of 77%, and AUC of 85%. Compared with SVM, Soft Voting increased accuracy by 7 percentage points, precision by 7 percentage points, recall by 6 percentage points, F1-score by 7 percentage points, and AUC by 6 percentage points. Compared with Random Forest, the improvements were 14 percentage points in accuracy, 14 percentage points in precision, 13 percentage points in recall, 13 percentage points in F1-score, and 18 percentage points in AUC. Compared with LightGBM, Soft Voting improved accuracy by 7 percentage points, precision by 7 percentage points, recall by 4 percentage points, F1-score by 4 percentage points, and AUC by 7 percentage points.

These results demonstrate that the advantage of the Soft Voting ensemble is not limited to the use of SMOTE. Even when the models were trained using the original class distribution, Soft Voting consistently achieved higher overall performance than each individual classifier. This finding provides additional evidence that combining SVM, Random Forest, and LightGBM allows the ensemble to benefit from the complementary learning characteristics of the three algorithms.

A comparison between the experiments with and without SMOTE also indicates that class balancing

contributed to improving the performance of most models. The Soft Voting model achieved an accuracy of 77% and an AUC of 85% without SMOTE, whereas the SMOTE-based experiment produced an accuracy of 80% and an AUC of 89%. Thus, the use of SMOTE was associated with increases of 3 percentage points in accuracy and 4 percentage points in AUC for the Soft Voting model. Precision increased from 78% to 80%, while recall and F1-score increased from 77% to 80%.

A similar tendency was observed for LightGBM, whose accuracy increased from 70% without SMOTE to 74% with SMOTE and whose AUC increased from 78% to 82%. Random Forest also showed improvements from 63% to 65% in accuracy and from 67% to 69% in AUC. In contrast, SVM showed only a small change, with accuracy increasing from 70% to 71%, while its AUC remained at 79%. These findings suggest that the effect of SMOTE varies across classifiers, with LightGBM and Soft Voting showing more noticeable improvements than SVM.

Overall, the two experimental scenarios show that the proposed Soft Voting ensemble consistently achieved the strongest classification performance both with and without SMOTE. However, its best results were obtained when SMOTE was applied exclusively to the training set, reaching 80% accuracy and an AUC of 89%. This comparison supports two aspects of the proposed approach: the contribution of the Soft Voting ensemble relative to the individual classifiers and the additional benefit associated with class balancing during model training.

To further examine the effect of class balancing on the proposed ensemble model, an additional comparison was conducted between the Soft Voting model trained with SMOTE and the same model trained without SMOTE. This comparison was designed to determine whether the application of SMOTE to the training data contributes to improvements in classification performance while maintaining the same ensemble configuration. The results of both experimental scenarios are summarized in Table 4 using accuracy, precision, recall, F1-score, and AUC as evaluation metrics.

Table 4
Comparison of Soft Voting Performance With and Without SMOTE

Scenario	Accuracy	Precision	Recall	F1-Score	AUC
Soft Voting without SMOTE	77%	78%	77%	77%	85%
Soft Voting with SMOTE	80%	80%	80%	80%	89%
Improvement	+3 pp	+2 pp	+3 pp	+3 pp	+4 pp

As shown in Table 4, the Soft Voting model achieved better performance when SMOTE was incorporated into the training process. Without SMOTE, the ensemble obtained an accuracy of 77%, precision of 78%, recall of 77%, F1-score of 77%, and AUC of 85%. After applying SMOTE exclusively to the training set, the accuracy increased to 80%, precision to 80%, recall to 80%, F1-score to 80%, and AUC to 89%.

The comparison indicates improvements of 3 percentage points in accuracy, 2 percentage points in precision, 3 percentage points in recall, 3 percentage points in F1-score, and 4 percentage points in AUC. The improvement in recall is particularly relevant for early Type 2 Diabetes Mellitus screening because a higher recall indicates a greater ability of the model to identify positive diabetes cases. Similarly, the increase in AUC from 85% to 89% indicates an improvement in the model's overall ability to discriminate between diabetes and non-diabetes cases across different classification thresholds.

