Optimizing Monkeypox Detection Using Advanced Class Imbalance Handling Methods: Smote, Smote-Enn, Smote-Tomek, Borderline-Smote

Fahlul Rizki[1,*], Widowati[2], Catur Edi Widodo[3]

[1,2,3]*Doctoral Program of Information System, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia*

(Received: December 5, 2025; Revised: February 1, 2026; Accepted: March 25, 2026; Available online: April 26, 2026)

Abstract

Monkeypox is a zoonotic viral disease with increasing global concern due to its rapid spread and potential public health impact. Accurate and timely detection is crucial, yet the development of machine learning-based detection systems is often challenged by class imbalance in clinical datasets, leading to biased predictions towards majority classes. This study systematically evaluates the effectiveness of various class imbalance handling techniques, including SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek, on the performance of ensemble learning algorithms, specifically Random Forest and Gradient Boosting, for monkeypox detection. Using a dataset of 25,000 synthetic patient records with 11 clinical features, models were trained and validated through stratified 5-fold cross-validation. Performance metrics including accuracy, precision, recall, F1-score, and Area Under the Curve (AUC), along with ROC analysis, were employed to assess the impact of each augmentation method. Results indicate that hybrid methods, particularly SMOTE-ENN, significantly improve recall and F1-score, improving the detection of clinically important monkeypox-positive cases while maintaining adequate discriminative ability. Standard SMOTE and SMOTE-Tomek provide stable performance across metrics, whereas Borderline-SMOTE shows lower recall despite high precision. These findings highlight the importance of selecting appropriate class imbalance handling strategies tailored to the clinical objective, emphasizing sensitivity in detecting positive monkeypox cases. The study provides practical guidance for implementing reliable and robust machine learning models in early monkeypox detection, contributing to improved clinical decision-making and public health interventions.

Keywords: Class Imbalance Handling, Data Augmentation, Monkeypox Detection, Machine Learning in Healthcare, Smote Base

1. Introduction

Monkeypox is a zoonotic viral disease caused by the monkeypox virus (MPXV), classified under the genus Orthopoxvirus in the family Poxviridae, and produces clinical manifestations similar to smallpox in humans [1], [2]. MPXV contains a double-stranded DNA genome enclosed in a brick-shaped structure that attaches to host cells via glycosaminoglycans [2] and typically infects animals such as rabbits, rodents, and non-human primates [1]. The recent global monkeypox outbreak has received considerable attention due to its rapid transmission and potential health impact [3], [4].

Monkeypox was declared a global health emergency following the largest recorded outbreak in history, with over 55,000 reported cases across 103 countries, including regions with no prior exposure [3]. On 23 July 2022, the Director-General of the World Health Organization (WHO) officially designated the monkeypox outbreak as a Public Health Emergency of International Concern (PHEIC) [4]. Consequently, reliable, fast, and accurate detection systems are essential to support disease control and management efforts [5], [6]. However, developing machine learning-based detection systems for monkeypox presents several challenges, particularly class imbalance within the datasets, which can substantially affect classification performance.

Class imbalance can degrade model performance by introducing bias towards the majority class [7], [8], [9]. This occurs because the model predominantly learns patterns from majority-class samples, resulting in poor representation of minority-class characteristics [10]. Detecting clinically important cases is critical in medical classification tasks,

***Corresponding author: Fahlul Rizki (ahlulrizki@students.undip.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i2.1286

regardless of class distribution [11]. In this study, although monkeypox-positive cases represent the majority class, improving their detection remains essential for early diagnosis and disease control.

In disease detection tasks, class imbalance may cause the model to achieve deceptively high evaluation scores while remaining biased towards the majority class [9]. Such biased models are at risk of missing minority-class cases, potentially leading to delayed diagnosis or diagnostic errors [10], [11]. These issues hinder interpretability and limit the practical applicability of predictive models in healthcare contexts [7]. Therefore, appropriate imbalance handling methods are needed to produce accurate and fair predictive outcomes [8].

To mitigate class imbalance, a variety of data-level strategies have been introduced, ranging from minority class augmentation and majority class reduction to approaches that integrate both mechanisms [11], [12]. Oversampling techniques aim to increase minority-class representation to achieve a more balanced ratio [13], undersampling mitigates class skew by reducing the influence of overrepresented classes through the removal of less relevant samples [14]. Hybrid techniques combining oversampling and undersampling achieve more balanced distributions while reducing noise, potentially improving classification performance [13], [14].

