DIABETES DETECTION USING STACKING TECHNIQUE: A COMBINATION OF XGBOOST, GRADIENT BOOSTING, AND META MODEL

Aden Rahmat[1], Danang Wahyu Utomo[2]
[1,2] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, Semarang, Jawa Tengah 50131.
Email: 111202113583@mhs.dinus.ac.id[1], danang.wu@dsn.dinus.ac.id[2]

Abstract - Type 2 diabetes mellitus is a chronic and progressively increasing global health issue that necessitates early detection to mitigate serious complications such as kidney failure, neuropathy, and cardiovascular disorders. While numerous studies have developed predictive models using machine learning techniques, many are limited by their reliance on single algorithms and inadequate handling of class imbalance. This research introduces a novel strategy by employing an ensemble stacking method that integrates Gradient Boosting, XGBoost, and Random Forest, with Random Forest acting as the meta-learner. The dataset, comprising 100,000 patient records, underwent preprocessing and was balanced using the SMOTE-Tomek approach to correct class distribution disparities. The stacking process is implemented in two phases: base models generate preliminary predictions, which are subsequently used as input for the meta-model to refine the final outcomes. The evaluation demonstrates that the stacking model achieves superior performance, recording 98% accuracy and an F1-score of 0.98, outperforming the individual models. The key distinction of this study lies in the effective application of ensemble stacking to enhance prediction accuracy, especially in dealing with imbalanced and complex medical data. This methodology has the potential to improve clinical decision support systems, making them more accurate and responsive.

Keywords – Diabetes, Smote-Tomek, XgBoost, GradientBoosting, Stacking.

I. INTRODUCTION

Diabetes mellitus (DM) is a chronic metabolic disease characterised by hyperglycemia due to impaired insulin secretion, insulin resistance, or a combination of both [1]. DM is classified as a noncommunicable disease (NCD) that significantly impacts the quality of life of those affected. The most common type is type 2 diabetes mellitus (T2DM), which accounts for over 90% of cases and is often undetected in the early stages due to its mild or nonspecific symptoms [2]. Delayed diagnosis increases the risk of serious complications, such as nephropathy, neuropathy, and cardiovascular disease, and places an economic burden on the healthcare system. In recent years, machine learning (ML) approaches have increasingly been used for predicting and classifying T2DM. ML has the ability to identify patterns from large and complex datasets [3]. Algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), and XGBoost have been applied for diabetes diagnosis, with varying levels of accuracy [4]. However, several challenges remain, including data imbalance and the limitations of single models in capturing the complexity of medical data [5]. One solution to improve predictive performance is through ensemble learning, particularly the stacking ensemble technique, which combines multiple base models through a meta-learner. Several studies have shown that stacking can improve the accuracy of chronic disease prediction compared to single models [6]. However, the application of this method in T2DM prediction remains relatively limited, especially when integrating SMOTE-Tomek Links as a data balancing technique [7].

II. SIGNIFICANCE OF THE STUDY

A. Previous Research

Numerous prior studies have explored the prediction and classification of diabetes using various machine learning techniques. For instance, research conducted by Yazan Jian (2021) developed a binary classification model to predict different diabetes-related complications. This study implemented stratified 10-fold cross-validation repeated 10 times, achieving a peak accuracy of 97.8% and an F1-score of 97.7% [4]. Similarly, Usama Ahmed (2022) proposed a hybrid classification model that combined Support Vector Machine (SVM) and Artificial Neural Network (ANN) for diabetes diagnosis. The dataset was divided using a 70:30 train-test split, and the resulting model was integrated with a fuzzy logic system, yielding a prediction accuracy of 94.87% [8]. In a separate study, Genta Dwigi Sepbriant compared the performance of the K-Nearest Neighbor (KNN) and XGBoost algorithms for product classification in the e-commerce domain. XGBoost, achieved the highest accuracy of 97.17%, outperforming the KNN model and demonstrating the effectiveness of ensemble learning methods like XGBoost in enhancing model accuracy [9]. Additionally, Wu et al. (2022) found that stacking ensemble techniques provided superior disease prediction performance compared to individual models. In that study, Random Forest was utilized as the meta-model due to its robustness against overfitting and strong generalization capabilities [10]. However, many of these studies did not explicitly address the issue of class imbalance, which can lead to a prediction bias favoring the majority class and hinder the accurate detection of positive cases. To mitigate this limitation, the SMOTE-Tomek Links method was employed in this study. SMOTE generates artificial samples for the minority class, while Tomek Links remove borderline instances to enhance class separation and reduce noise. Building on these research gaps, the present study proposes a stacking ensemble model that integrates Gradient Boosting, XGBoost, and Random Forest algorithms, alongside the SMOTE-Tomek Links technique for data balancing. This approach is aimed at improving the classification accuracy of type 2 diabetes. Moreover, the findings of this study have practical implications for hospitals and clinics, as they can support earlier detection of diabetes risk, ultimately aiding in complication prevention and more effective patient management.