These results suggest that the class imbalance present in the original training data influenced the learning process of the ensemble model. By generating synthetic minority-class samples exclusively within the training set, SMOTE provided a more balanced representation of diabetes and non-diabetes cases during model development. This allowed the constituent classifiers to learn patterns from both classes more evenly before their predicted probabilities were combined through the Soft Voting mechanism.

Nevertheless, the improvement should be interpreted as the performance difference observed under the experimental settings used in this study. The comparison demonstrates that the Soft Voting configuration with SMOTE produced better results than the corresponding configuration without SMOTE, but further validation using cross-validation and independent clinical datasets would be necessary to determine whether the improvement remains consistent across different samples and data distributions.

Overall, the results presented in Table 4 provide additional evidence for the proposed model configuration. Soft Voting already outperformed the individual SVM, Random Forest, and LightGBM classifiers in the experiment without SMOTE, while the inclusion of SMOTE further increased its overall predictive performance. Therefore, the combination of class balancing and heterogeneous ensemble learning produced the best performance observed in this study, with an accuracy of 80% and an AUC of 89%. The following section presents the GUI implementation testing.



Fig. 6 Data Input on GUI

The fig. 6 shows the Graphical User Interface (GUI) of the Type 2 Diabetes Mellitus early detection system developed using a Soft Voting-based machine learning model. This GUI was designed to facilitate users in entering patient clinical data before the prediction process is performed by the system.

The GUI contains several health parameters used as prediction inputs, including pregnancies, glucose level, blood pressure, skin thickness, insulin level, body mass index (BMI), diabetes pedigree function (DPF), and patient age. All of these parameters are the main features used in the diabetes prediction dataset and have an influence on the risk of Type 2 Diabetes Mellitus.

In the displayed input example, the patient has a glucose level of 148 mg/dL, a BMI value of 33.60, and an age of 50 years. These values indicate several risk factors associated with the possibility of diabetes. After all data are entered, users can press the “PREDICTION PROCESS” button to execute the machine learning model.

The system then processes the data using the Soft Voting model that combines the LightGBM, Random Forest, and Support Vector Machine (SVM) algorithms. The prediction results are used to determine whether the patient has a risk of Type 2 Diabetes Mellitus or not. The implementation of this GUI aims to provide ease of use so that the developed model not only has good classification performance but can also be practically applied as an early diabetes detection tool.



Figure 7. Test Results on GUI

The figure 7 shows the final result display of the Type 2 Diabetes Mellitus early detection system based on the Soft Voting machine learning model. After users input patient clinical data on the input page, the system processes the data using the combination of LightGBM, Random Forest, and Support Vector Machine (SVM), then displays the diabetes risk classification results.

In the prediction result display, the system provides a “HIGH RISK (RESIKO TINGGI)” status, indicating that the patient has a high possibility of experiencing Type 2 Diabetes Mellitus based on the entered health data. This information is presented in a simple and understandable manner to assist users and healthcare professionals in performing rapid initial screening.

In addition to displaying prediction results, the system also provides a feature to display the factors that contribute most significantly to the classification result. In the “DOMINANT RISK FACTORS (Faktor Resiko Dominan)” section, the system displays several key parameters influencing the prediction, namely Pregnancies, Glucose, BloodPressure, and SkinThickness. These factors are attributes that have a major influence on the

machine learning decision-making process.

The existence of the dominant risk factor feature adds value to the system because it improves the interpretability of prediction results. Users not only obtain the diabetes risk classification results but can also understand which health parameters are the main causes of the high risk of Type 2 Diabetes Mellitus in patients. Therefore, the developed system functions not only as a prediction tool but also as an analytical aid in the early health screening process.