Existing studies have applied oversampling, undersampling, and hybrid strategies to address class imbalance. Study [15], employed SMOTE to generate synthetic minority samples, improving classification performance especially in macro average recall and F1-score though precision and accuracy showed limited gains. Study [16] applied Borderline-SMOTE to generate samples near decision boundaries, reducing class overlap and improving performance under class imbalance conditions.

Study [17] applied the hybrid SMOTE-ENN approach to a diabetes dataset, where SMOTE increased minority representation and ENN removed noisy majority samples. Combined with Random Forest, SMOTE-ENN yielded significantly improved classification performance compared to SMOTE alone or SVM, demonstrating its effectiveness for balancing data and enhancing disease prediction. Last study [18] applied SMOTE-Tomek to a flood forecasting dataset, improving minority class representation and removing ambiguous boundary samples, which led to higher mean F1-score and better performance under class imbalance conditions.

Although techniques such as SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek demonstrate promising results for handling class imbalance, comprehensive evaluations comparing these methods on monkeypox datasets using ensemble algorithms and imbalance-sensitive metrics remain scarce. This study therefore systematically evaluates the impact of various imbalance handling techniques on the performance of Random Forest and Gradient Boosting to identify effective and robust approaches. Random Forest and Gradient Boosting were employed because they effectively capture nonlinear dependencies among features and maintain stable performance in the presence of noisy and diverse clinical data [19], [20].

The dataset used in this study was obtained from Kaggle and consists of 25,000 samples with 11 clinical and historical features, along with a binary classification target (positive vs. negative monkeypox) with a class ratio of 1.75:1. In this dataset, monkeypox-positive cases constitute the majority class, while negative cases represent the minority class.

To address class imbalance in the monkeypox dataset, this study aims to optimize detection performance through the application and comparison of multiple imbalance handling techniques. Specifically, we evaluate the effects of oversampling methods (SMOTE and Borderline-SMOTE) and hybrid combinations of oversampling and undersampling (SMOTE-ENN and SMOTE-Tomek) on machine learning classifiers such as Random Forest [21] and Gradient Boosting [22]. This work presents a quantitative comparison of model improvements through a set of complementary performance indicators, supported by ROC curve analysis to examine classification discrimination. Furthermore, we compare the effectiveness of each imbalance handling method in improving minority-class recognition. The findings are expected to inform the selection of appropriate imbalance handling strategies for developing reliable and fair monkeypox detection systems in the presence of class imbalance.

2. Literature Review

2.1. Machine Learning for Infectious Disease Detection

Machine learning in healthcare leverages artificial intelligence to automatically analyze medical data in support of diagnosis and clinical decision-making [23], [24]. Its application in infectious disease detection improves diagnostic accuracy, supports appropriate treatment selection, and strengthens disease prevention efforts [24], [25]. Moreover, machine learning can help reduce medical costs, predict disease outbreaks, prevent infections, and improve patient quality of life [24]. Through this approach, disease detection can be performed faster and more accurately than manual methods, which are more prone to human error [21].

Studies have demonstrated that machine learning is effective for detecting infectious diseases based on clinical symptoms, including Monkeypox [23], tuberculosis, COVID-19, HIV, meningitis, malaria, and bacterial infections [26]. Features used for such predictive tasks may include medical images, clinical data, or biological variables relevant to diagnosis or disease prediction [21], [23], [27]. Despite the diversity of feature types, machine learning models can identify complex patterns across symptoms that are difficult to analyze manually, thereby improving accuracy in infectious disease detection [26].

2.2. Class Imbalance in Medical Datasets

A healthcare dataset is a collection of patient medical records containing information such as diagnoses, laboratory test results, treatments, and demographic attributes [28], [29]. Such datasets are used to train classification models to recognize patterns and relationships between features and target classes [29]. However, class imbalance in healthcare datasets can degrade model performance when detecting disease cases [9]. Moreover, class imbalance may lead machine learning models to become biased toward the majority class [7], [8]. As a result, the model becomes less sensitive to minority cases, producing seemingly good performance metrics while failing to detect critical cases effectively [30], [31]. This situation increases the risk of delayed medical intervention and can worsen health outcomes, including mortality risk [32].

Class imbalance can be addressed through oversampling, undersampling, or hybrid techniques [11], [12]. A balanced dataset enables models to produce fairer, more consistent, and more reliable predictions for medical decision-making [33]. This is essential for improving diagnostic accuracy and the quality of clinical decisions, particularly for rare or high-risk diseases [33].