B. Research Method

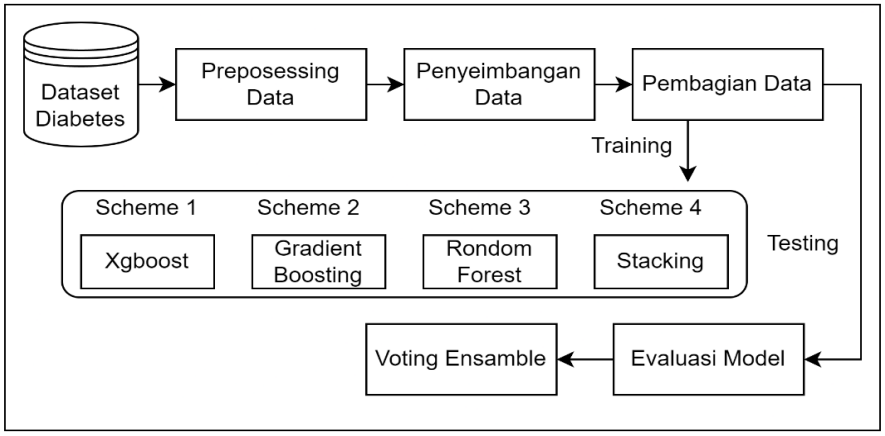

Figure 1. Research Methodology

As depicted in Figure 1, this research utilizes a machine learning framework to create a predictive model for type 2 diabetes mellitus. The process begins with an examination of the dataset to detect and handle missing values and duplicate entries. Records containing missing information are removed, and duplicate data points are eliminated to prevent bias during the training process. After cleansing, the dataset is balanced using the SMOTE-Tomek Links method to mitigate class imbalance issues. Specifically, SMOTE generates artificial samples for the minority class, whereas

Tomek Links eliminates overlapping instances between the majority and minority classes to enhance the distinction between them. This process helps the model learn from a more balanced representation of both classes. The refined dataset is then split into training (80%) and testing (20%) subsets. Model development employs a stacking ensemble technique that integrates several base learners, including Gradient Boosting and XGBoost, with Random Forest functioning as the meta-learner. This ensemble method is designed to improve predictive performance by leveraging the unique strengths of each algorithm. The effectiveness of the model is assessed through evaluation metrics such as accuracy, precision, recall, and F1-score. Ultimately, the ensemble model's results are compared with those of individual base models, highlighting significant performance gains and underscoring the advantages of ensemble learning in medical data analysis.

C. Dataset

The Diabetes Prediction Dataset comprises 100,000 patient records that include both demographic and clinical information, as well as each individual's diabetes status (positive or negative). Key attributes in the dataset consist of age, gender, body mass index (BMI), a history of hypertension, heart-related conditions, smoking habits, HbA1c readings, and blood glucose levels [11]. This dataset, sourced from the Kaggle platform and curated by Mustafa I, is highly relevant for the medical field, especially in supporting healthcare professionals in identifying individuals who may be at higher risk of developing diabetes. The data can facilitate the creation of tailored and effective treatment plans. Furthermore, it enables researchers to examine the influence of various risk factors on diabetes development, offering deeper insights into strategies for prevention and early intervention. The comprehensive nature of this dataset also makes it a reliable foundation for training machine learning models to predict diabetes outcomes based on patients' health and demographic characteristics [12].

Table I. Dataset 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

D. Data Preprocessing

Data preprocessing serves as a foundational stage in getting datasets ready for building machine learning models [13]. This process involves a series of tasks aimed at refining, restructuring, and adapting the data to meet the specific requirements of the modeling phase [14]. Initially, missing values are addressed either by discarding incomplete entries or by imputing them based on trends observed within the existing data. The dataset is then examined for duplicate records, which are removed to eliminate redundancy and minimize bias during training. Numerical features are normalized or standardized to maintain a consistent scale across variables, ensuring that differences in magnitude do not distort model performance. When encountering class imbalance, the SMOTE-Tomek Links method is utilized to equalize the class distribution. SMOTE creates synthetic samples for the underrepresented class, whereas Tomek Links eliminates borderline instances that may blur the separation between classes. Additionally, categorical features are transformed into numerical values through encoding techniques like one-hot encoding or label encoding. Lastly, the dataset is divided into training and testing portions to enable proper assessment of the model's ability to generalize. This thorough preprocessing framework is essential for improving data quality and boosting the predictive performance of machine learning models.