DISCUSSIONS

The experimental results demonstrate that the proposed Soft Voting ensemble model provides promising performance for the early detection of Type 2 Diabetes Mellitus. Under the SMOTE-based experimental configuration, the model achieved an accuracy of 80%, with precision, recall, and F1-score values of 0.80 and an AUC of 0.89. These results indicate that the proposed ensemble is capable of distinguishing between diabetes and non-diabetes cases with relatively balanced classification performance.

Recall is particularly important in the context of early diabetes screening because it reflects the model's ability to correctly identify patients belonging to the diabetes class. A recall of 0.80 indicates that the model was able to identify a substantial proportion of positive diabetes cases. Nevertheless, False Negative predictions remain an important consideration because individuals with diabetes who are incorrectly classified as non-diabetes may not be identified for further examination. Therefore, reducing False Negative cases should remain an important objective in future model development.

In addition to classification accuracy, the proposed Soft Voting model achieved an AUC value of 0.89. The AUC provides information regarding the model's discriminative capability across different classification thresholds and complements the information provided by accuracy. The obtained AUC therefore indicates that the ensemble model has a strong ability to discriminate between diabetes and non-diabetes cases based on the predicted probabilities.

To evaluate the contribution of the Soft Voting strategy, the proposed ensemble was directly compared with its constituent models, namely SVM, Random Forest, and LightGBM, under the same SMOTE-based experimental configuration. As presented in Table 2, LightGBM achieved the best performance among the individual classifiers, with an accuracy of 74%, precision of 75%, recall of 74%, F1-score of 74%, and AUC of 82%. SVM achieved an accuracy of 71% and an AUC of 79%, while Random Forest achieved an accuracy of 65% and an AUC of 69%.

The Soft Voting model outperformed all three individual classifiers across the reported evaluation metrics. Compared with LightGBM, which was the strongest individual classifier, Soft Voting increased accuracy from 74% to 80%, precision from 75% to 80%, recall from 74% to 80%, F1-score from 74% to 80%, and AUC from 82% to 89%. These differences correspond to improvements of 6 percentage points in accuracy, 5 percentage points in precision, 6 percentage points in recall, 6 percentage points in F1-score, and 7 percentage points in AUC.

Compared with SVM, Soft Voting improved accuracy by 9 percentage points, precision by 8 percentage points, recall by 9 percentage points, F1-score by 10 percentage points, and AUC by 10 percentage points. The largest difference was observed relative to Random Forest, where Soft Voting improved accuracy, precision, recall, and F1-score by 15 percentage points and AUC by 20 percentage points. These results provide quantitative evidence that the Soft Voting mechanism achieved better overall predictive performance than the use of any of the individual classifiers under the same experimental configuration.

The improved performance of Soft Voting may be associated with the complementary learning characteristics of the constituent classifiers. LightGBM and Random Forest are tree-based learning methods, whereas SVM constructs decision boundaries based on class separation. Rather than relying on the output of a single classifier, Soft Voting combines the predicted class probabilities produced by the three models to determine the final prediction. This mechanism enables the ensemble to incorporate different representations of the underlying data patterns and may reduce the influence of prediction errors produced by an individual classifier.

To further determine whether the performance advantage was primarily associated with the ensemble mechanism rather than only with class balancing, an additional experiment was conducted without applying SMOTE. In this experiment, SVM, Random Forest, LightGBM, and Soft Voting were trained using the original class distribution of the training data. As presented in Table 3, Soft Voting continued to achieve the highest overall performance, with an accuracy of 77%, precision of 78%, recall of 77%, F1-score of 77%, and AUC of 85%.

Without SMOTE, SVM achieved an accuracy of 70%, precision of 71%, recall of 71%, F1-score of 70%, and AUC of 79%. Random Forest obtained an accuracy of 63%, precision of 64%, recall of 64%, F1-score of 64%, and AUC of 67%. LightGBM achieved an accuracy of 70%, precision of 71%, recall of 73%, F1-score of 73%, and AUC of 78%. Although SVM and LightGBM obtained the same accuracy, LightGBM demonstrated higher recall and F1-score, while SVM achieved a slightly higher AUC.