2.3. Oversampling Techniques: SMOTE and Borderline-SMOTE

Oversampling increases the representation of minority classes to balance data distribution and improve model performance[13], [34]. One of the most widely used oversampling approaches is SMOTE. SMOTE is a data augmentation approach that mitigates class imbalance by creating artificial minority-class instances through interpolation among similar samples, enhancing diversity in the training data [35], [36].

A key advantage of SMOTE is its ability to generate realistic synthetic samples without simply duplicating minority data, which enhances model learning capacity [35], [37]. However, SMOTE may produce noisy samples, increase class overlap, and potentially lead to overfitting [35], [37]. To address these limitations, an extended variant of SMOTE known as Borderline-SMOTE was introduced [38]. This approach improves the quality of oversampling by focusing the synthesis process on minority samples located near class decision boundaries [38]. Borderline-SMOTE is especially relevant for monkeypox detection, as clinical symptom profiles frequently occur near decision boundaries, where high-quality synthetic samples are essential for enhancing class separability.

2.4. Hybrid Resampling Methods: SMOTE-ENN and SMOTE-TOMEK

Hybrid methods are approaches that combine two sampling techniques to address data imbalance [17]. These methods are particularly important in healthcare datasets because they enhance classification performance by mitigating class distribution bias [17]. SMOTE-ENN is a hybrid resampling strategy that integrates synthetic oversampling with instance removal to address class imbalance in classification problems [17]. This approach balances the dataset by synthesizing minority-class samples through SMOTE and subsequently eliminating noisy instances using Edited Nearest Neighbors (ENN), resulting in improved data quality, classification accuracy, and model generalization [14],

[17]. SMOTE-ENN is well-suited for imbalanced healthcare datasets as it can effectively eliminate noise and outliers [14].

SMOTE-Tomek is a hybrid resampling method that combines synthetic minority oversampling with Tomek Links to remove borderline instances between classes [39]. Through minority class augmentation and boundary refinement, SMOTE-Tomek produces a dataset with reduced overlap and improved learnability [39]. This method effectively reduces imbalance and class overlap, thereby enhancing the model's performance, particularly for minority classes [39].

2.5. Ensemble Learning for Imbalanced Classification

Ensemble learning techniques improve prediction accuracy by integrating multiple base models, among which Random Forest and Gradient Boosting are commonly applied [21]. Random Forest achieves this by combining numerous decision trees trained on randomly selected samples and feature sets, thereby reducing variance and overfitting [21]. This method is well-suited for clinical data because it can handle complex feature interactions, missing values, and data variability while maintaining strong generalization capability [21].

Gradient Boosting is another ensemble approach that constructs decision trees sequentially, where each subsequent model focuses on correcting the errors of its predecessors to improve accuracy [22]. Gradient Boosting is particularly suitable for clinical data analysis due to its ability to capture complex non-linear patterns while maintaining robust performance under imbalanced class distributions [22].

Combining ensemble learning with resampling techniques has been shown to be effective for disease prediction tasks in the presence of class imbalance [17]. This approach enables models to better learn from both majority and minority classes, thereby improving minority-class prediction performance [17]. Random Forest and Gradient Boosting offer complementary characteristics: Random Forest is robust to noise and data variability, while Gradient Boosting is more sensitive to complex minority-class patterns. Therefore, evaluating both models under different resampling techniques is essential to obtain a robust and accurate disease detection system.

2.6. Research Gap

Although previous studies have demonstrated the effectiveness of class imbalance handling techniques such as SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek in improving classification performance across various healthcare domains, research that specifically compares these techniques in the context of monkeypox detection remains limited. Moreover, most existing works tend to evaluate a single resampling technique or a single classification algorithm in isolation, without conducting a systematic comparative analysis using different ensemble algorithms and imbalance-sensitive evaluation metrics.

These limitations hinder the development of a comprehensive understanding of which imbalance handling methods are most effective and robust for supporting machine-learning-based monkeypox detection systems. Therefore, a structured comparative study involving multiple resampling techniques and ensemble models on imbalanced monkeypox datasets is needed to address this research gap.

3. Methodology

This study uses the Monkeypox Patients Dataset from Kaggle, which contains 25,000 tabular patient records with a class distribution of 15,909 positive cases and 9,091 negative cases. The dataset includes 11 clinical symptom and medical history features that serve as predictor variables for monkeypox classification. The complete sequence of the research stages is shown in figure 1.

