Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14439

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Analysis of Homogeneous and Heterogeneous Ensembles for Diabetes Classification Optimization

Muhammad Naufal Maulana1)*, **Muljono**2), **Eka Putra Agus Meindiawan**3)
1,2,3) Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
1)111202113237@mhs.dinus.ac.id, 2)muljono@dsn.ac.id, 3)111202113269@mhs.dinus.ac.id

Submitted :Jan 7, 2025 | **Accepted** : Feb 9, 2025 | **Published** : Feb 15, 2025

Abstract: Diabetes mellitus is a chronic disease with an increasing prevalence worldwide, including in Indonesia, reaching 11.7% by 2023. Early prediction of this disease is essential for more effective management. This study aims to develop a diabetes mellitus prediction model using an ensemble learning approach, including homogeneous (boosting and bagging) and heterogeneous (stacking and blending) techniques. In this study, the boosting algorithm using AdaBoost with Random Forest as the base estimator showed the highest accuracy of 98%, with balanced precision and recall. The bagging technique, which also uses Random Forest as the base estimator, achieved 97% accuracy, although slightly lower than boosting. The stacking technique, which combines XGBoost, Gradient Boosting, and Random Forest as base learners, with Random Forest as the meta-model, yields similar accuracy of 98%, but with lower prediction error, demonstrating its ability to cope with more complex data. Blending, which uses a similar approach but with training on the entire dataset, gave 98% accuracy with shorter processing time and more efficient memory usage than stacking.

Keywords: Diabetes; Boosting; Bagging; Stacking; Blending

INTRODUCTION

Diabetes mellitus, according to the World Health Organization (WHO), is a chronic degenerative disease that occurs when the pancreas does not produce enough insulin or the body is unable to use insulin effectively (Tanwar & Bhatia, 2024). This condition is characterized by hyperglycemia, an increase in blood glucose levels, which is one of the main indicators of the disease. Insulin is a hormone that regulates blood sugar levels. Diabetes is one of the fastest growing diseases in the world.

Currently, there are an estimated 537 million adults in the world aged 20 to 79 years old living with diabetes, equivalent to 10.5% of the population in that age range. This number is projected to increase to 643 million by 2030 and reach 783 million by 2045(Mujumdar & Vaidehi, 2019). According to the 10th edition of the International Diabetes Federation (IDF) report, the incidence of diabetes in the Southeast Asia (SEA) region has increased significantly over the past two decades, surpassing all previous predictions (Ogurtsova et al., 2017).

In Indonesia, according to a report from the Ministry of Health, the prevalence of diabetes in 2023 will reach 11.7%, an increase from 8.5% in 2018 (Rif'at et al., 2023). This highlights the importance of utilizing technology in diabetes management. By utilizing digital technology, such as machine learning, prediction and identification of deficiencies in patient care can be done more effectively.

Ensemble learning is an approach that combines the advantages of several algorithms to produce a more accurate model. The two main techniques in ensemble learning are bagging and boosting. Bagging (Bootstrap Aggregating) trains multiple models on a randomly selected subset of data and combines their predictions to produce the final result. Meanwhile, boosting trains models sequentially, with subsequent models focused on correcting the previous model's errors, thereby reducing bias and improving accuracy. One example of algorithms in boosting is AdaBoost and Gradient Boosting.

One of the advanced techniques in ensemble learning is stacking, which combines predictions from multiple base models using meta models to improve overall accuracy. Stacking has proven effective in medical data analysis, including prediction of diseases such as diabetes mellitus, due to its ability to handle complex and multidimensional data. Previous research shows that algorithms such as Gradient Boosting, XGBoost, and Random Forest are often used in medical studies (Mohammed & Kora, 2023). A research by (Wu et al., 2022) shows that stacking produces higher prediction accuracy than individual methods in disease prediction. Random Forest is

*name of corresponding author

often chosen as the meta model in stacking due to its ability to handle overfitting and produce more stable prediction results.

Besides stacking, blending is a similar ensemble technique but differs in the way it is learned. In blending, base models are trained on the entire dataset, and their predictions are used as features for meta models trained on different subsets of the data. Research by (Yorke-smith & Dumančić, 2023) showed that blending can give better results than stacking in some cases, especially in high-noise datasets.

A major challenge in medical data analysis is data imbalance, where positive cases (such as diabetics) are far fewer than negative cases. This imbalance can cause bias in the model. To address the problem, the SMOTE-Tomek Links technique is used, which generates synthetic data for minority classes and removes sample pairs that are too close together to improve separation between classes. The combination of SMOTE and Tomek Links helped to increase the proportion of minority data and reduce outliers, thus improving the sample quality for model training (Chandra et al., 2023).

This research focuses on the use of homogeneous ensemble (boosting and bagging) and heterogeneous ensemble (stacking and blending) approaches. The dataset used includes various features related to diabetes risk factors.Through preprocessing, including handling imbalanced data with the SMOTE-Tomek technique, models were built and evaluated.Thus, the research aims to develop and evaluate a machine learning model that can predict the risk of diabetes mellitus with high accuracy and identify the main risk factors that influence the occurrence of diabetes.

LITERATURE REVIEW

Machine Learning approaches are widely used in clinical epidemiological studies related to diabetes, due to their advantages in predicting and classifying patient characteristics by recognizing patterns in data sets(Mengcan et al., 2021). Machine Learning itself uses an algorithm model to work properly. Algorithms used in machine learning are divided into three types, namely supervised learning, unsupervised learning, and reinforcement learning. Each type of machine learning has a different process. The amount of research conducted shows great interest in studying methods and designing machine learning models that can predict and classify the characteristics of diabetes mellitus patients.