E. Data Balancing

This research addresses the problem of class imbalance by applying both SMOTE (Synthetic Minority Over-sampling Technique) and Tomek Links [15]. SMOTE is an oversampling approach that augments the minority class by generating artificial samples based on existing data points within that class [16]. The technique randomly selects a sample from the minority class, identifies its closest neighbors, and creates synthetic examples by interpolating between the selected sample and its neighbors using a random scaling factor. In contrast, Tomek Links is an undersampling method aimed at refining the majority class by identifying pairs of data points from different classes that are mutual nearest neighbors [17]. These paired samples, referred to as Tomek Links, are removed from the majority class to reduce overlap and enhance class separation. By integrating SMOTE with Tomek Links, the dataset becomes better balanced and cleaner, which contributes to improved model training and greater predictive reliability.

F. Data Spliting

After the preprocessing phase and the implementation of dataset balancing, the data was partitioned into training and testing sets using stratified sampling to maintain the original class ratio, following an 80:20 division. Post-SMOTE-Tomek Links application, the total sample count increased to 174,538, with 139,630 records (80%) allocated for training and 34,908(20%) reserved for testing. This segmentation distinguishes the dataset into two functional parts: one for building the model and the other for assessing its performance [18]. The 80:20 ratio allows the model to learn from a majority of the data while ensuring sufficient data remains for reliable evaluation. Stratified sampling ensures proportional representation of each class in both subsets, enabling a fair and consistent assessment of the model's predictive accuracy.

G. Ensamble Learning

Ensemble learning is a machine learning approach that integrates multiple algorithms to enhance overall model performance [19]. This strategy encompasses techniques such as bagging, boosting, and stacking. Bagging operates by generating multiple models from different subsets of the dataset, with final predictions determined through averaging or majority voting. Boosting, on the other hand, builds models in a sequential manner, where each subsequent model attempts to correct the errors of its predecessor by assigning greater weight to misclassified instances. Stacking involves combining the outputs of several base models using a meta-model, which is trained to optimise prediction accuracy by learning how to best integrate the individual model predictions. The key advantage of stacking lies in its ability to reduce both bias and variance while mitigating the risk of overfitting, particularly when applied to large and heterogeneous datasets. In this study, the stacking method was employed to predict type 2 diabetes by integrating multiple algorithms through a meta-model, thereby enhancing predictive accuracy. This ensemble approach offers improved reliability compared to relying on a single model, as it harnesses the complementary strengths of diverse base learners.

H. Boosting Technique

Boosting is a powerful ensemble technique in machine learning that enhances prediction accuracy by sequentially integrating multiple weak learners. In this approach, models are trained in a series where each subsequent learner focuses more on the instances that were previously misclassified, typically by assigning greater weights to those samples [20]. The method aims to incrementally reduce the overall error by optimizing a loss function at each step of the training process. Initially, a simple model is fitted to the data, and its performance is evaluated to identify areas of weakness. A new learner is then trained to better capture the patterns in the misclassified data points. This cycle continues either for a predetermined number of iterations or until performance gains plateau and the loss stabilizes. Research has shown that boosting is particularly effective at addressing bias in high-bias models, making it suitable for improving predictive accuracy on complex datasets that

are challenging to classify [21]. Its flexibility allows it to be paired with various base algorithms, further enhancing its utility in diverse predictive modelling tasks. By leveraging the collective strengths of multiple weak models, boosting leads to the creation of a more resilient and precise overall model.

