Journal of Applied Data Sciences
Vol. 7, No. 1, January 2026, pp. 541-553

Robust Predictive Model for Heart Disease Diagnosis Using Advanced Machine Learning Techniques

Rini Sovia[1,*], M. Khairul Anam[2], Irzal Arief Wisky[3], Randy Permana[4], Nadya Alinda Rahmi[5], Ruri hartika Zain[6]

[1,6]*Information Technology, Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung, Padang and 25145, Indonesia*

[2]*Informatics, Universitas Samudra, Jl, Prof. Dr. Syarief Thayeb, Meurandeh, Langsa and 24416, Indonesia*

[3,5]*Information System, Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung, Padang and 25145, Indonesia*

[4]*Informatics Engineering, Universitas Putra Indonesia YPTK Padang, Jl. Raya Lubuk Begalung, Padang and 25145, Indonesia*

(Received: July 10, 2025; Revised: September 5, 2025; Accepted: December 10, 2025; Available online: January 14, 2026)

Abstract

This study presents a hybrid ensemble learning framework designed to enhance the predictive accuracy, robustness, and generalizability of heart disease classification models. The framework integrates three base classifiers: Decision Tree (DT), Gaussian Naive Bayes (GNB), and K Nearest Neighbor (KNN), which are combined using a stacking ensemble method with Logistic Regression (LR) as the meta learner. Each classifier contributes a distinct analytical perspective: DT models nonlinear relationships, GNB provides probabilistic reasoning, and KNN captures similarity-based patterns. Logistic Regression aggregates their outputs to produce a unified predictive decision. To mitigate class imbalance commonly observed in clinical datasets, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic samples of the minority class, improving the model's ability to recognize underrepresented cases. Hyperparameter optimization is performed using the Optuna framework, which applies the algorithm to efficiently explore parameter configurations. The proposed model was evaluated on a publicly available heart disease dataset and achieved an accuracy of 99.61 %, precision of 99.62 %, recall of 99.59 %, F1 score of 99.60 %, and specificity of 99.58 %, corresponding to a false positive rate of only 0.42 percent. These results demonstrate the framework's strong ability to accurately identify heart disease cases while minimizing misclassification. The integration of SMOTE, stacking, and Optuna optimization contributes to its superior performance and robustness. Consequently, this approach shows strong potential for integration into clinical decision support systems to assist healthcare professionals in reliable and timely diagnosis.

Keywords: Heart Disease, Ensemble Learning, Stacking, Optuna, Classification

1. Introduction

In the development of machine learning-based prediction systems, ensemble learning has become one of the most effective strategies to improve model accuracy, stability, and generalization capability [1], [2]. Ensemble learning combines several base learners to produce a final prediction that is more robust than any single model [3], [4]. Two common ensemble methods are voting and stacking. Voting aggregates predictions from multiple models in parallel through a majority voting mechanism (hard voting) or by averaging probabilities (soft voting) [5]. In contrast, stacking combines the outputs of base learners as inputs to a meta learner, allowing the final model to be trained on a richer combination of patterns [6]. Nevertheless, despite its theoretical advantages, the application of ensemble learning in operational settings still faces several challenges.

One central issue is inconsistent performance when a model is evaluated on data whose distribution differs from the training data, often described as a generalization gap. Moreover, ensemble models that do not incorporate strategies for handling class imbalance tend to be biased toward the majority class and may fail to detect minority class patterns that are frequently more critical [7]. This problem is salient in heart disease prediction, where the number of patients

***Corresponding author: Rini Sovia (rini_sovia@upiyptk.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i1.1092

with the disease is typically smaller than the number of healthy individuals [8]. If not addressed, such imbalance can reduce the model's sensitivity to at risk patients [9]. Prior studies have shown that models with high training accuracy can experience substantial performance drops on test data, especially when proper class balancing and parameter tuning are not employed [10]. Other findings also suggest that ensemble models without SMOTE tend to favor the majority class and yield low recall for heart disease cases, limiting clinical utility [11].

Heart disease remains one of the leading causes of mortality worldwide and a major focus within health systems. Early detection is essential to prevent severe complications and to improve treatment effectiveness. Therefore, developing a heart disease prediction system that is accurate, fair, and reliable across classes is an important objective. Previous work has used single algorithms such as Decision Tree, Naive Bayes, and K Nearest Neighbor for heart disease prediction [12], [13]. Some studies have explored ensemble techniques such as voting [14], stacking [15], bagging [16], or boosting [17]. However, many of these studies did not perform systematic hyperparameter optimization, which materially affects final model performance.

Robustness in clinical implementation is defined here as the stability of model performance under nonideal data conditions. In this study, robustness is operationally specified as performance stability under three scenarios that commonly arise in clinical datasets, namely feature noise, the presence of outliers, and domain shift across sources or over time. Robustness is quantified as standardized changes in Accuracy, Precision, Recall, F1, and Specificity relative to an undisturbed test set. In addition, the False Positive Rate is reported alongside these metrics to provide a balanced characterization of diagnostic benefit and risk. Uncertainty in the estimates is conveyed through repeated training with different random seeds

To address these limitations, this study proposes a hybrid ensemble learning approach using stacking. The model is constructed using three base algorithm Decision Tree (C4.5), Gaussian Naive Bayes, and K-Nearest Neighbor as base learners, and Logistic Regression as the meta-learner in the stacking phase. To handle class imbalance, commonly found in medical datasets, the SMOTE is applied during preprocessing [18], [19]. Additionally, Optuna is used as the hyperparameter tuning algorithm to ensure that the constructed model operates under optimal configurations [20], [21].

This study focuses on structured, tabular data for heart disease prediction and evaluates the proposed approach on a publicly available dataset. Within this scope, the findings indicate that the model is a promising candidate for practical use in settings with similar data characteristics and pipelines. This recommendation is grounded in the reported metrics (Accuracy, Precision, Recall, F1, Specificity, and FPR). Future work aimed at extending these findings includes external validation on additional cohorts and observational evaluation under routine data conditions.