. Research by Umair Muneer Butt in 2021 using MLP outperformed other classifiers with 86.08% accuracy and LSTM improved significant predictions with 87.26% accuracy for diabetes. Moreover, a comparative analysis of the proposed approach was also carried out with existing state-of-the-art techniques, demonstrating the adaptability of the proposed approach in many healthcare applications.(Butt et al., 2021).

Research in 2022 by Victor Chang using the Naive Bayes classification model outperformed the random forest model and the J48 decision tree model in terms of accuracy. The Naive Bayes model on the 3-factor data subset performed as well as the random forest model on the full dataset with an accuracy of 79.13% compared to 79.57% which was the highest accuracy in this experiment(Chang et al., 2023).

Research by Roshi Saxena in 2022 which studied multilayer perceptron, decision tree, K-nearest neighbor, and random forest classifiers and several feature selection techniques were applied to classifiers to detect diabetes at an early stage. The raw data is subjected to preprocessing techniques, thus removing outliers and imputation of missing values based on averaging and then ultimately hyperparameter optimization. Experiments were conducted on Indian PIMA dataset using Weka 3.9 and the accuracy achieved for multilayer perceptron was 77.60%, for decision tree was 76.07%, for K-nearest neighbor was 78.58%, and for random forest was 79.8%, which is by far the best accuracy for random forest classifier(Saxena et al., 2022).

And the last there is research by Hang Lai in 2019 by building predictive models using Logistic Regression and Gradient Boosting Machine (GBM) techniques. The area under the receiver operating characteristic curve (AROC) was used to evaluate the discriminative ability of these models. We used adjusted threshold method and class weighting method to improve sensitivity - the proportion of Diabetes Mellitus patients correctly predicted by the models as well as compared these models with other machine learning techniques such as Decision Tree and Random Forest. The AROC result for the proposed GBM model is 84.7% with 71.6% sensitivity and the AROC for the proposed Logistic Regression model is 84.0% with 73.4% sensitivity. The GBM and Logistic Regression models performed better than the Random Forest and Decision Tree models(Lai et al., 2019).

METHOD

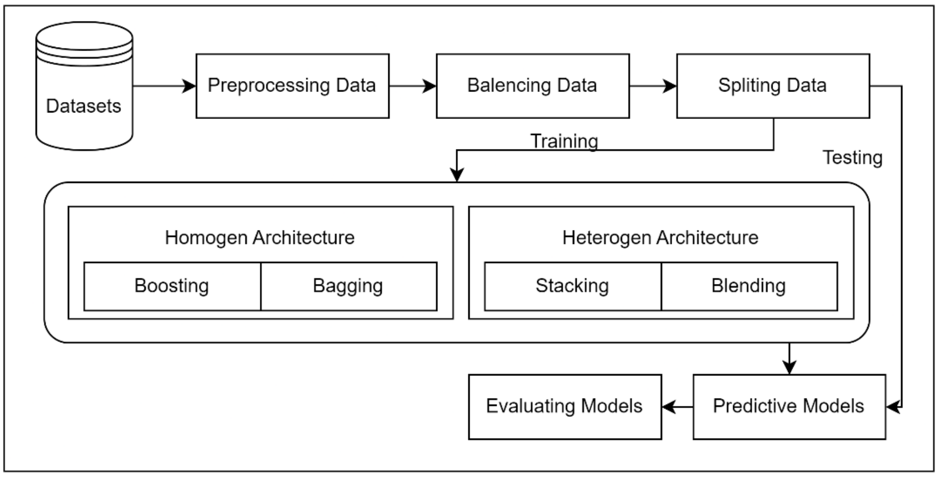

Figure 1. Research Methode

This research method begins with the collection of a dataset containing various features related to diabetes mellitus risk factors. The first step is data processing, which includes data cleaning and handling missing values. After that, data balancing is performed using the SMOTE-Tomek technique to overcome the problem of class imbalance in the dataset. Next, the dataset is divided into two parts, namely training data (train) and test data (test). In the next stage, classification is performed using two homogeneous ensemble approaches, such as boosting and bagging, and heterogeneous ensemble approaches, such as stacking and blending.

In homogeneous ensembles, different models are trained on different subsets of data to improve stability and accuracy. For example, in bagging, models are trained using subsets of data retrieved by bootstrapping techniques, while in boosting, models are trained sequentially to correct previous model errors. On the other hand, stacking and blending approaches are used to combine prediction results from several different models (base models) through meta models, with the aim of improving prediction performance.

Once the models are built and evaluated using relevant metrics such as accuracy, comparisons are made to evaluate the resource usage, such as RAM or memory usage, as well as the training and prediction process length of each model. This comparison aims to provide an overview of the efficiency and speed of each method, as well as provide considerations regarding practicality in implementation on resource-constrained systems. Finally, ensemble learning is applied to combine the results of the various models that have been trained, in order to produce higher accuracy and improve the ability to predict the risk of diabetes mellitus.

DATASET

This Diabetes Prediction Dataset consists of 100,000 patient data that includes medical and demographic information as well as their diabetes status (positive or negative). The dataset includes features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level(Alam et al., 2014). This dataset is taken from Kaggle provided by Mustafa I. This dataset is very useful for medical personnel to identify individuals at risk of developing diabetes, while supporting the development of more personalized treatment plans. In addition, this dataset can also be utilized by researchers to analyze the correlation between various risk factors and the likelihood of developing diabetes, providing additional insights into the prevention and treatment of this disease This dataset can be used to build machine learning models to predict diabetes based on a patient's medical and demographic profile(Mujumdar & Vaidehi, 2019).