The Soft Voting model also outperformed all three individual classifiers when SMOTE was not applied. Compared with SVM, Soft Voting improved accuracy by 7 percentage points and AUC by 6 percentage points. Compared with Random Forest, the improvements reached 14 percentage points in accuracy and 18 percentage

points in AUC. Compared with LightGBM, Soft Voting increased accuracy from 70% to 77% and AUC from 78% to 85%. These findings demonstrate that the performance advantage of Soft Voting was still observed under the original class distribution, providing additional evidence that the improvement is associated with the ensemble strategy and not solely with the use of SMOTE.

The effect of class balancing on the Soft Voting model was further examined by directly comparing its performance with and without SMOTE, as presented in Table 4. Without SMOTE, Soft Voting achieved an accuracy of 77%, precision of 78%, recall of 77%, F1-score of 77%, and AUC of 85%. When SMOTE was applied exclusively to the training set, performance increased to 80% accuracy, 80% precision, 80% recall, 80% F1-score, and an AUC of 89%.

The comparison in Table 4 therefore shows improvements of 3 percentage points in accuracy, 2 percentage points in precision, 3 percentage points in recall, 3 percentage points in F1-score, and 4 percentage points in AUC. These results indicate that class balancing was associated with an additional improvement in the predictive performance of the Soft Voting model. In particular, the increase in recall suggests an improvement in the model's ability to identify positive diabetes cases, while the increase in AUC indicates better overall discrimination between diabetes and non-diabetes observations.

Importantly, SMOTE was applied only to the training data after the train-test split. The testing data were retained in their original distribution and were not involved in synthetic sample generation. This procedure was used to prevent information leakage and to ensure that model evaluation was performed using original observations that were not involved in the oversampling procedure. Thus, the comparison between the SMOTE and non-SMOTE scenarios provides a more appropriate assessment of the contribution of class balancing to the proposed model.

Overall, the two experimental scenarios provide evidence for two different components of the proposed approach. First, Soft Voting consistently outperformed SVM, Random Forest, and LightGBM both with and without SMOTE, demonstrating the contribution of the ensemble mechanism. Second, the comparison between the two Soft Voting configurations shows that the application of SMOTE to the training data was associated with additional improvements in accuracy, precision, recall, F1-score, and AUC. The best overall result was therefore obtained from the combination of Soft Voting and SMOTE, with an accuracy of 80% and an AUC of 0.89.

When compared with previous studies, the performance obtained in this study is not the highest reported for diabetes prediction. (Erilin et al., 2022), for example, reported an improvement in accuracy from 77% to 82% using Logistic Regression combined with SMOTE and Grid Search. (Trisna & Ayu Mahessya, 2025) reported an accuracy of 97% using K-Nearest Neighbor with K=3. Furthermore, (Pratama & Utomo, 2026) reported accuracies of 97.31% for XGBoost and 97.26% for LightGBM, while their blending approach achieved 97.51%. However, these results should not be compared solely on the basis of accuracy because the studies used different datasets, preprocessing procedures, sampling strategies, model configurations, and experimental designs.

The contribution of the present study therefore does not depend solely on achieving the highest absolute accuracy reported in the literature. Instead, the experimental results demonstrate that the heterogeneous combination of LightGBM, Random Forest, and SVM through Soft Voting consistently outperformed the corresponding individual classifiers under the same experimental conditions. Furthermore, the additional experiment with and without SMOTE provides evidence regarding the contribution of class balancing to the final ensemble performance.

Another practical contribution of this study is the implementation of the prediction model through a Graphical User Interface (GUI). The developed interface allows users to enter clinical attributes such as pregnancies, glucose level, blood pressure, skin thickness, insulin level, BMI, Diabetes Pedigree Function, and age. The system processes these variables using the Soft Voting model and provides a prediction of diabetes risk. This implementation extends the study beyond model evaluation by demonstrating how the proposed model can be incorporated into a practical early screening application.