I. XGBoost Technique

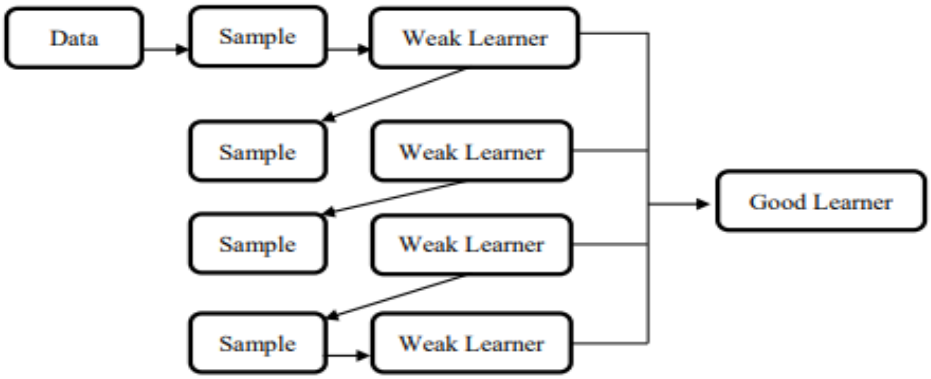

Figure 2. XGBoost Structure

As shown in Figure 2. XGBoost Structure, XGBoost is a gradient boosting framework optimized using decision trees as its basic learner, designed for high performance and efficiency. XGBoost develops a gradient boosting framework and is well suited for classification and regression problems due to its ability to find the best decision trees [22]. According to Jafarzadeh et al., XGBoost outperforms many other machine learning algorithms. Its advantages include regularization to prevent overfitting, parallel processing to accelerate the training process, and the ability to automatically handle missing values in data. This method has proven highly effective in various machine learning competitions and applications, producing accurate and efficient predictions [23].

J. Gradient Boosting

Gradient Boosting is a robust ensemble method that enhances model accuracy by sequentially combining multiple weak learners, typically decision trees. This approach builds each subsequent model to address the shortcomings of its predecessor by optimizing a loss function such as mean squared error for regression tasks or log-loss for classification problems [24]. The technique works by computing the gradient of the loss function relative to the previous predictions and incrementally updating the model to reduce prediction errors. Renowned for its ability to effectively capture complex, non-linear relationships in data, Gradient Boosting consistently delivers high predictive performance. Nonetheless, it demands careful tuning of hyperparameters including the number of estimators, maximum tree depth, and learning rate to avoid issues such as overfitting. Owing to its adaptability and high accuracy, Gradient Boosting is frequently applied in domains such as medical risk assessment, fraud detection, and the classification of complex datasets.

K. Meta Model

A meta model is a machine learning model used in the second stage of the stacking ensemble technique to combine predictions from base models. The main task of a meta model is to learn the patterns of relationships between the outputs of base models and generate more accurate final predictions [25]. Meta models play an important role in improving the overall performance of models by leveraging the advantages of several base algorithms [26]. One of the most commonly used meta models is Random Forest (RF). With its hierarchical structure, Random Forest is able to capture complex data relationships while providing predictions that are easy to interpret. Although simple, Random Forest can be optimized through parameters such as maximum depth or splitting

criteria to avoid overfitting. When used as a meta model, RF provides fast and effective predictions, making it ideal for use with moderately complex data.

L. Stacking Technique

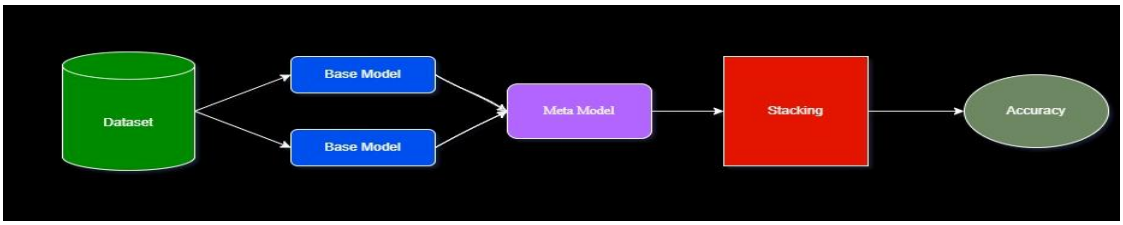

Figure 3. Stacking Structure

As illustrated in Figure 3, the stacking architecture also referred to as stacked generalization is an ensemble learning strategy that integrates multiple base learners through a secondary model known as a meta-learner or meta-classifier [27]. In this framework, the base models are initially trained on the full training dataset, and their output predictions serve as input features for training the meta-model. The underlying principle of stacking lies in the recognition that different algorithms may capture different aspects of the data. By leveraging the diverse predictions from these base models, the meta-model learns to optimize the final output by emphasizing their respective strengths and mitigating individual limitations. This layered approach often leads to improved predictive performance compared to relying on a single learning algorithm.

M. Perfomence Analisys

The confusion matrix is a widely utilized tool in machine learning and statistics for assessing the performance of classification algorithms [28]. Typically displayed in a square format, the confusion matrix summarizes the outcomes of a classification task by detailing the number of correct and incorrect predictions made by the model. While it is primarily used for binary classification problems, the confusion matrix can also be adapted to accommodate multi-class classification scenarios [29]. The matrix comprises four key elements, as illustrated in Table II: TruePositive(TP), FalsePositive(FP), TrueNegative(TN), and FalseNegative(FN). TP and TN represent instances where the model's predictions are correct, whereas FP and FN denote incorrect predictions. Among the commonly employed evaluation metrics derived from the confusion matrix, accuracy is frequently used to measure the overall correctness of the model's predictions. Accuracy is calculated using the values of TP, FP, TN, and FN, as outlined in Table II.