Table 1 . Database Structure

gender	age	hypertension	heart_disease	Smoking_hystori	bmi	HbA1c_level	Blood_glucose_level	diabetes
Female	80.0	0	0	Never	25.19	6.6	140	0
Female	54.0	0	0	No info	27.32	6.6	80	0
Male	28.0	0	0	Never	27.32	5.7	158	0
Female	36.0	0	0	Current	23.45	5.0	155	0
Male	76.0	1	1	Currennt	20.14	4.8	155	0

DATA PREPROCESSING

Data processing is a crucial step in preparing a dataset before it is used for machine learning model development. This stage includes various processes to clean, process, and format the data to fit the needs of the model(Agnitia LEstari et al., 2021). The first step is to handle missing data, where empty values are removed by deleting rows or estimating empty values based on the available data patterns. After that, the dataset is checked to detect duplicate data, and if found, the duplicate rows are removed to avoid bias in model training.

The next step is normalization or standardization of numeric features to align the scale between features, so that the model is not affected by scale differences. In the case of datasets with class imbalance, the SMOTE-Tomek Links technique is used to balance the distribution of data between classes. SMOTE (Synthetic Minority Oversampling Technique) adds synthetic samples for minority classes, while Tomek Links removes overlapping samples between majority and minority classes, resulting in a cleaner and more balanced dataset.

. Categorical variables are then converted into a format that can be processed by the model using methods such as one-hot encoding or label encoding. Finally, the dataset is divided into training data and testing data to optimally evaluate the performance of the model. This entire process aims to improve the quality of the data while supporting the model's performance in making predictions.

DATA BALANCING

SMOTE (Synthetic Minority Over-sampling Technique) and Tomek Links were used in this study to overcome class imbalance(Fareed et al., 2022). SMOTE is an oversampling strategy that increases the sample size of a minority class by using synthetic data generated from the existing sample(Muljono et al., 2024). This technique works by taking a random sample from the minority class, finding its nearest neighbor, and then generating synthetic data by subtracting the sample from the nearest neighbor and multiplying it by a random number. In contrast, Tomek Links is an undersampling strategy that reduces the number of samples in the majority class. This method finds pairs of samples from different classes that are each other's nearest neighbors(Wang et al., 2019). These sample pairs, known as Tomek Links, were then eliminated from the majority class.

DATA SPLITING

The next step was to divide the data into training and testing subsets using the label stratification method with a ratio of 80:20 after the preprocessing and data imbalance handling processes were completed. After the application of SMOTE-Tomek Links, the total amount of data increased to 174,538 samples. This split resulted in 139,630 samples for training data (80%) and 34,908 samples for testing data (20%). The Train-Test Split separates the processed data into two parts: train data to train the model and test data to evaluate the model's performance(Kahloot & Ekler, 2021). With a ratio of 80:20, 80% of the data is used to build the model, while the remaining 20% is used to measure its effectiveness. The label stratification method ensures that the class distribution in the original dataset remains consistent in both subsets, so that the test data can represent all classes well.

ENSAMBLE LEARNING

Ensemble learning is an approach in machine learning that combines several “weak learners” to form one stronger model with better predictive performance. A “weak learner” is a model that performs slightly better than a random guess.(Gomes et al., 2017). The main goal in designing an ensemble is to ensure that each ensemble member has unique characteristics, especially with regard to misclassification. If the ensemble members tend to err on different instances, then they complement each other, and the combination of their predictions may result in higher performance than individual models. In ensemble learning, the term “combination” or voting is often used to describe the way the predictions of various members are combined to produce the final prediction. The ensemble architecture refers to how the classifiers are organized in the ensemble, while the voting method determines how the prediction results from each member are used to form the final decision.(Kumar et al., 2022). Although increasing diversity in the ensemble does not necessarily guarantee improved performance, this approach remains a very active area of research, with many new taxonomies and methods constantly being developed to improve the accuracy and effectiveness of models, both in the context of batch learning and in data streams.

BOOSTING

Boosting is an ensemble technique in machine learning that combines multiple base models (weak learners) that are trained sequentially to produce more accurate predictions. This process is done by incrementally optimizing a certain loss function at each iteration.(Manconi et al., 2022). In the boosting method, the input data is used to train the first weak model, and the error generated by the model is analyzed. The next model is then trained to correct the error by giving more weight to the hard-to-classify data. This process continues up to a certain number of iterations or until the loss function reaches the optimal result.