 ISSN 2723-6471

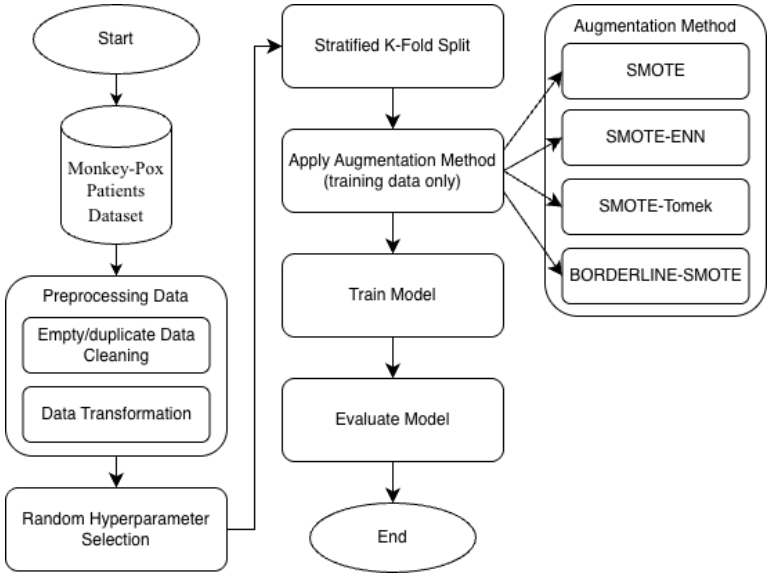

Figure 1. Workflow Diagram

Based on figure 1, each stage of the study is explained systematically to provide a clear overview of the workflow carried out in this research:

3.1. Dataset

The data used in this study were collected from an open-access dataset hosted on Kaggle and consists of 25,000 tabular patient records, comprising 15,909 (63.64%) positive monkeypox cases and 9,091 (36.36%) negative cases, with a class ratio of approximately 1.75:1. Each record represents an individual patient and contains clinical symptom information relevant to monkeypox diagnosis.

This dataset is synthetic, generated based on clinical reports published in The BMJ under the title "Clinical features and novel presentations of human monkeypox in a central London center during the 2022 outbreak: descriptive case series," with the aim of supporting research and machine learning model development. However, because the dataset is synthetically generated from reported clinical cases, the distribution of symptoms and feature correlations may not perfectly reflect real-world clinical populations. Therefore, the results of this study should be interpreted primarily as a methodological evaluation of imbalance handling techniques rather than as a clinically validated diagnostic system.

The dataset includes one categorical feature, Systemic Illness, along with several binary variables that describe clinical conditions such as rectal pain, sore throat, penile swelling, oral lesions, and other infection-related symptoms. It also contains a single target variable, MonkeyPox, which indicates whether the case is positive or negative. Therefore, this dataset can be effectively used to evaluate the performance of various clinical symptom–based classification algorithms.

3.2. Preprocessing Data

Data preprocessing constitutes a critical phase in ensuring robust machine learning performance. In this study, multiple preprocessing procedures were conducted to enhance the quality and reliability of the clinical data used for monkeypox prediction.

The initial preprocessing step involved checking the dataset for missing values and duplicate records. A total of 6,216 missing values were identified in the Systemic Illness feature, and the corresponding records were removed to ensure data consistency and prevent potential bias during model training. No duplicate records were found. Boolean symptom features were converted into numerical values (True/False to 1/0), including the target variable MonkeyPox, which was transformed from positive/negative labels into binary form. The categorical feature Systemic Illness was encoded using one-hot encoding to preserve categorical information, while the Patient_ID attribute was removed to avoid potential data leakage.

These preprocessing procedures resulted in a final dataset of 18,784 records, consisting of 12,585 positive and 6,199 negative monkeypox cases, providing a structured and consistent dataset for training the Random Forest and Gradient Boosting models.

3.3. Random Hyperparameter Selection

The Random Hyperparameter Selection stage aims to obtain an optimal configuration of the Random Forest model without exhaustively evaluating every possible hyperparameter combination. In this stage, the hyperparameter search space is first defined, including the number of decision trees (n_estimators), tree depth (max_depth), minimum number of samples required for node splitting and leaf formation, feature selection strategy, and class weighting scheme. These parameters play an essential role in controlling the learning process and enabling the model to capture complex patterns from the clinical symptom data related to monkeypox.