Table II. Confusion Matrix

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

III. RESULTS AND DISCUSSION

The initial stage of this study began with data cleaning to ensure data quality before building a type 2 diabetes prediction model. The dataset, which initially consisted of 100,000 rows, was analyzed for missing values and duplicate data.

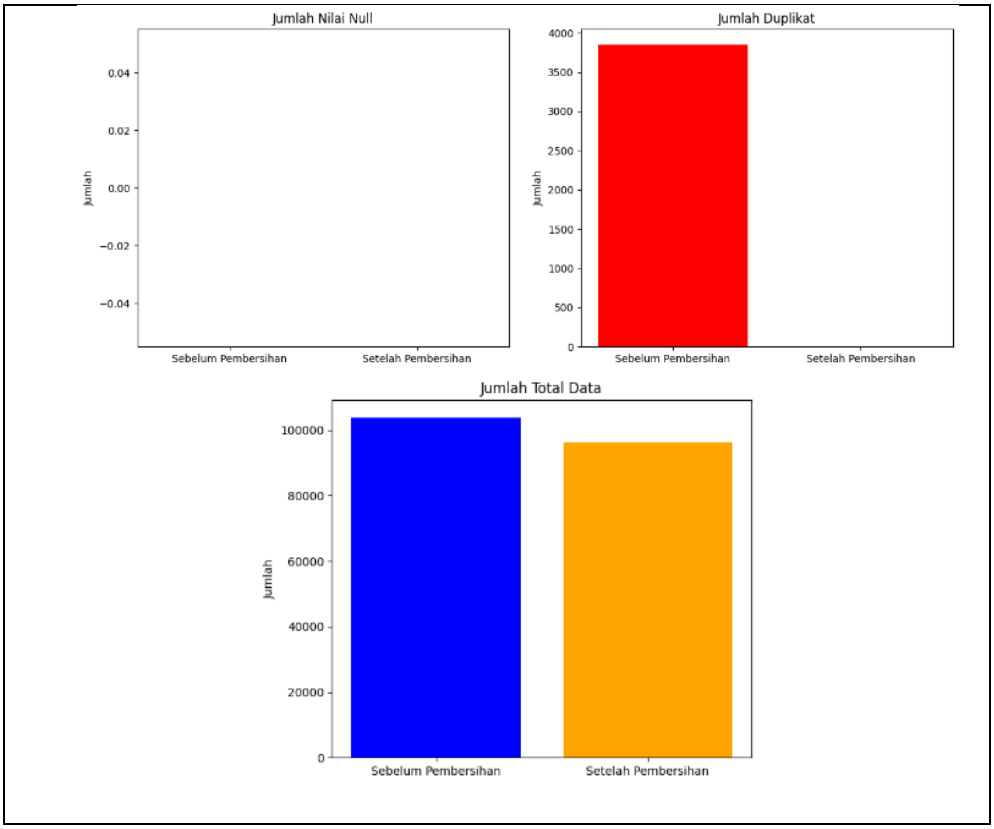

Figure 4. Data Preprocessing

as shown in Figure 4. No missing values were found across all attributes, so no imputation process was required. However, 7,708 rows were identified as duplicates and removed, leaving 96,146 unique data rows.