The GUI also presents dominant risk factors associated with the prediction result. In the example presented in the system, parameters such as Pregnancies, Glucose, Blood Pressure, and Skin Thickness were displayed as dominant factors contributing to the prediction. This information may improve the interpretability of the prediction by helping users understand which clinical variables are associated with the generated result. However, these factors should be interpreted as model-related predictive information rather than as clinical diagnoses or causal explanations. Therefore, the developed system is more appropriately positioned as an early screening and decision-support tool rather than as a replacement for professional medical assessment.

Despite the improved performance obtained by the proposed approach, further evaluation remains necessary. Future studies should evaluate the model using repeated cross-validation, larger and more diverse clinical datasets, and independent external datasets. Further investigation of hyperparameter optimization, alternative class-balancing strategies, probability calibration, feature selection, and ensemble weighting may also provide additional improvements. In particular, future experiments should compare False Negative rates for each individual classifier

and the Soft Voting ensemble to determine whether the ensemble provides a consistent advantage in detecting positive diabetes cases.

Overall, the findings demonstrate that the proposed Soft Voting ensemble provides better classification performance than SVM, Random Forest, and LightGBM when evaluated under comparable experimental conditions. The ensemble remained superior even without SMOTE, achieving 77% accuracy and an AUC of 85%, while the inclusion of SMOTE further increased its performance to 80% accuracy and an AUC of 89%. These findings support both the contribution of the ensemble strategy and the additional benefit of class balancing in the proposed early detection framework for Type 2 Diabetes Mellitus.

CONCLUSION

This study developed a machine learning-based early screening approach for Type 2 Diabetes Mellitus using a Soft Voting ensemble that combines LightGBM, Random Forest, and Support Vector Machine. To address class imbalance while preventing data leakage, SMOTE was applied exclusively to the training data after the train-test split, while the testing data remained unchanged for final evaluation. Under this configuration, the proposed Soft Voting model achieved an accuracy of 80%, precision of 80%, recall of 80%, F1-score of 80%, and an AUC of 0.89.

A direct comparison with the individual classifiers demonstrated that Soft Voting achieved the best overall performance under the same experimental conditions. LightGBM, the strongest individual classifier in the SMOTE-based experiment, achieved an accuracy of 74% and an AUC of 0.82, while SVM achieved an accuracy of 71% and an AUC of 0.79, and Random Forest obtained an accuracy of 65% and an AUC of 0.69. Compared with LightGBM, Soft Voting improved accuracy by 6 percentage points and AUC by 7 percentage points, while also achieving higher precision, recall, and F1-score. These findings provide quantitative evidence that combining the three classifiers through Soft Voting produces better overall classification performance than relying on any individual classifier under the evaluated conditions.

The additional experiment without SMOTE further demonstrated the contribution of the ensemble strategy. Without class balancing, Soft Voting continued to outperform the individual classifiers, achieving an accuracy of 77%, precision of 78%, recall of 77%, F1-score of 77%, and an AUC of 0.85. When SMOTE was applied to the training data, Soft Voting performance increased to 80% accuracy, 80% precision, 80% recall, 80% F1-score, and an AUC of 0.89. Therefore, the comparison between the two experimental scenarios shows improvements of 3 percentage points in accuracy, 2 percentage points in precision, 3 percentage points in recall, 3 percentage points in F1-score, and 4 percentage points in AUC. These results indicate that the performance advantage of Soft Voting is not solely dependent on class balancing, while the inclusion of SMOTE provides an additional improvement in overall predictive performance.

The implementation of the proposed model through a Graphical User Interface further demonstrates its practical potential as an early screening and decision-support system. The GUI enables users to input clinical parameters, obtain diabetes risk predictions, and view dominant factors associated with the generated prediction. However, the developed system is intended to support early screening and should not be considered a replacement for professional medical diagnosis.