2. Literature Review

This review focuses on heart disease prediction using structured tabular data. The emphasis is on supervised classification methods and evaluation with standard metrics such as Accuracy, Precision, Recall, F1, Specificity, and the False Positive Rate (FPR). The goal is to synthesize representative studies, highlight methodological patterns and limitations, and motivate a hybrid stacking approach with imbalance handling and transparent optimization.

The prediction of heart disease using machine learning has been extensively studied over the past few years. Various algorithms and techniques have been proposed to improve classification accuracy, robustness, and interpretability. However, differences in data quality, feature selection, and optimization approaches often lead to varying performance across studies. To understand the current landscape of research in this area, several key studies were reviewed and summarized in table 1.

These studies cover a range of machine learning methods, from traditional single classifiers to advanced ensemble and optimization-based models. Each method brings unique strengths. For example, Support Vector Machine (SVM) is known for handling high dimensional spaces effectively, while ensemble approaches such as Random Forest (RF), CatBoost, and Bagging are designed to reduce overfitting and improve model stability. Metaheuristic and hyperparameter optimization techniques have also been introduced to enhance convergence and fine tune model parameters, ensuring that models achieve their best possible configuration on medical datasets.

The datasets used in these studies generally consist of cardiovascular or heart disease records containing clinical parameters such as age, blood pressure, cholesterol levels, and electrocardiogram results. Accuracy is typically used as the main evaluation metric, though some studies also include Recall or F1 to measure sensitivity to heart disease cases. The diversity of methods and results reflects ongoing progress and experimentation in balancing model complexity, interpretability, and generalization for medical prediction tasks.

Table 1. Summary of Previous Studies on Heart Disease Prediction

Source	Method	Dataset	Accuracy (%)	Key Finding
[22]	CatBoost Tuned	Cardiovascular	90.94	Improved feature interactions
[23]	Decision Tree	Cardiovascular	99.16	Simple interpretable model
[9]	SVM with Jellyfish Optimization	Heart Disease	98.47	Metaheuristic optimization improved convergence
[24]	Bagged with Random Forest	Heart Disease	94.34	Reduced overfitting and moderate accuracy
[25]	XGBoost with GridSearchCV	Heart Disease	87.02	Standard hyperparameter tuning
[26]	Random Forest	Heart Disease	98.60	High recall for heart patients

The results summarized in table 1 illustrate the wide performance spectrum of existing models. Reference [22] achieved 90.94 percent accuracy using a tuned CatBoost model, indicating that boosting algorithms can enhance feature interactions through iterative refinement. Reference [23] reported 99.16 percent accuracy with a Decision Tree, showing that simple and interpretable models can perform competitively on well-structured datasets. Reference [9] combined SVM with a Jellyfish Optimization Algorithm and achieved 98.47 percent accuracy, demonstrating that metaheuristic optimization can improve convergence by efficiently exploring hyperparameter spaces. Similarly, reference [24] implemented a Bagged Random Forest, obtaining 94.34 percent accuracy and confirming that ensemble bagging reduces variance and overfitting, although its generalization remains moderate compared to more complex models. [25] applied XGBoost with GridSearchCV and achieved 87.02 percent accuracy; while lower than others, this underscores the limitations of traditional grid search in exploring complex parameter spaces. In contrast, reference [26] achieved 98.60 percent accuracy using Random Forest and reported high Recall for heart patients, which is critical in screening contexts where missed positives have serious consequences. Building on the comparative evidence in table 1, recurring limitations remain in prior work, which are synthesized below to motivate the proposed hybrid stacking approach.

Although prior work on heart disease prediction using structured data has reported encouraging results, several recurring limitations are evident across studies (for example, [12], [13], [14], [15], [22], [23], [24], [25], [26]). First, many evaluations rely on a single source dataset, which restricts generalizability and makes findings sensitive to idiosyncrasies in data acquisition or preprocessing. Second, class imbalance is handled inconsistently; when applied, balancing strategies are often not coupled with transparent reporting of their configuration and impact on Recall for the minority class. Third, hyperparameter optimization is frequently under specified or narrow in scope, with incomplete disclosure of search protocols and budgets, which hinders reproducibility and fair comparison. Fourth, robustness is rarely operationalized beyond headline Accuracy; few studies quantify performance stability under nonideal conditions such as feature noise, outliers, or distribution shift, and reporting of Specificity and FPR is often incomplete. Fifth, some works emphasize overall Accuracy while providing limited analysis of error profiles such as false positives versus false negatives, which is essential for assessing diagnostic benefit and risk trade-offs. Finally, potential data handling pitfalls such as applying resampling before data partitioning or using overlapping preprocessing between training and test splits are not always ruled out explicitly, raising the possibility of optimistic estimates.

3. Methodology

Before outlining the methodological workflow employed in this study, it is important to understand the systematic approach applied to achieve optimal results. This research was designed with careful consideration of the primary challenges in healthcare data classification, particularly the issue of class imbalance commonly found in disease diagnosis datasets. To address this, a combination of preprocessing techniques, data balancing methods, machine learning algorithm selection, and hyperparameter optimization was carried out in a step-by-step and structured manner. These stages are described in detail in the following sections, starting with a dataset description, followed by data preprocessing, the application of the SMOTE, and finally, the implementation of ensemble models and optimization using Optuna. Figure 1 is the methodology used in this research.

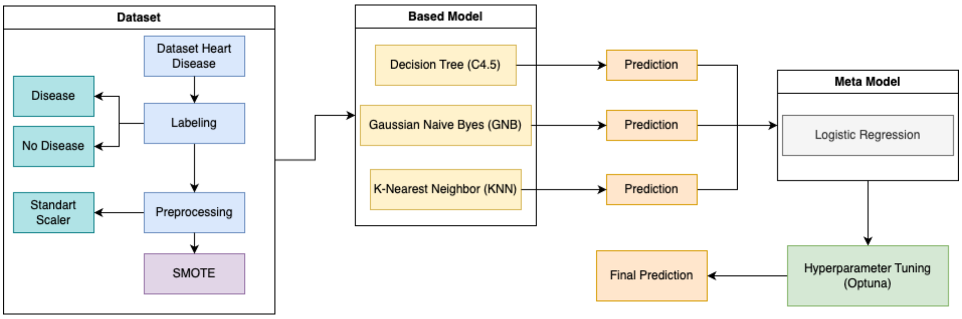

Figure 1. Methodology Flow

3.1. Dataset

The dataset used in this study was obtained from the Kaggle platform and is related to heart disease detection. It consists of 14 attributes representing various clinical information, including risk factors associated with heart disease. These attributes include age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, electrocardiogram results, maximum heart rate, exercise-induced angina, ST segment depression, ST slope, number of major vessels observed through fluoroscopy, and thalassemia test results. The main classification label in this dataset is indicated by the “target” attribute, which shows whether a patient has heart disease (1) or not (0).