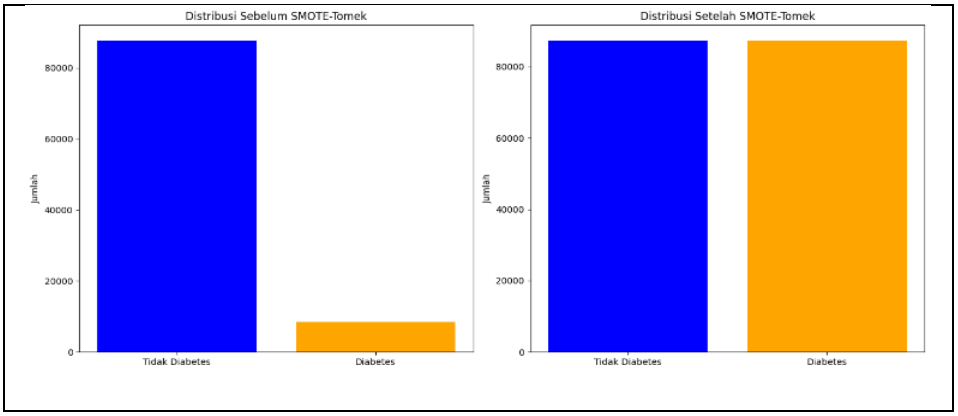

Figure 5. Diabetes Data Distribution

In addition, Figure 5 illustrates the data distribution, highlighting a notable imbalance between the non-diabetic class, which contains 87,664 instances, and the diabetic class, with only 8,482 instances. Such an imbalance can lead to model bias, where predictions are disproportionately skewed toward the majority class. To mitigate this issue, the SMOTE-Tomek technique was implemented. SMOTE (Synthetic Minority Oversampling Technique) augmented the minority class by generating synthetic samples, while Tomek Links helped clean the data by removing borderline instances from the majority class. As a result, the dataset was balanced, with each class comprising 87,269 instances. Following this adjustment, the dataset was partitioned into training and testing sets using an 80:20 stratified split to preserve class proportions across both subsets. This stratification ensures a reliable and unbiased evaluation of model performance. The study employed four machine learning models: XGBoost, Gradient Boosting, Random Forest, and a stacking ensemble model that utilized Random Forest as the meta-learner. Model performance was assessed

using a confusion matrix along with key evaluation metrics, including precision, recall, F1-score, and accuracy.

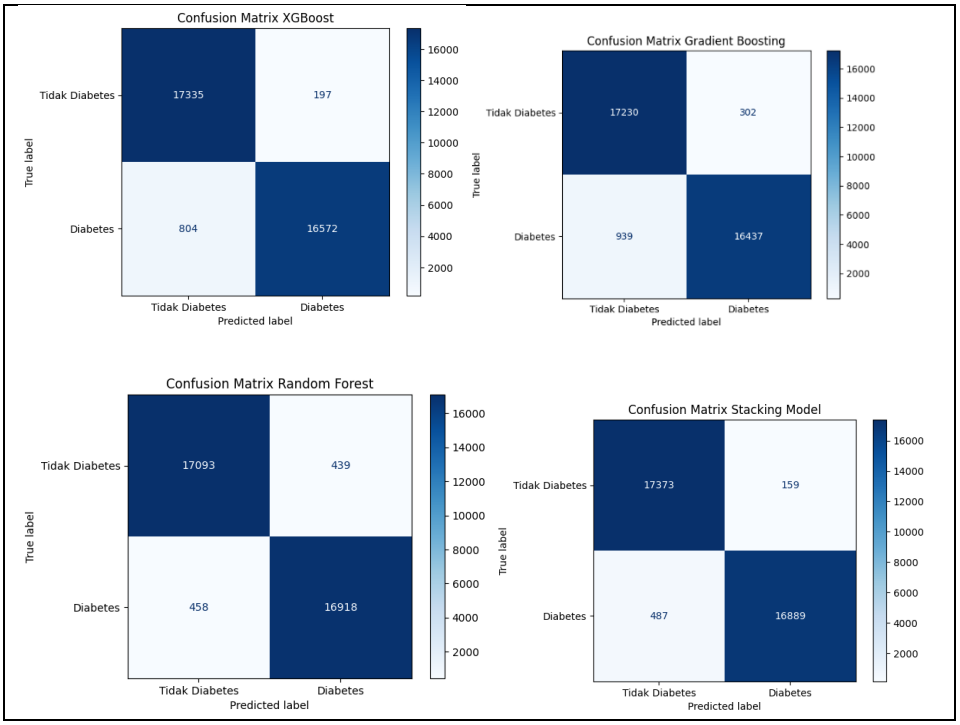

Figure 6. Confusion Matrix

The confusion matrix results Figure 6. Confusion Matrix show that the XGBoost model has strong performance with only 198 false positives and 868 false negatives. This demonstrates the model's reliability in identifying non-diabetic and diabetic patients with minimal error. However, Gradient Boosting produced 354 false positives and 956 false negatives—its performance was slightly lower, especially in classifying non-diabetic patients. Meanwhile, Random Forest showed a good balance between the two classes with 446 false positives and 542 false negatives, respectively. Although its accuracy is high, the relatively large number of False Positives indicates a potential for error in identifying healthy individuals as diabetic patients. The stacking ensemble model showed the best performance with only 166 false positives and 529 false negatives. This indicates the highest accuracy and lowest classification error compared to other models. Based on Table III, the stacking model achieved an f1-score of 0.98 for both classes, as well as an overall accuracy of 98%. This advantage stems from its ability to combine the strengths of the base models (XGBoost, Gradient Boosting, and Random Forest), enabling it to handle data variability and reduce bias from each individual model.