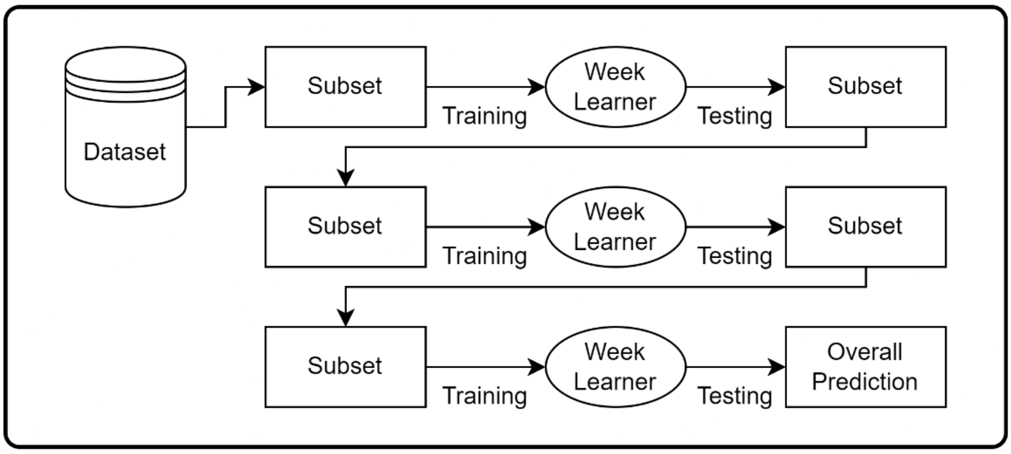

Figure 2. Boosting Structure

Other research shows that boosting methods focus more on bias reduction than variation, by improving the performance of initial models that have high bias (A & Homayouni, 2021). This technique is designed to tackle complex or difficult-to-classify data, so it can significantly improve model performance. Moreover, boosting is flexible and can be applied with a variety of machine learning algorithms, making it an effective choice for predictive data analysis. This approach helps to create more reliable models by utilizing the collective strength of multiple weak models.

BAGGING

Bagging (Bootstrap Aggregating) is an ensemble method in machine learning that aims to improve model accuracy by reducing variance(Ngo et al., 2022). This method involves creating multiple training models from different subsets of data, which are randomly selected by bootstrapping(Tuysuzoglu & Birant, 2020). Each model is trained using different subsets of data, and the prediction results from each model are combined, usually by voting for classification or averaging for regression, to produce a more stable final prediction.

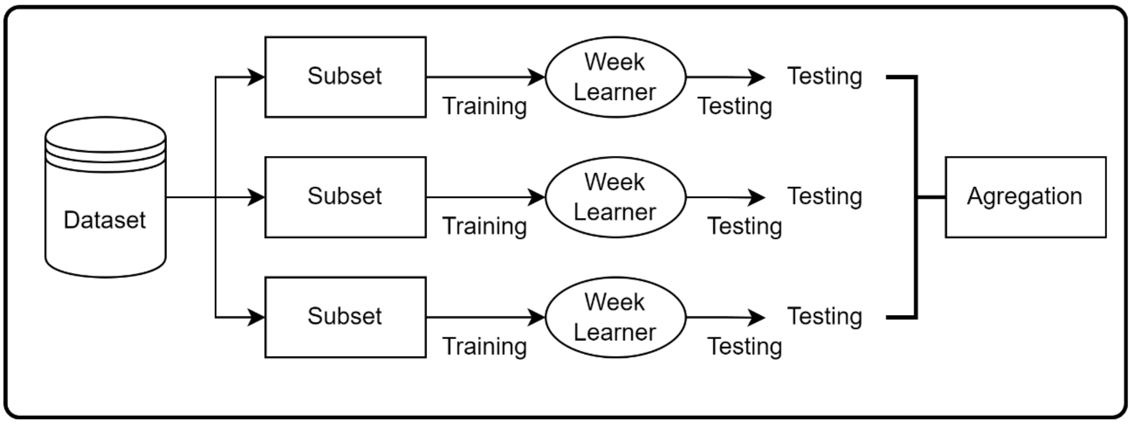

Figure 3. Bagging Structure

One example of the application of bagging is Random Forest, which combines multiple decision trees. Using this technique, Random Forest can reduce variance and avoid overfitting the training data, making it more effective in handling large and complex datasets. Bagging is particularly useful for improving the performance of models that are prone to overfitting, such as decision trees, by producing more accurate predictions.

STACKING

Stacking is an ensemble method in machine learning that combines predictions from multiple base models to improve prediction accuracy(Chatzimparpampas et al., 2021). Unlike bagging and boosting, stacking involves using a meta model (stacker) that is trained to combine the prediction results of the base models. The process consists of two stages: first, several base models are trained on the training data and generate predictions, and then those predictions are used as features to train the meta model that generates the final predictions.

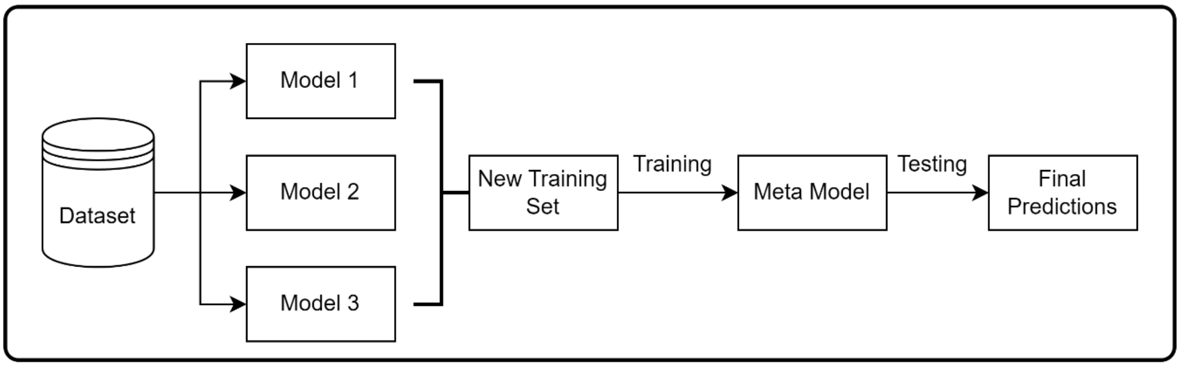

Figure 4. Stacking Structure

The main advantage of stacking is its ability to combine the strengths of different types of models, such as decision trees, logistic regression, or neural networks, resulting in a more robust model compared to a single model. Stacking is often more effective in improving model performance, especially when the various base models have different strengths and weaknesses.