This dataset was selected due to its open-access nature, clinical relevance, and widespread use in previous research as a benchmark for developing heart disease prediction models. Furthermore, the combination of numerical and categorical data allows for the flexible application of various machine learning methods. Although the class distribution in the data is relatively balanced, preprocessing steps such as normalization and class imbalance handling were still carried out to ensure optimal model training outcomes. Given these characteristics, the dataset is deemed suitable to support the research objective of building an accurate and reliable heart disease classification model.

3.2. Preprocessing

In In the preprocessing stage, numerical features were normalized to place them on a comparable scale and to prevent any single feature from dominating the learning process [27]. This study uses StandardScaler from scikit learn, which transforms each feature to have zero mean and unit variance [28]. StandardScaler is preferred because the features in the heart disease dataset come from different units and ranges, yet do not exhibit pervasive extreme outliers. Under these conditions, standardization preserves the natural spread and relative distances among samples, which is important for distance-based learners and variance sensitive models used in this study.

Compared with MinMaxScaler, standardization does not compress values into fixed bounds. MinMaxScaler can shrink between sample distances and may distort neighbourhood relations when there are new or slightly out of range values at inference time. In contrast, StandardScaler maintains a stable geometry of the feature space, which supports K Nearest Neighbors and contributes to stable optimization for Logistic Regression. Compared with RobustScaler, which centers and scales by median and interquartile range, StandardScaler is more appropriate when outliers are not dominant, since it retains information carried by variance that can be useful for classifiers that rely on distance and covariance structure.

Operationally, the scaler is fit only on the training data and then applied to the test data to avoid leakage. This step is applied after the train test split and before model fitting so that downstream procedures operate on consistently scaled variables. The use of StandardScaler also aligns with the oversampling step, since maintaining relative distances in the scaled space helps Synthetic Minority Oversampling Technique generate interpolated minority samples that are coherent with the learned geometry. Together, these considerations make StandardScaler the most suitable choice for the data characteristics and model families evaluated in this study.

Previous studies have also shown that StandardScaler achieves fairly good accuracy compared to MinMaxScaler and RobustScaler, namely 95% accuracy [29]. In comparison, studies using MinMaxScaler achieved 93% accuracy [30].

ISSN 2723-6471

Another study using RobustScaler achieved an accuracy of 85.23% [31]. These various studies were also taken into consideration when choosing StandardScaler.

In the preprocessing stage, redundancy reduction was carried out to ensure that the data remained relevant and non-repetitive. Duplicate records were removed based on identical values across all features and labels, while highly similar records were filtered by eliminating one from each pair with extremely high similarity to maintain class balance. Feature redundancy was also minimized by removing features with very low variance, eliminating one from pairs with Pearson correlation above 0.90, and checking for multicollinearity using the Variance Inflation Factor (VIF), where features with VIF values greater than 10 were iteratively removed. All processes were performed on the training data and applied consistently to the test data to prevent data leakage. These steps helped clarify the data structure and enhance model stability during the learning process.

3.3. Synthetic Minority Oversampling Technique (SMOTE)

After completing the data normalization stage, the next step is to address the issue of class imbalance in the target label. In the heart disease dataset used, there is an imbalance between patients diagnosed with heart disease and those who are not, with one class being more dominant. This condition can lead machine learning models to become biased toward the majority class, resulting in reduced performance when identifying minority class instances which are often more critical in medical contexts [32].

To overcome this issue, the Synthetic Minority Oversampling Technique was applied. This method generates new synthetic samples for the minority class through interpolation between existing samples and their nearest neighbours [33]. The oversampling process is performed only on the training set to avoid data leakage into the testing set [34]. In this study, the dataset contains 1025 samples before oversampling and 1052 samples after oversampling with Synthetic Minority Oversampling Technique.

The application of Synthetic Minority Oversampling Technique in this study aims to create a balanced distribution between the positive class patients with heart disease and the negative class patients without heart disease, allowing the model to learn fairly from both classes [35]. The resulting balance is expected to enhance the model's ability to recognize patterns from the minority class and to improve evaluation metrics such as recall and F1 score. All experiments used an 80:20 split of the dataset into training and test sets.

3.4. Base Model

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

The next stage is to build the base models using three different machine learning algorithms: Decision Tree (C4.5), GNB, and KNN. The Decision Tree algorithm works by splitting the data based on the most informative attributes and presents the decision-making process in an easily interpretable tree structure [36]. Gaussian Naive Bayes assumes that each feature is independent and normally distributed, making it a lightweight and efficient algorithm, especially suitable for medium-dimensional datasets [37]. K-Nearest Neighbor is an instance-based learning algorithm that classifies data points based on their distance to the nearest neighbors [38].

Each of these models produces predictions on the test data, which are then utilized in the subsequent ensemble process. The mathematical formulations of these base models used in this study are shown in the following Eq. (1), Eq. (2), and Eq. (3).