Future research should evaluate the proposed approach using larger, more diverse, and independent clinical datasets to assess its robustness and generalizability. Further experiments should also investigate repeated cross-validation, hyperparameter optimization, alternative class balancing techniques, probability calibration, feature selection, and different ensemble weighting strategies. In addition, the False Negative performance of each individual classifier and the Soft Voting ensemble should be examined separately to determine whether the ensemble provides a consistent advantage in detecting positive diabetes cases. Overall, the findings support the effectiveness of combining heterogeneous classifiers through Soft Voting and demonstrate that the integration of Soft Voting with training-only SMOTE produced the best predictive performance observed in this study.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology (KEMENDIKTISAINTEK) for supporting this research through the PDP research grant scheme. The authors also extend their appreciation to STIKES Maharatu for the support, facilities, and contributions provided during the completion of this research.

REFERENCES

- Anam, M. K., Firdaus, M. B., Suandi, F., Lathifah, Nasution, T., & Fadly, S. (2024). Performance Improvement of Machine Learning Algorithm Using Ensemble Method on Text Mining. *ICFTSS 2024 - International Conference on Future Technologies for Smart Society*, 90–95. <https://doi.org/10.1109/ICFTSS61109.2024.10691363>



- Anam, M. K., Lestari, T. P., Efrizoni, L., Handayani, N. S., & Andhika, I. (2025). Sentiment Analysis Optimization Using Ensemble of Multiple SVM Kernel Functions. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 9(4), 905–914. <https://doi.org/10.29207/resti.v9i4.6708>
- Atiyah, O. S. (2024). Using Soft Voting Classifier of Machine Learning Algorithms for Earthquake Prediction. *Al-Kitab Journal for Pure Sciences*, 9(01), 1–13. <https://doi.org/10.32441/kjps.09.01.p1>
- Enríquez-Ortega, D., Chulde-Fernández, B., Pozo-Coral, P., Vaca, A., Zhinin-Vera, L., Almeida-Galárraga, D., Ramírez-Cando, L., Tirado-Espín, A., Cadena-Morejón, C., Villalba-Meneses, F., Guevara, C., & Acosta-Vargas, P. (2025). Enhancing Diabetes Diagnosis Through Machine Learning: A Comparative Study. *Applied Sciences (Switzerland)*, 15(1), 1–18. <https://doi.org/10.3390/app151810087>
- Erlin, Marlim, Y. N., Junadhi, Suryati, L., & Agustina, N. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2), 88–96. <https://doi.org/10.22146/jnteti.v11i2.3586>
- Ginting, J. B., Ginting, R., & Hartono. (2022). Deteksi dan Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Machine Learning (Scooping Review). *Jurnal Keperawatan Priority*, 5(2), 93–105. <https://doi.org/10.34012/jukep.v5i2.2671>
- Hamdani, Randi N.A., & Anam, M. K. (2024). Comparison of Support Vector Machine and Random Forest Algorithms for Analyzing Online Loans on Twitter social media. *JAIA-Journal Of Artificial Intelligence And Applications*, 4(1), 8–16. <https://doi.org/10.33372/jaia.v4i1.1087>
- Haryadi, A. N. P., Setiawan, W., & Fatah, D. A. (2024). Penerapan Data Mining untuk Klasifikasi Penyakit Diabetes Menggunakan Metode Decision Tree. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 7(6), 1484–1495. <https://doi.org/10.32672/jnkti.v7i6.8145>