BLENDING

Blending is an ensemble method similar to stacking, but with the main difference in how the data is used to train the meta model. In blending, the training data is divided into two parts: one part is used to train the base model, and the other part is used to train the meta model. Once the base model is trained, their predictions on the unseen part of the data are used as input for the meta model, which then generates the final predictions.

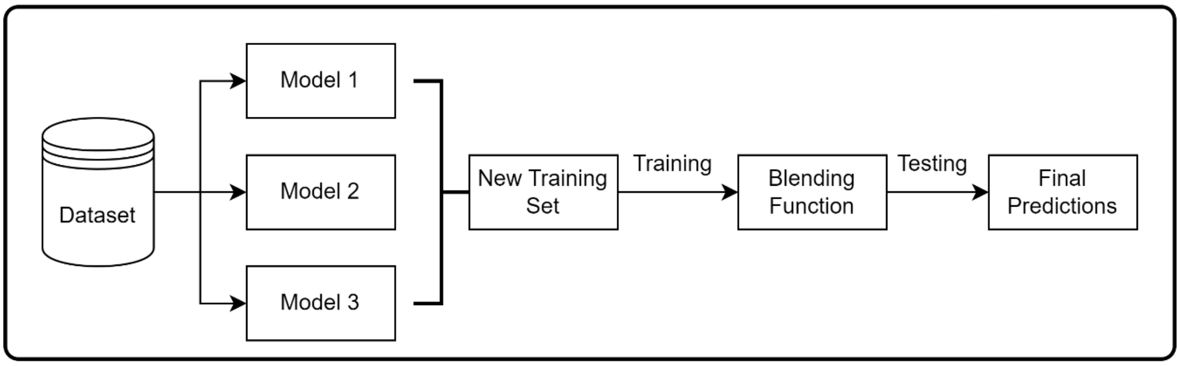

Figure 5. Blending Structure

Blending is simpler compared to stacking as it only requires one data split for meta model training, whereas stacking involves more complex data splits for validation and training. Nonetheless, blending is also effective in improving prediction accuracy by combining different underlying models. This technique is often used when there is a need to avoid overfitting, relying on the meta model to make better decisions based on the results from the base models.

PERFORME ANALISYS

Confusion matrix is a tool in machine learning and statistics used to assess the performance of classification algorithms(Zhang et al., 2021). It is a square matrix that is often used to summarize the results of a classification problem. The confusion matrix provides a detailed breakdown of the correct and incorrect predictions made by the classification model. It is particularly useful for binary (two-class) classification problems, but can also be used for multi-class classification(Nur et al., 2024). In the confusion matrix, there are four terms that describe the results of the classification process. As seen in Table II, these terms are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP and TN indicate correct classification results, while FP and FN indicate incorrect classification results. The formula for calculating accuracy is as in table 2. Confusion Matrix

Table 2 . Confussion Matrix

Precision	**Sensitifity**	**F1-Score**	**Accuracy**
$= \frac{TP}{TP + FP}$	$= \frac{TP}{TP + FN}$	$= 2x \frac{Precision \times Sensitifity}{Precision + Sensitifity}$	$= \frac{TP + TN}{TP + TN + FP + FN}$

RESULT

In this research, the first step is preprocessing the dataset, which starts with removing duplicate data.

Table 3. Deleting Duplicate Data

Jumlah Data	**Data Duplikat**	**Setelah Penghapusan Data**
100000	3854	96146

Table 3 shows that at the initial stage, the dataset used contained a number of duplicates. The presence of duplicates in the dataset can have a significant effect on the analysis results, as the same data repeated over and over again can give disproportionate weight to the model, which ultimately decreases the accuracy of the model. Therefore, the first step taken was to remove the duplicate data. Based on the results obtained, the dataset that had 100,000 initial entries, after the removal of duplicate data, was reduced to 96,146 entries. This ensures that the model built uses clean and more representative data. Next, data transformation was performed. In the data transforming stage, the categorical smoking history column in the data frame is converted into a numeric format, so that the data can be used in further analysis. The table shows that the smoking history data has been converted into numeric format.

Table 4. After Tranformasi Data

Gender	**Age**	**Hypertention**	**Heart Disease**	**Smoking History**	**bmi**	**HbAc Level**	**Blood Gucose Level**	**Diabetes**
0	80	0	1	4	25.19	6.6	140	0
1	54	0	0	0	27.32	6.6	80	0
2	28	0	0	4	27.32	5.7	158	0
3	36	0	0	1	23.45	5.0	155	0
4	76	1	1	1	20.14	4.8	155	0

The dataset used in this study initially amounted to 100,000 entries, which consisted of two classes, namely Normal and Diabetes. Of the total data, 91,500 entries were categorized as Normal and the other 8,500 entries were classified as Diabetes. This class imbalance may cause the model to be more inclined to classify the data into the majority class (Normal), potentially reducing the model's performance in detecting the minority class (Diabetes). To overcome this, the first step is to remove duplicate data. At this stage, duplicate data was identified and removed, resulting in 96,146 unique data entries. The removal of duplicate data is essential to ensure that each entry in the dataset represents only valid information, which will prevent distortions in the machine learning analysis and model.