Decision Tree

Let $\ell(x)$ denote the leaf reached by input x, with $n\ell$ samples in that leaf and nc, ℓ samples of class $c\in\{0,1\}$. The DT estimates the class-1 probability by the class frequency in the reached leaf,

$$\hat{p}^{(DT)}(x) = \frac{n_{1,\ell(x)}}{n_{\ell(x)}}, \quad \hat{y}_1 = \left[\hat{p}^{(DT)}(x) \geq \frac{1}{2}\right] \tag{1}$$

Gaussian Naïve Bayes

With class prior π_c and, for each feature $j = 1, \dots, d$, class-conditional $x_j|y = c \sim \mathcal{N}(\mu_{cj})^2$ the posterior is

$$P(y = c|x) \propto \pi_c \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}\,\sigma_{cj}} exp\left(-\frac{(x_j - \mu_{cj})^2}{2\sigma_{cj}^2}\right), \text{Normalized so that } \sum_c P(y = c|x) = 1. Hence,$$
$$\hat{p}^{(GNB)}(x) = P(y = 1|x), \quad \hat{y}_2 = \left[\hat{p}^{(GNB)}(x) \geq \frac{1}{2}\right] \quad (2)$$

K-Nearest Neighbor

Let $\mathcal{N}_k(x)$ be the set of k nearest neighbors of x. Using distance weights $w_j = 1/(\left|\left|x - x_j\right|\right| + \in$ with small $\in > 0$,

$$\hat{p}^{(KNN)}(x) = \frac{\sum_{j\in\mathcal{N}_k(x)} w_j \; 1[y_j = 1]}{\sum_{j\in\mathcal{N}_k(x)} w_j}, \quad \hat{y}_3 = \left[\hat{p}^{(KNN)}(x) \geq \frac{1}{2}\right] \quad (3)$$

3.5. Stacking Model

Once the prediction results from each base model (Decision Tree, Gaussian Naive Bayes, and K-Nearest Neighbor) were obtained, the next step was to construct a meta-model using the stacking technique [39]. In this approach, the prediction results from each base model for a data sample x_i are arranged into a new feature vector, which is visible in Eq. (4):

$$s(x) = \left[\hat{p}^{(DT)}(x), \hat{p}^{(GNB)}(x), \hat{p}^{(KNN)}(x)\right]^T$$
With parameters β_0, β_1 (4)
$$\hat{p}^{(Stack)}(x) = \sigma\left(\beta_0 + \beta^T s(x)\right), \quad \hat{y} = 1\left[\hat{p}^{(STACK)}(x) \geq \frac{1}{2}\right]$$

$\hat{p}^{(DT)}(x), \hat{p}^{(GNB)}(x), \hat{p}^{(KNN)}(x)$ represent the predicted probabilities from the three base learners: Decision Tree, Gaussian Naïve Bayes, and K-Nearest Neighbors. $s(x)$ is a second-level feature (level-2 feature) for the meta-model. β_0 is the bias (intercept) term. $\beta = [\beta_1 + \beta_2 + \beta_3]^T$ are the weights that indicate how much influence each base model's output has in the final prediction. $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid (logistic) activation function, which converts the linear combination of inputs into a probability value between 0 and 1. $\hat{p}^{(Stack)}(x)$ represents the final probability of the positive class as predicted by the stacking model.

3.6. Optuna

To ensure the model performs optimally, hyperparameter tuning was carried out using the Optuna framework. Optuna is an automated optimization algorithm based on the Tree-structured Parzen Estimator (TPE), which efficiently searches for the best hyperparameter configuration [40].

4. Results and Discussion

4.1. Result

This study aims to develop a heart disease prediction model based on data obtained from the Kaggle platform. Before any resampling, preprocessing fitting, or model training, the dataset was split into training and test sets using an 80:20 ratio (80% for training and 20% for testing) to ensure an unbiased final evaluation. The preprocessing stage was conducted to ensure that the data was clean and ready for use in model training. All numerical features were normalized using the StandardScaler technique, which transforms the data distribution to have a mean of 0 and a standard deviation of 1. To prevent data leakage, StandardScaler was fitted on the training set only, and the learned scaling parameters were then applied to transform the test set. This process is essential to prevent features with large value ranges from dominating the model training process, especially in algorithms that are sensitive to scale, such as KNN or SVM.

One of the main challenges in this dataset is the class imbalance in the target variable, where the number of patients diagnosed with heart disease is smaller than those without the condition. To address this issue, the SMOTE was employed. SMOTE was applied only to the training set after preprocessing, while the test set was kept unchanged and was never oversampled. This technique successfully generated a balanced class distribution by adding synthetic samples to the minority class without compromising data quality. The application of SMOTE has proven effective in improving the model's performance in detecting heart disease patients, particularly in terms of recall and F1-score metrics. Figure 2 shows the results of the SMOTE test with Stacking. After SMOTE, the base learners were trained on the resampled training data, and the stacking ensemble was constructed using KNN, GNB, and DT as base learners and Logistic Regression as the meta-learner. To avoid leakage in the stacking layer, the meta-learner was trained using out-of-fold predictions generated via cross-validation on the training data, ensuring that the meta features were produced without using the same samples for both training and prediction.

```
              precision    recall  f1-score   support

           0       1.00      0.99      1.00       526
           1       0.99      1.00      1.00       526

    accuracy                           1.00      1052
   macro avg       1.00      1.00      1.00      1052
weighted avg       1.00      1.00      1.00      1052

Accuracy : 0.9961977186311787
-------------------------------------------
```

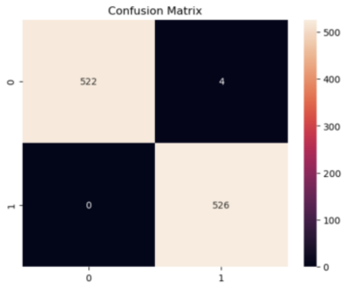

Figure 2. Testing using SMOTE + Stacking + Optuna

Figure 2 reports the test performance of the SMOTE + Stacking + Optuna configuration. Compared with SMOTE + Stacking, the optimized model achieved an accuracy of 99.61% (vs. 99.21%), with precision, recall, and F1-score close to 1.00 for both classes. From the confusion matrix, 522 out of 526 samples in class 0 were correctly classified (4 misclassified as class 1), and all 526 samples in class 1 were correctly classified (no false negatives). Although the performance is near-ceiling, this result does not by itself indicate overfitting because the 80:20 train–test split was performed before any resampling, SMOTE was applied only to the training set, and Optuna tuning and stacking training were conducted within the training data using cross-validation. All reported metrics were computed on a held-out test set that was never used during preprocessing fitting, resampling, model selection, or optimization. In addition, all preprocessing transformations were fitted only on the training data, and the test set was used only once for final evaluation, further minimizing the risk of optimistic bias. Accuracy increased from 99.21% to 99.61%. The number of misclassifications in class 0 remained unchanged (4 errors), but class 1 improved to perfect recall. The macro average and weighted average values for precision, recall, and F1-score all increased to 1.00, indicating balanced performance across both classes.