Table III. Evaluated Models

XGBoost	Class	Precision	Recall	F1-Score	Accuracy
	Diabetes	0.99	0.95	0.97	0.97
	Normal	0.96	0.99	0.97	0.97
Gradient Boosting	**Class**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
	Diabtes	0.98	0.95	0.96	0.96
	Normal	0.95	0.98	0.97	0.96
Rondom Forest	**Class**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
	Diabtes	0.97	0.97	0.97	0.97
	Normal	0.97	0.97	0.97	0.97
Stacking	**Class**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
	Diabtes	0.99	0.97	0.98	0.98
	Normal	0.97	0.99	0.98	0.88

From these results, it can be concluded that ensemble methods, particularly stacking, not only improve accuracy but also maintain a balance between recall and precision. This is particularly important in a medical context, where classification errors, especially False Negatives can have serious implications for patient health. Additionally, this approach has proven effective in handling initially imbalanced datasets, making it highly relevant for predicting chronic diseases such as type 2 diabetes mellitus.

IV. CONCLUSION

The experimental findings revealed that the stacking ensemble method outperformed individual modelsnamely XGBoost, Gradient Boosting, and Random Forest—in predicting type 2 diabetes mellitus. The primary strength of the stacking approach lies in its capacity to integrate the complementary advantages of multiple base learners into a cohesive prediction framework, effectively reducing bias and enhancing overall accuracy. In this research, XGBoost and Gradient Boosting acted as competent base models capable of capturing intricate data relationships, while Random Forest, employed as the meta-learner, contributed to the model's robust and reliable predictive performance. The stacking model showed superior performance by achieving 98% accuracy, with consistently high values of precision, recall, and F1-score across both classes (diabetic and non-diabetic), reaching an F1-score of 0.98. This is particularly important given the class imbalance in the dataset, where diabetic samples were underrepresented. The stacking ensemble's ability to maintain balance between the two classes while improving overall prediction accuracy makes it a more reliable and efficient approach than individual models. Overall, the application of the stacking ensemble in this research significantly enhanced prediction performance and demonstrated flexibility in handling imbalanced datasets with complex features. This approach is expected to support the development of more accurate and reliable diagnostic systems for detecting type 2 diabetes mellitus in the medical field.

REFERENCES

[1] D. S. Prawitasari, "Diabetes Melitus dan Antioksidan," *KELUWIH J. Kesehat. dan Kedokt.*, vol. 1, no. 1, pp. 48–52, 2019, doi: 10.24123/kesdok.v1i1.2496.

[2] U. Galicia-garcia, A. Benito-vicente, S. Jebari, and A. Larrea-sebal, "Costus ignus: Insulin plant and it's preparations as remedial approach for diabetes mellitus," *Int. J. Mol. Sci.*, pp. 1–34, 2020.

[3] M. I. N. Mengcan, C. Xiaofang, and X. I. E. Yongfang, "Constrained voting extreme learning machine and its application," vol. 32, no. 1, pp. 209–219, 2021, doi: 10.23919/JSEE.2021.000018.

[4] J. Flamino, R. DeVito, B. K. Szymanski, and O. Lizardo, "A Machine Learning Approach to Predicting Continuous Tie Strengths," 2021, [Online]. Available: http://arxiv.org/abs/2101.09417

[5] M. Hamama, "A Study imbalance handling by various data sampling methods in binary classification," 2021, [Online]. Available: http://arxiv.org/abs/2105.10959

[6] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 2, pp. 757–774, 2023, doi: 10.1016/j.jksuci.2023.01.014.

[7] W. Chandra, B. Suprihatin, and Y. Resti, "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction," *Symmetry (Basel).*, vol. 15, no. 4, 2023, doi: 10.3390/sym15040887.

[8] U. Ahmed *et al.*, "Prediction of Diabetes Empowered With Fused Machine Learning," *IEEE Access*, vol. 10, pp. 8529–8538, 2022, doi: 10.1109/ACCESS.2022.3142097.

[9] G. D. Sepbriant and D. W. Utomo, "Ensemble Learning pada Kategorisasi Produk E-Commerce Menggunakan Teknik Boosting," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 2, pp. 123–133, 2024, doi: 10.14421/jiska.2024.9.2.123-133.

[10] H. Wu *et al.*, "ScHiCStackL: A stacking ensemble learning-based method for single-cell Hi-C classification using cell embedding," *Brief. Bioinform.*, vol. 23, no. 1, pp. 1–10, 2022, doi: 10.1093/bib/bbab396.