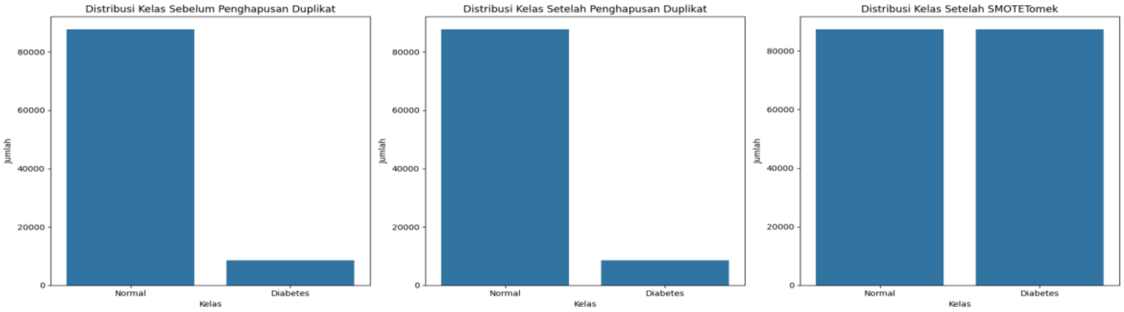

Figure 6. Preprosesing Dataset

*name of corresponding author

After the duplicate data is successfully removed, the next step is dataset balancing, because there is an imbalance between the amount of data for the Normal and Diabetes classes. This class imbalance can affect the accuracy of the model, as the model tends to classify more data into the class that has more data. To overcome this problem, the SMOTETomek technique is applied, which combines two approaches: Synthetic Minority Over-sampling Technique (SMOTE) and Tomek links. SMOTE generates a synthetic sample of the minority class (Diabetes), while the Tomek links technique serves to remove overlapping or close data between the majority class and the minority class, which can reduce potential errors in classification.

The result of applying SMOTETomek is a more balanced class distribution, with each class (Normal and Diabetes) having 87,269 data entries. With a balanced data distribution, the model built is expected to predict both classes more accurately, as well as reduce the bias that may occur towards the majority class. These steps aim to maximize the performance of the machine learning model in detecting diabetes effectively and precisely. The dataset that has gone through this balancing process is then divided into two subsets using the train-test split method with a ratio of 80:20, so that 139,631 data are used as training data and 34,908 data are used as testing data. The label stratification method is applied to ensure that the class distribution remains consistent between the training and testing data, so that the model evaluation results can represent the performance of the entire data population.

Several machine learning algorithms were applied to build a diabetes mellitus prediction model using an ensemble learning approach. The homogeneous ensemble learning approach involves boosting (such as AdaBoost) and bagging techniques, while the heterogeneous ensemble learning approach uses stacking and blending techniques to combine the strengths of multiple algorithms.

DISCUSSIONS

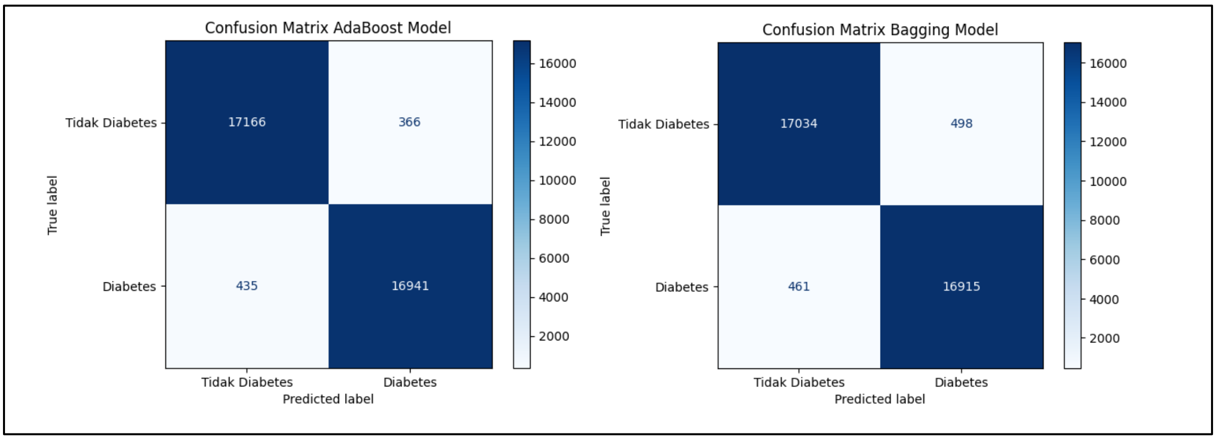

Figure 7. Confusion Matrix Homogen Models

The boosting approach using AdaBoost with Random Forest as the base estimator showed excellent results, achieving an accuracy rate of 98%. This model has a precision, recall, and F1-score of 0.98 for both classes, non-diabetes and diabetes. These results demonstrate the model's excellent ability to recognize both classes with minimal error. However, boosting requires a considerable training time of 386.50 seconds, with RAM usage of 2676.15 MB, making it more resource-intensive compared to some other methods.

The bagging approach with Random Forest as the base estimator also delivered high performance, with an accuracy rate of 97%. The precision, recall, and F1-score for both classes are 0.97, showing a good balance in predictions. However, bagging requires a longer training time compared to boosting, at 466.21 seconds, and higher RAM usage, reaching 3887.77 MB. This makes bagging a stable method but less efficient in resource usage compared to other methods such as stacking and blending.