This improvement is attributed to the use of Optuna, a Bayesian-based hyperparameter optimization framework that helps find the best parameter combination for the stacking model. As a result, the integration of SMOTE + Stacking + Optuna not only balances the data distribution and combines the strengths of multiple models, but also enhances overall performance through optimal hyperparameter tuning. Importantly, hyperparameter optimization was conducted only on the training set to prevent information from the test set influencing model selection. These results demonstrate that this approach is highly effective and suitable as a machine learning-based solution for heart disease detection, offering both high accuracy and sensitivity. The following section presents a comparison with single algorithms and stacking-based approaches.

To ensure reproducibility and interpretability, this study documents the Optuna configuration as follows. Optimization used 100 trials with the TPE sampler. The search space included the number of neighbors for KNN (3–15), the maximum depth for Decision Tree (3–30), and the smoothing parameter α for Gaussian Naive Bayes (0.01–1.0, log-uniform). Trials were executed in parallel on four computational threads, and each candidate was evaluated under the same training–validation protocol used throughout the study. Specifically, Optuna evaluated each candidate using cross-validation on the training set; within each fold, preprocessing was fitted only on the fold's training portion, SMOTE was applied only to the fold's training portion (never to the validation portion), and the candidate model was scored on the untouched validation portion. After selecting the best hyperparameters, the final stacking model was refit on the full training set and evaluated on the held-out test set. These details address the previously missing computational specification and enable replication of the optimization procedure. Table 2 shows the results of all the tests carried out.

Table 2. Testing All Model

Model Algorithm	Accuracy	Precision	Recall	F1-Score
KNN	90.25%	90%	91%	91%
GNB	83.23%	83%	83%	83%
DT	86.56%	87%	86%	86%
SMOTE + KNN	96.32%	97%	96%	97%
SMOTE + GNB	85.34%	84%	84%	85%
SMOTE + DT	88.60%	88%	88%	88%
SMOTE + Stacking	99.21%	99%	99%	99%
SMOTE + KNN + Optuna	98.43%	98%	98%	98%
SMOTE + GNB + Optuna	89.88%	89%	90%	90%
SMOTE + DT + Optuna	92.43%	92%	92%	92%
SMOTE + Stacking + Optuna	99.61%	99.62%	99.59%	99.60%

Based on the overall evaluation results presented in table 2, KNN delivered the best performance, achieving an accuracy of 90.25%, precision of 90%, recall of 91%, and F1-score of 91%, indicating a strong balance between detecting positive cases and maintaining prediction accuracy. DT ranked second with an accuracy of 86.56%, precision of 87%, recall of 86%, and F1-score of 86%, reflecting stable performance though slightly lower than KNN. Meanwhile, GNB produced the lowest results, with an accuracy of 83.23% and equal values for precision, recall, and F1-score at 83%, showing this model's limitation in capturing complex data patterns. Overall, these findings indicate that KNN performs best for this dataset, DT provides consistent and reliable results, while GNB is relatively less effective in modeling nonlinear relationships within the data.

Subsequently, the dataset was balanced using SMOTE, and each algorithm was tested again, yielding varying performance levels in terms of accuracy, precision, recall, and F1-score. In general, models enhanced with Optuna for hyperparameter optimization demonstrated improved performance compared to their standard versions. The SMOTE + KNN model achieved a high accuracy of 96.32%, with both precision and F1-score reaching 97%, indicating consistent performance in identifying both positive and negative cases. In contrast, the SMOTE + GNB model yielded the lowest accuracy of 85.34%, with balanced precision and recall around 84–85%, highlighting its limitations in handling data complexity even after class balancing was applied.

After applying hyperparameter optimization using Optuna, model performance further improved. For instance, SMOTE + KNN + Optuna reached 98.43% accuracy, while SMOTE + DT + Optuna achieved 92.43%. The SMOTE + Stacking + Optuna model achieved the best overall performance, reaching 99.61% accuracy with precision, recall, and F1-score close to 1.00, indicating that stacking combined with Optuna optimization is effective for this classification task under the studied data conditions. Because the full pipeline confines preprocessing fitting, SMOTE resampling, stacking training, and Optuna optimization strictly to the training data, the reported test metrics reflect generalization rather than leakage-driven inflation.

In this study, a sensitivity analysis of the main hyperparameters for each algorithm was not yet conducted. Parameters such as the number of neighbors in KNN, the maximum depth in DT, and the smoothing value in GNB play a crucial role in determining model accuracy and stability. Without testing variations in these parameters, this research cannot fully confirm how configuration changes might influence model behavior and performance. Therefore, future studies are planned to include a comprehensive sensitivity analysis by varying key parameter values and observing their impact on performance metrics such as accuracy, precision, recall, and F1-score. This step will provide deeper insights into the robustness and consistency of each model under different settings.

Overall, this experiment confirms that the combination of SMOTE, ensemble learning, and hyperparameter optimization can improve classification performance for heart disease risk prediction. To examine whether the near-ceiling accuracy could be explained by chance, we computed a 95% confidence interval (CI) for test accuracy and compared paired predictions between SMOTE + Stacking + Optuna and the baseline SMOTE + Stacking using McNemar's test on the same test set. In the revised manuscript, we report the chi-square statistic (df = 1) and the exact p-value derived from the discordant pair counts b and c using the continuity-corrected form $X^2 = (|b - c| - 1)^2/(b + c)$, enable assessment of robustness.

4.2. Discussion

This study demonstrates the significant impact of combining multiple techniques to improve the accuracy and robustness of heart disease prediction models. One of the most critical challenges in healthcare datasets, particularly for disease diagnosis, is class imbalance. In the case of heart disease prediction, the number of patients diagnosed with the condition is smaller compared to those without it. This imbalance can result in biased models that are more likely to predict the majority class, leading to poor performance in detecting less frequent, but more critical, cases.

To address this, the SMOTE was applied. SMOTE generates synthetic samples for the minority class by interpolating between existing samples and their nearest neighbors. This approach is more advanced than simple oversampling, as it helps to ensure that the newly generated samples are realistic and contribute to the model's ability to learn the patterns of the minority class. In this study, SMOTE played a crucial role in improving the performance of the model, particularly in recall and F1-score metrics, ensuring that the model could effectively identify heart disease cases.