[11] U. Alam, O. Asghar, S. Azmi, and R. A. Malik, "General aspects of diabetes mellitus," *Handb. Clin. Neurol.*, vol. 126, pp. 211–222, 2014, doi: 10.1016/B978-0-444-53480-4.00015-1.
[12] A. Mujumdar and V. Vaidehi, "ScienceDirect ScienceDirect ScienceDirect ScienceDirect Diabetes Prediction using Machine Learning Aishwarya Mujumdar Diabetes Prediction using Machine Learning Aishwarya Mujumdar Aishwarya," *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.
[13] T. Li, "Optimizing the configuration of deep learning models for music genre classification," *Heliyon*, vol. 10, no. 2, p. e24892, 2024, doi: 10.1016/j.heliyon.2024.e24892.
[14] M. Agnitia LEstari, M. Tabrani, and S. Ayumida, "Sistem Informasi Pengolahan Data Administrasi Kependudukan Pada Kantor Desa Pucung Karawang," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 13, no. 3, pp. 14–21, 2021, doi: 10.35969/interkom.v13i3.50.
[15] M. M. S. Fareed *et al.*, "ADD-Net: An Effective Deep Learning Model for Early Detection of Alzheimer Disease in MRI Scans," *IEEE Access*, vol. 10, pp. 96930–96951, 2022, doi: 10.1109/ACCESS.2022.3204395.
[16] Muljono, S. A. Wulandari, H. Al Azies, M. Naufal, W. A. Prasetyanto, and F. A. Zahra, "Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI," *IEEE Access*, vol. 12, no. November 2023, pp. 9292–9307, 2024, doi: 10.1109/ACCESS.2024.3353788.
[17] Z. Wang, C. Wu, K. Zheng, X. Niu, and X. Wang, "SMOTETomek-Based Resampling for Personality Recognition," *IEEE Access*, vol. 7, pp. 129678–129689, 2019, doi: 10.1109/ACCESS.2019.2940061.
[18] K. M. Kahloot and P. Ekler, "Algorithmic Splitting: A Method for Dataset Preparation," *IEEE Access*, vol. 9, pp. 125229–125237, 2021, doi: 10.1109/ACCESS.2021.3110745.
[19] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain Adaptive Ensemble Learning," *IEEE Trans. Image Process.*, vol. 30, no. 8, pp. 8008–8018, 2021, doi: 10.1109/TIP.2021.3112012.
[20] A. Manconi, G. Armano, M. Gnocchi, and L. Milanesi, "A Soft-Voting Ensemble Classifier for Detecting Patients Affected by COVID-19," *Appl. Sci.*, vol. 12, no. 15, 2022, doi: 10.3390/app12157554.
[21] P. D. A and S. Homayouni, "Bagging and Boosting Ensemble Classifiers for Classification of Comparative Evaluation," 2021.
[22] M. Kumar, S. Singhal, S. Shekhar, B. Sharma, and G. Srivastava, "Optimized Stacking Ensemble Learning Model for Breast Cancer Detection and Classification Using Machine Learning," *Sustain.*, vol. 14, no. 21, 2022, doi: 10.3390/su142113998.
[23] S. S. Azmi and S. Baliga, "An Overview of Boosting Decision Tree Algorithms utilizing AdaBoost and XGBoost Boosting strategies," *Int. Res. J. Eng. Technol.*, vol. 07, no. May, pp. 6867–6870, 2020, [Online]. Available: www.irjet.net
[24] G. Biau and B. C. L. Rouvière, "Accelerated gradient boosting," *Mach. Learn.*, vol. 108, no. 6, pp. 971–992, 2019, doi: 10.1007/s10994-019-05787-1.
[25] M. O. Olusanya, R. E. Ogunsakin, M. Ghai, and M. A. Adeleke, "Accuracy of Machine Learning Classification Models for the Prediction of Type 2 Diabetes Mellitus: A Systematic Survey and Meta-Analysis Approach," *Int. J. Environ. Res. Public Health*, vol. 19, no. 21, 2022, doi: 10.3390/ijerph192114280.
[26] N. Heama, P. Jongpradist, P. Lueprasert, and S. Suwansawat, "Investigation on tunnel responses due to adjacent loaded pile by 3D finite element analysis," *Int. J. GEOMATE*, vol. 12, no. 31, pp. 63–70, 2017, doi: 10.21660/2017.31.6542.
[27] S. N. Alexandropoulos *et al.*, "Stacking Strong Ensembles of Classifiers To cite this version : HAL Id : hal-02331304 Stacking strong ensembles of classi ers," pp. 0–12, 2019.
[28] H. Zhang *et al.*, "Recurrence Plot-Based Approach for Cardiac Arrhythmia Classification Using Inception-ResNet-v2," *Front. Physiol.*, vol. 12, no. May, pp. 1–13, 2021, doi: 10.3389/fphys.2021.648950.
[29] A. Nur, A. Thohari, A. Karima, K. Santoso, and R. Rahmawati, "Crack Detection in Building Through Deep Learning Feature Extraction and Machine Learning Approach," vol. 8, no. 1, pp. 1–6, 2024.