- Ikhwan Saputra, E., Khairul Anam, M., Yenni, H., Hamdani, & Zamsuri, A. (2025). Optimalisasi Algoritma Support Vector Machine pada Aspect-Based Sentiment Analysis Menggunakan GridSearchCV. *Zonasi (Jurnal Sistem Informasi)*, 7(1), 271–279. <https://doi.org/10.31849/zn.v7i1.17800>
- Iparraguirre-Villanueva, O., Espinola-Linares, K., Castañeda, R. O. F., & Cabanillas-Carbonell, M. (2023). Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes. *Diagnostics*, 13(14), 1–16. <https://doi.org/10.3390/diagnostics13142383>
- Jabbar, H. G. (2024). Advanced Threat Detection Using Soft and Hard Voting Techniques in Ensemble Learning. *Journal of Robotics and Control (JRC)*, 5(4), 1104–1116. <https://doi.org/10.18196/jrc.v5i4.22005>
- Karlos, S., Kostopoulos, G., & Kotsiantis, S. (2020). A soft-voting ensemble based co-training scheme using static selection for binary classification problems. *Algorithms*, 13(1), 1–19. <https://doi.org/10.3390/a13010026>
- Kull, K., Khan, M. A., Asad, B., Naseer, M. U., Kallaste, A., & Vaimann, T. (2025). Adaptive Deep Belief Networks and LightGBM-Based Hybrid Fault Diagnostics for SCADA-Managed PV Systems: A Real-World Case Study. *Electronics (Switzerland)*, 14(18), 1–29. <https://doi.org/10.3390/electronics14183649>
- Nugroho H, Y. D., Zakiyabarsi, F., & Paramita, A. J. (2025). Implementasi SMOTE-ENN dan Borderline SMOTE Terhadap Performa LightGBM Pada Imbalanced Class. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(1), 51–59. <https://doi.org/10.36341/rabit.v10i1.5436>
- Pratama, N. A., & Utomo, D. W. (2026). Deteksi Diabetes Mellitus dengan Menggunakan Teknik Ensemble XGBoost dan LightGBM. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 11(1), 1–12. <https://doi.org/10.14421/jiska.4908>
- Rahayu, K., Anam, M. K., Efrizoni, L., Nurjayadi, & Fitri, T. A. (2024). Optimasi Teknik Voting Pada Sentimen Analisis Pemilihan Presiden 2024 Menggunakan Machine Learning. *The Indonesian Journal of Computer Science*, 13(4), 6289–6302. <https://doi.org/10.33022/ijcs.v13i4.4119>
- Rai, N., Kaushik, N., Kumar, D., Raj, C., & Ali, A. (2022). Mortality prediction of COVID-19 patients using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 3, 172–179. <https://doi.org/10.1016/j.ijcce.2022.09.001>
- Suandi, F., Anam, M. K., Firdaus, M. B., Fadli, S., Lathifah, L., Yumami, E., Saleh, A., & Hasibuan, A. Z. (2024). Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning Algorithms. *International Conference on Applied Engineering*, 126–138. https://doi.org/10.2991/978-94-6463-620-8_10
- Susanto, E. R., Inzaghi, M. R., Amarudin, A., & Neneng, N. (2025). Evaluasi Kinerja Model Random Forest Dalam Memprediksi Diabetes Berdasarkan Dataset Kesehatan di Indonesia. *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(7), 1857–1866. <https://doi.org/10.52436/1.jpti.871>

- Trisna, N., & Ayu Mahessya, R. (2025). Prediksi Risiko Diabetes Tahap Awal Menggunakan Machine Learning dengan Algoritma K-Nearest Neighbor. *Jurnal Pustaka Data (Pusat Akses Kajian Database, Analisa Teknologi, Dan Arsitektur Komputer)*, 5(2), 481–487. <https://doi.org/10.55382/jurnalpustakadata.v5i2.1550>
- Van Fc, L. L., Anam, M. K., Firdaus, M. B., Yunefri, Y., & Rahmi, N. A. (2024). Enhancing Machine Learning Model Performance in Addressing Class Imbalance. *COGITO Smart Journal*, 10(1), 478–490. <https://doi.org/10.31154/cogito.v10i1.626.478-490>