The Stacking ensemble method, which combined DT, GNB, and KNN as base learners with Logistic Regression as the meta-learner, further enhanced the model's predictive capability. By combining the strengths of various algorithms, Stacking allowed for more stable and accurate predictions. This is particularly important in healthcare, where models must perform consistently across different types of input data and avoid overfitting to specific patterns in the training data. The results showed that the Stacking model achieved an impressive accuracy of 99.21%, indicating that the combination of base models and a meta-learner resulted in a well-rounded and reliable prediction system.

The integration of Optuna for hyperparameter tuning proved to be another key factor in enhancing the model's performance. Optuna's TPE optimization method efficiently searches for the best combination of hyperparameters, ensuring that the model performs at its highest potential. The addition of Optuna in the SMOTE + Stacking framework improved the accuracy to 99.61%, with perfect precision, recall, and F1-score for the heart disease class. This improvement highlights the importance of hyperparameter optimization, as small adjustments in model parameters can have a significant effect on the model's ability to generalize and classify correctly.

Compared to previous studies that used different methods for heart disease prediction, this study's approach, integrating SMOTE, Stacking, and Optuna, outperformed existing models in terms of classification accuracy and sensitivity. While other studies used models like Decision Trees or Random Forest, none combined these techniques in the same systematic way. This combination of SMOTE, Stacking, and Optuna provides a powerful solution to the class imbalance problem while improving the model's overall predictive performance.

In conclusion, the results from this study demonstrate that the combination of SMOTE, Stacking, and Optuna is highly effective in tackling the challenges of class imbalance and model optimization in heart disease prediction. The model showed exceptional performance, making it a strong candidate for clinical decision support systems, where accuracy and sensitivity to the minority class are critical. The findings underscore the importance of using a holistic approach that combines data balancing, ensemble learning, and hyperparameter optimization to develop robust and reliable predictive models in healthcare. Future research can explore the application of this framework to other medical datasets or investigate alternative ensemble techniques to further enhance model performance. Table 3 presents a comparison between the results of this study and previous research.

Table 3. Testing All Model

Researcher	Dataset	Model	Accuracy
[22]	Cardiovascular Diseases	Catboost_Tuned	90.94%
[23]	Cardiovascular Diseases	Decision Tree	99.16%
[9]	Heart Diseases	SVM- Jellyfish Optimization Algorithm	98.47%
[24]	Heart Diseases	Bagged + RF	94.34%
[25]	Heart Diseases	XGBoost + GridSearch CV	87.02
[39]	Heart Diseases	Random Forest	85.25%
[26]	Heart Diseases	Random Forest	98.60%
[40]	Heart Diseases	Random Forest	95.63%
Proposed	Heart Diseases	SMOTE+Stacking	99.21%
Proposed	Heart Diseases	SMOTE + Stacking + Optuna	99.61%

ISSN 2723-6471

Table 3 presents a comparison of the performance of various machine learning algorithms used for diagnosing heart and cardiovascular diseases, based on the accuracy achieved. Studies [22] and [23] employed cardiovascular disease datasets, using the CatBoost_Tuned algorithm with an accuracy of 90.94%, and Decision Tree, which surprisingly achieved 99.16%. Meanwhile, other studies focused on heart disease datasets and applied a variety of algorithms. Study [9], which utilized SVM combined with the Jellyfish Optimization Algorithm, demonstrated strong performance with an accuracy of 98.47%. Other approaches, such as Bagged + Random Forest used in [24], yielded 94.34%, while XGBoost with GridSearchCV in [25] achieved 87.02%. Several other studies implemented Random Forest directly, with varying results: [39] recorded 85.25%, [26] achieved 98.60%, and [40] obtained 95.63%. Notably, the two proposed models in this study demonstrated the best performance: the SMOTE + Stacking model achieved 99.21%, and the combined SMOTE + Stacking + Optuna model attained the highest accuracy at 99.61%. These findings indicate that ensemble strategies and hyperparameter tuning using Optuna can significantly enhance classification accuracy for heart disease prediction compared to previous methods. The absolute difference between these two configurations is 0.40 percentage points, which constitutes a small incremental gain at a near-ceiling accuracy level. Accordingly, the role of Optuna is characterized as an optimization refinement that improves parameter selection consistency within a narrow margin, rather than a substantial performance jump.

Although the achieved accuracy is high, we acknowledge potential overfitting given the relatively small dataset and its frequent reuse in prior studies. We mitigated this risk by using a separated train and test split, fitting the scaler on training data only, applying Synthetic Minority Oversampling Technique only to the training set to avoid leakage, and keeping a consistent training protocol. Even so, confirmation of model durability requires external validation on additional cohorts, assessment under distribution shift, and evaluation in operational conditions. Accordingly, these results are interpreted as preliminary within the scope of structured data from a single source.

5. Conclusion

This study shows that combining the Synthetic Minority Oversampling Technique for class balance, ensemble learning through stacking with Decision Tree, Gaussian Naive Bayes, and K Nearest Neighbors, and hyperparameter optimization using Optuna can achieve high performance for heart disease prediction on structured data. The best configuration, SMOTE plus Stacking plus Optuna, reached 99.61 percent accuracy with very high precision, recall, and F1 score. The gain from SMOTE plus Stacking to SMOTE plus Stacking plus Optuna is incremental within a narrow range and chiefly improves consistency in parameter selection, so the result is interpreted as a moderate yet stable improvement for similar data contexts.

This study has important limitations. Evaluation was conducted on a single public dataset, so generalizability across populations and healthcare settings cannot yet be established; the findings are therefore preliminary and specific to structured data from one source. In addition, the study relies on externally sourced (non-local) secondary data rather than prospectively collected institutional data, indicating the need for further development and validation using in-house datasets that reflect local clinical practice. The model has not been tested on real time clinical streams or integrated with complex electronic health record systems, and interpretability for the stacking design remains limited in clinical environments that require transparent decision rationale.