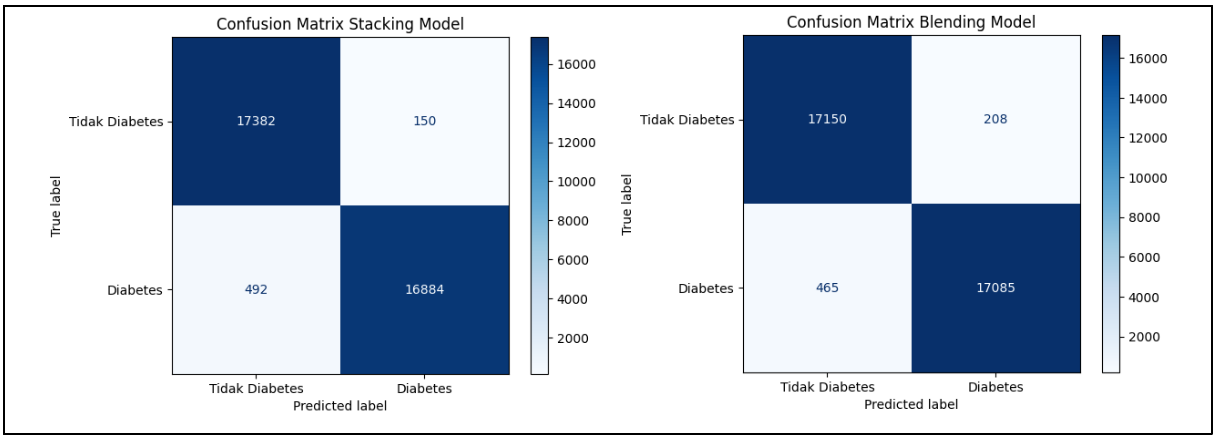

Figure 8. Confussion Matrix Heterogen Models

The stacking approach, which combines XGBoost, Gradient Boosting, and Random Forest as base learners with Random Forest as the meta-model, delivered excellent results with an accuracy of 98%. The precision, recall, and F1-score ranged from 0.97 to 0.99 for both classes, demonstrating the model's high reliability in identifying non-diabetes and diabetes. This approach was also more efficient compared to boosting and bagging, with a shorter training time of 189.23 seconds and RAM usage of 155.26 MB. This makes stacking a highly competitive method in terms of both performance and resource efficiency.

The blending approach, which also uses XGBoost, Gradient Boosting, and Random Forest as base learners and Random Forest as the meta-model, showed nearly identical performance to stacking, achieving an accuracy of 98%. The precision, recall, and F1-score for both classes were also very high, ranging from 0.97 to 0.99. However, blending proved to be the most efficient method, with a very short training time of 37.20 seconds and RAM usage of only 87.66 MB. Therefore, blending is an excellent choice when time and resource efficiency are top priorities.

Table 5. Evaluated Models

Boosting						
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Memory (MB)**	**Processing Time (Seconds)**
Diabetes	**0.98**	**0.97**	**0.98**	**0.98**	**2676.15 MB**	**386.50**
Normal	0.98	0.98	0.98	0.98		
Bagging						
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Memory (MB)**	**Processing Time (Seconds)**
Diabtes	0.97	0.97	0.97	0.97	3887.77	466.21
Normal	0.97	0.97	0.97	0.97		
Rondom Forest						
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Memory (MB)**	**Processing Time (Seconds)**
Diabtes	0.99	0.97	0.98	0.98	155.26	189.23
Normal	0.97	0.99	0.98	0.98		
Stacking						
Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Memory (MB)**	**Processing Time (Seconds)**
Diabtes	0.99	0.97	0.98	0.98	87.66	37.20
Normal	0.97	0.99	0.98	0.98		

CONCLUSION

This study demonstrates that ensemble learning approaches, both homogeneous (boosting and bagging) and heterogeneous (stacking and blending), have great potential in improving the accuracy of diabetes mellitus risk prediction. The data processing steps, including dataset balancing using the SMOTE-Tomek technique, proved effective in addressing the class imbalance issue, resulting in more balanced and accurate models. Based on the evaluation results, all applied ensemble learning approaches showed excellent performance in predicting type 2 diabetes, with high accuracy levels. The boosting approach using AdaBoost with Random Forest as the base estimator achieved the best accuracy (98%), but required considerable training time and RAM usage. The bagging approach with Random Forest showed good stability but was less efficient compared to other methods.

The stacking technique offers a balance between high accuracy (98%) and resource efficiency, making it a competitive method. Meanwhile, blending delivers performance comparable to stacking but far excels in time efficiency and RAM usage, with the shortest training time (37.20 seconds) and lowest memory usage (87.66 MB).

Overall, blending emerges as the most efficient choice, while stacking is suitable for consistent prediction needs with resource efficiency. Boosting and bagging remain relevant for cases where accuracy or stability is prioritized.

ACKNOWLEDGMENT

The author would like to express their deepest gratitude to Dian Nuswantoro University for their invaluable support in this research. The provision of facilities, resources, and academic guidance has been extremely helpful in ensuring the smooth progress of this study.

REFERENCES

A, P. D., & Homayouni, S. (2021). *Bagging and Boosting Ensemble Classifiers for Classification of Comparative Evaluation*.