Future research will prioritize external validation on multi center cohorts with variation in demographics, prevalence, and measurement protocols, including assessment under temporal and site related distribution shift. Complementary metrics such as ROC AUC, PR AUC, and calibration will be added to accompany accuracy, precision, recall, F1, specificity, and the false positive rate. Subsequent work will also include a dedicated interpretability study using methods such as SHAP and LIME for per feature attributions on the base models and the meta learner, the development of clinician facing explanations, and prospective testing on operational data to assess acceptability and practical benefit. In addition, we plan to incorporate rigorous cross-validation protocols in future studies to provide more robust estimates of generalization performance.

ISSN 2723-6471

6. Declarations

6.1. Author Contributions

Conceptualization: R.S. and M.K.A.; Methodology: M.K.A.; Software: I.A.W.; Validation: R.S., RHZ; Formal Analysis: R.P.; Investigation: N.A.R.; Resources: R.S.; Data Curation: M.K.A., RHZ; Writing Original Draft Preparation: I.A.W., R.S., and N.A.R.; Writing Review and Editing: R.S.; Visualization: M.K.A.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. R. Sarra, I. I. Gorial, R. R. Manea, A. E. Korial, M. Mohammed, and Y. Ahmed, "Enhanced Stacked Ensemble-Based Heart Disease Prediction with Chi-Square Feature Selection Method," *Journal of Robotics and Control (JRC),* vol. 5, no. 6, pp. 1753–1763, 2024, doi: 10.18196/jrc.v5i6.23191.

[2] M. K. Anam, M. B. Firdaus, F. Suandi, Lathifah, T. Nasution, and S. Fadly, "Performance Improvement of Machine Learning Algorithm Using Ensemble Method on Text Mining," *in ICFTSS 2024 - International Conference on Future Technologies for Smart Society, Kuala Lumpur: Institute of Electrical and Electronics Engineers Inc.,* vol. 1, no. 1, pp. 90–95, Sep. 2024. doi: 10.1109/ICFTSS61109.2024.10691363.

[3] P. Natha and P. RajaRajeswari, "Advancing Skin Cancer Prediction Using Ensemble Models," *Computers,* vol. 13, no. 7, pp. 1–24, Jul. 2024, doi: 10.3390/computers13070157.

[4] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare,* vol. 11, no. June, pp. 1-24, 2023, MDPI. doi: 10.3390/healthcare11121808.

[5] M. K. Anam, T. P. Lestari, H. Yenni, T. Nasution, and M. B. Firdaus, "Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble," *ECTI Transactions on Computer and Information Technology (ECTI-CIT),* vol. 19, no. 2, pp. 159–167, Mar. 2025, doi: 10.37936/ecti-cit.2025192.257815.

[6] L. L. Van FC, M. K. Anam, S. Bukhori, A. K. Mahamad, S. Saon, and R. L. V. Nyoto, "The Development of Stacking Techniques in Machine Learning for Breast Cancer Detection," *Journal of Applied Data Sciences,* vol. 6, no. 1, pp. 71–85, Jan. 2025, doi: 10.47738/jads.v6i1.416.

[7] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class," *Indonesian Journal of Information Systems,* vol. 7, no. 1, pp. 38–52, 2024, doi: 10.24002/ijis.v7i1.8879.

[8] S. Hadianti and W. A. G. Kodri, "Optimization of The Machine Learning Approach using Optuna in Heart Disease Prediction," *Journal Medical Informatics Technology,* vol. 1, no. 3, pp. 59–64, Sep. 2023, doi: 10.37034/medinftech.v1i3.15.

[9] A. A. Ahmad and H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm," *Diagnostics,* vol. 13, no. 14, pp. 1–17, Jul. 2023, doi: 10.3390/diagnostics13142392.

[10] A. Bustillo, R. Reis, A. R. Machado, and D. Y. Pimenov, "Improving the accuracy of machine-learning models with data from machine test repetitions," *J Intell Manuf,* vol. 33, no. 1, pp. 203–221, Jan. 2022, doi: 10.1007/s10845-020-01661-3.

[11] H. Nizam-Ozogur and Z. Orman, “A heuristic-based hybrid sampling method using a combination of SMOTE and ENN for imbalanced health data,” *Expert Syst,* vol. 41, no. 8, pp. 1–22, Aug. 2024, doi: 10.1111/exsy.13596.

[12] A. Masitha and M. Kunta Biddinika, “Preparing Dual Data Normalization for KNN Classfication in Prediction of Heart Failure,” *KLIK,* vol. 4, no. 3, pp. 1227–1234, 2023, doi: 10.30865/klik.v4i3.1382.

[13] Widiharto, M. Arief Soeleman, and A. Syukur, “Performance Improvement of Naïve Bayes Algorithm Based on Information Gain and Forward Selection Features Selection for Heart Disease Classification,” *IOSR Journal of Mobile Computing & Application,* vol. 9, no. 6, pp. 54–64, 2022, doi: 10.9790/0050-09065464.

[14] T. Karadeniz, H. H. Maraş, G. Tokdemir, and H. Ergezer, “Two Majority Voting Classifiers Applied to Heart Disease Prediction,” *Applied Sciences (Switzerland),* vol. 13, no. 6, pp. 1–13, Mar. 2023, doi: 10.3390/app13063767.

[15] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, “A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients,” *Decision Analytics Journal,* vol. 7, no. Jun., pp. 1–13, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.

[16] W. Agwil, H. Fransiska, and N. Hidayati, “Analysis of Student Graduation Timeliness Using Bagging CART,” *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika,* vol. 6, no. 2, pp. 155–166, Dec. 2020, doi: 10.24853/fbc.6.2.155-166.

[17] R. M. Syafei and D. A. Efrilianda, “Machine Learning Model Using Extreme Gradient Boosting (XGBoost) Feature Importance and Light Gradient Boosting Machine (LightGBM) to Improve Accurate Prediction of Bankruptcy,” *Recursive Journal of Informatics,* vol. 1, no. 2, pp. 64–72, Sep. 2023, doi: 10.15294/rji.v1i2.71229.

[18] M. M. R. Mubarak, Y. H. Chrisnanto, and P. N. Sabrina, “Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce,” *Enrichment: Journal of Multidisciplinary Research and Development,* vol. 1, no. 8, pp. 463–477, Nov. 2023, doi: 10.55324/enrichment.v1i8.69.

[19] N. G. Ramadhan, “Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus,” *Scientific Journal of Informatics,* vol. 8, no. 2, pp. 276–282, Nov. 2021, doi: 10.15294/sji.v8i2.32484.

[20] P. Srinivas and R. Katarya, “hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost,” *Biomed Signal Process Control,* vol. 73, no. 1, pp. 1–10, Mar. 2022, doi: 10.1016/j.bspc.2021.103456.

[21] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” *in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery,* vol. 2019, no. July., pp. 2623–2631. doi: 10.1145/3292500.3330701.

[22] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, “Advanced machine learning techniques for cardiovascular disease early detection and diagnosis,” *J Big Data,* vol. 10, no. 1, pp. 1–29, Dec. 2023, doi: 10.1186/s40537-023-00817-1.

[23] K. M. M. Uddin, R. Ripa, N. Yeasmin, N. Biswas, and S. K. Dey, “Machine learning-based approach to the diagnosis of cardiovascular vascular disease using a combined dataset,” *Intell Based Med,* vol. 7, no. Jan., pp. 1–15, Jan. 2023, doi: 10.1016/j.ibmed.2023.100100.

[24] H. Khan, A. Bilal, M. A. Aslam, and H. Mustafa, “Heart Disease Detection: A Comprehensive Analysis of Machine Learning, Ensemble Learning, and Deep Learning Algorithms,” *Nano Biomed Eng,* vol. 16, no. 4, pp. 677–690, Dec. 2024, doi: 10.26599/nbe.2024.9290087.

[25] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques,” *Algorithms,* vol. 16, no. 2, pp. 1–14, Feb. 2023, doi: 10.3390/a16020088.

[26] D. Salama AbdElminaam, N. Mohamed, H. Wael, A. Khaled, and A. Moataz, “MLHeartDisPrediction: Heart Disease Prediction using Machine Learning,” *Journal of Computing and Communication,* vol. 2, no. 1, pp. 50–65, Jan. 2023, doi: 10.21608/jocc.2023.282098.

[27] C. Liu, L. Yang, and J. Qu, “A structured data preprocessing method based on hybrid encoding,” *in Journal of Physics: Conference Series, IOP Publishing Ltd,* vol. 2021, no. January, pp. 1–9, 2021. doi: 10.1088/1742-6596/1738/1/012060.

[28] A. P. Joshi and B. V. Patel, “Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Oriental journal of computer science and technology,* vol. 13, no. 0203, pp. 78–81, Jan. 2021, doi: 10.13005/ojcst13.0203.03.

[29] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, and Z. Siddique, “Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance,” *Technologies (Basel),* vol. 9, no. 3, pp. 1–17, Sep. 2021, doi: 10.3390/technologies9030052.

[30] M. A. Latief, L. R. Nabila, W. Miftakhurrahman, S. Ma'rufatullah, and H. Tantyoko, "Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification," *International Journal of Engineering and Computer Science Applications (IJECSA),* vol. 3, no. 1, pp. 11–18, Feb. 2024, doi: 10.30812/ijecsa.v3i1.3758.

[31] G. Husain et al., "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms,* vol. 18, no. 1, pp. 1–16, Jan. 2025, doi: 10.3390/a18010037.

[32] F. P. Arifianti and A. Salam, "XGBoost and Random Forest Optimization using SMOTE to Classify Air Quality," *Advance Sustainable Science, Engineering and Technology,* vol. 6, no. 1, pp. 1–8, Nov. 2024, doi: 10.26877/asset.v6i1.18136.

[33] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences (Switzerland),* vol. 13, no. 6, pp. 1–34, Mar. 2023, doi: 10.3390/app13064006.

[34] P. Rim and E. Liu, "Optimizing the C4.5 Decision Tree Algorithm using MSD-Splitting," *(IJACSA) International Journal of Advanced Computer Science and Applications,* vol. 11, no. 10, pp. 41–47, 2020, doi: 10.14569/IJACSA.2020.0111006.

[35] T. P. Handayani, W. Hasyim, and Nursetia, "Preliminary Evaluation of Gaussian Naive Bayes for Multi-Label Hate Speech and Abusive Language Detection on Indonesian Twitter," *Journal of International Multidisciplinary Research,* vol. 1, no. 1, pp. 159–165, Nov. 2023, doi: 10.62504/jimr532.

[36] X. T. Dang and T. T. Le, "KNN-SMOTE: An Innovative Resampling Technique Enhancing the Efficacy of Imbalanced Biomedical Classification," in Machine Learning and Other Soft Computing Techniques: Biomedical and Related Applications, N. Hoang Phuong, N. T. Huyen Chau, and V. Kreinovich, Eds., Cham: Springer Nature Switzerland, vol. 2024, no. August, pp. 111–121, 2024. doi: 10.1007/978-3-031-63929-6_11.

[37] B. L. V. S. R. Krishna, V. Mahalakshmi, and G. K. M. Nukala, "A Stacking Model for Outlier Prediction using Learning Approaches," *International Journal of Intelligent Systems and Applications in Engineering,* vol. 12, no. 2s, pp. 629–638, 2023.

[38] J. P. Lai, Y. L. Lin, H. C. Lin, C. Y. Shih, Y. P. Wang, and P. F. Pai, "Tree-Based Machine Learning Models with Optuna in Predicting Impedance Values for Circuit Analysis," *Micromachines (Basel),* vol. 14, no. 2, pp. 20, Feb. 2023, doi: 10.3390/mi14020265.

[39] H. Khan and P. Singh, "Heart Disease Analysis and Prediction Using EDA and ML Classifiers," *in ACM International Conference Proceeding Series, Association for Computing Machinery,* vol. 2023, no. November, pp. 1–6, 2023. doi: 10.1145/3647444.3647857.

[40] Md. J. N. Nayeem, S. Rana, and Md. R. Islam, "Prediction of Heart Disease Using Machine Learning Algorithms," *European Journal of Artificial Intelligence and Machine Learning,* vol. 1, no. 3, pp. 22–26, Nov. 2022, doi: 10.24018/ejai.2022.1.3.13.