Agnitia LEstari, M., Tabrani, M., & Ayumida, S. (2021). Sistem Informasi Pengolahan Data Administrasi Kependudukan Pada Kantor Desa Pucung Karawang. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, *13*(3), 14–21. https://doi.org/10.35969/interkom.v13i3.50

Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). General aspects of diabetes mellitus. *Handbook of Clinical Neurology*, *126*, 211–222. https://doi.org/10.1016/B978-0-444-53480-4.00015-1

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. *Journal of Healthcare Engineering*, *2021*. https://doi.org/10.1155/2021/9930985

Chandra, W., Suprihatin, B., & Resti, Y. (2023). Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction. *Symmetry*, *15*(4). https://doi.org/10.3390/sym15040887

Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, *35*(22), 16157–16173. https://doi.org/10.1007/s00521-022-07049-z

Chatzimparmpas, A., Martins, R. M., Kucher, K., & Kerren, A. (2021). StackGenVis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1547–1557. https://doi.org/10.1109/TVCG.2020.3030352

Fareed, M. M. S., Zikria, S., Ahmed, G., Mui-Zzud-Din, Mahmood, S., Aslam, M., Jillani, S. F., Moustafa, A., & Asad, M. (2022). ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI Scans. *IEEE Access*, *10*, 96930–96951. https://doi.org/10.1109/ACCESS.2022.3204395

Gomes, H. M., Barddal, J. P., Enembreck, A. F., & Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys*, *50*(2). https://doi.org/10.1145/3054925

Kahloot, K. M., & Ekler, P. (2021). Algorithmic Splitting: A Method for Dataset Preparation. *IEEE Access*, *9*, 125229–125237. https://doi.org/10.1109/ACCESS.2021.3110745

Kumar, M., Singhal, S., Shekhar, S., Sharma, B., & Srivastava, G. (2022). Optimized Stacking Ensemble Learning Model for Breast Cancer Detection and Classification Using Machine Learning. *Sustainability (Switzerland)*, *14*(21). https://doi.org/10.3390/su142113998

Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocrine Disorders*, *19*(1), 1–9. https://doi.org/10.1186/s12902-019-0436-6

Manconi, A., Armano, G., Gnocchi, M., & Milanesi, L. (2022). A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19. *Applied Sciences (Switzerland)*, *12*(15). https://doi.org/10.3390/app12157554

Mengcan, M. I. N., Xiaofang, C., & Yongfang, X. I. E. (2021). *Constrained voting extreme learning machine and its application*. *32*(1), 209–219. https://doi.org/10.23919/JSEE.2021.000018

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Mujumdar, A., & Vaidehi, V. (2019). Diabetes Prediction using Machine Learning Algorithms. *Procedia Computer Science*, *165*, 292–299. https://doi.org/10.1016/j.procs.2020.01.047

Muljono, Wulandari, S. A., Azies, H. Al, Naufal, M., Prasetyanto, W. A., & Zahra, F. A. (2024). Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI. *IEEE Access*, *12*(November 2023), 9292–9307. https://doi.org/10.1109/ACCESS.2024.3353788

Ngo, G., Beard, R., & Chandra, R. (2022). Evolutionary bagging for ensemble learning. *Neurocomputing*, *510*, 1–14. https://doi.org/10.1016/j.neucom.2022.08.055

Nur, A., Thohari, A., Karima, A., Santoso, K., & Rahmawati, R. (2024). *Crack Detection in Building Through Deep Learning Feature Extraction and Machine Learning Approach*. *8*(1), 1–6.

Ogurtsova, K., da Rocha Fernandes, J. D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., Cavan, D., Shaw, J. E., & Makaroff, L. E. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Research and Clinical Practice*, *128*, 40–50. https://doi.org/10.1016/j.diabres.2017.03.024

Rif'at, I. D., Hasneli N, Y., & Indriati, G. (2023). Gambaran Komplikasi Diabetes Melitus Pada Penderita Diabetes Melitus. *Jurnal Keperawatan Profesional*, *11*(1), 52–69. https://doi.org/10.33650/jkp.v11i1.5540

Saxena, R., Sharma, S. K., Gupta, M., & Sampada, G. C. (2022). A Novel Approach for Feature Selection and Classification of Diabetes Mellitus: Machine Learning Methods. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/3820360

Tanwar, A., & Bhatia, P. K. (2024). A Review on Diabetes Prediction Using Machine Learning Techniques. *Lecture Notes in Electrical Engineering*, *1185*(09), 513–524. https://doi.org/10.1007/978-981-97-1682-1_41

Tuysuzoglu, G., & Birant, D. (2020). Enhanced bagging (eBagging): A novel approach for ensemble learning. *International Arab Journal of Information Technology*, *17*(4), 515–528. https://doi.org/10.34028/iajit/17/4/10

Wang, Z., Wu, C., Zheng, K., Niu, X., & Wang, X. (2019). SMOTETomek-Based Resampling for Personality Recognition. *IEEE Access*, *7*, 129678–129689. https://doi.org/10.1109/ACCESS.2019.2940061

Wu, H., Wu, Y., Jiang, Y., Zhou, B., Zhou, H., Chen, Z., Xiong, Y., Liu, Q., & Zhang, H. (2022). ScHiCStackL: A stacking ensemble learning-based method for single-cell Hi-C classification using cell embedding. *Briefings in Bioinformatics*, *23*(1), 1–10. https://doi.org/10.1093/bib/bbab396

Yorke-smith, N., & Dumančić, S. (2023). *Model Stacking Performance Comparisons for Lifetime Estimation of CMOS ICs*. *November 2022*.

Zhang, H., Liu, C., Zhang, Z., Xing, Y., Liu, X., Dong, R., He, Y., Xia, L., & Liu, F. (2021). Recurrence Plot-Based Approach for Cardiac Arrhythmia Classification Using Inception-ResNet-v2. *Frontiers in Physiology*, *12*(May), 1–13. https://doi.org/10.3389/fphys.2021.648950