Instead of employing a computationally expensive exhaustive search approach such as grid search, this study applies a random search method with 25 iterations, where at each iteration one hyperparameter configuration is randomly selected from the predefined search space. The number of iterations was limited to 25 to balance computational cost and search efficiency. Preliminary experiments indicated that performance improvements became marginal after approximately 20 iterations, suggesting that 25 iterations were sufficient to obtain near-optimal hyperparameter configurations while maintaining reasonable computational cost. Previous studies have shown that random search can achieve competitive performance with relatively few iterations because it explores the search space more efficiently than exhaustive grid search [40], [41]. This approach allows efficient and effective exploration of different model configurations while reducing computational overhead. Table 1 and table 2 describe the hyperparameter search space used in this process.

Table 1. Random Forest Hyperparameter Distribution

Hyperparameter	Parameter Distribution
n_estimators	900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700
max_depth	5, 10, 20, 30, 40
min_samples_split	2, 5, 10, 15, 20
min_samples_leaf	1, 2, 4, 6, 8, 10
max_features	'sqrt', 'log2'
bootstrap	True, False
class_weight	None, 'balanced', 'balanced_subsample'
criterion	'log_loss', 'entropy'

Table 1 presents the hyperparameter distribution used during the Random Hyperparameter Selection process for the Random Forest model. Each hyperparameter was evaluated across predefined ranges to explore different model configurations, including the number of trees (n_estimators), tree depth (max_depth), minimum sample requirements, feature selection strategies (max_features), bootstrap settings, class weighting schemes, and splitting criteria. Larger values for n_estimators were included to examine whether larger ensembles could improve model stability and reduce variance when learning from clinical symptom data that may contain noise and variability. In ensemble learning, increasing the number of trees generally improves performance up to a certain point before reaching diminishing returns. Therefore, a wide range (100–1700) was considered to allow the random search procedure to identify an appropriate trade-off between predictive performance and computational cost.

Table 2. Gradient Boosting Hyperparameter Distribution

Hyperparameter	Parameter Distribution
n_estimators	900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700
learning_rate	0.01, 0.05, 0.1, 0.2
max_depth	5, 10, 20, 30, 40
min_samples_split	2, 5, 10, 15, 20
min_samples_leaf	1, 2, 4, 6, 8, 10

max_features	'sqrt', 'log2'
subsample	0.6, 0.7, 0.8, 0.9, 1.0

Table 2 presents the hyperparameter distribution used in the random search process to optimize the Gradient Boosting model. The search space includes the number of estimators, learning rate, tree depth, and various sample and feature settings to regulate model complexity. Variations in the subsample parameter were also included to enhance generalization and reduce overfitting on the clinical monkeypox data.

Stratified k-fold cross-validation was employed to evaluate each hyperparameter set, preserving class distribution in all folds. Performance metrics were averaged across folds, allowing the random hyperparameter search to select models that demonstrate both strong predictive capability and consistent behavior in monkeypox detection tasks.

3.4. Stratified K-Fold Split

At this stage, stratified k-fold cross-validation was employed to partition the dataset into multiple training and testing subsets while maintaining the original class distribution in each fold. This stratification ensures that the proportions of positive and negative monkeypox cases are consistently preserved, which is essential for unbiased model evaluation in the presence of class imbalance.

In this study, the value of k was set to 5, resulting in five stratified folds with data shuffling enabled to enhance sample variability. During each iteration, the model was trained on four folds and evaluated on the remaining fold. This process was repeated until each fold had served as the validation set exactly once, allowing performance to be assessed across diverse training–testing configurations rather than a single data split. To avoid data leakage, the resampling techniques and hyperparameter optimization were applied only to the training portion within each fold, while the validation fold remained untouched for model evaluation.

The use of five-fold stratified cross-validation contributes to more stable and representative evaluation outcomes, as performance metrics including accuracy, precision, recall, F1-score, and AUC are averaged across all folds. Consequently, observed performance differences can be attributed more reliably to the applied class imbalance handling techniques and model configurations, rather than to randomness in data partitioning. This enhances the robustness and validity of the machine learning-based monkeypox detection framework.

3.5. Apply Augmentation Method

In the Apply Augmentation Method stage, class imbalance handling techniques were applied only to the training data within each cross-validation fold, while the validation data remained unchanged. This approach ensures that synthetic samples are generated exclusively from the training subset and prevents information from the validation data from influencing the training process, thereby avoiding potential data leakage.

In this study, several data balancing techniques were applied independently, namely SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek. SMOTE generates synthetic samples for the minority class (negative cases) to balance the class distribution. Borderline-SMOTE focuses the synthesis near decision boundaries to improve discrimination in difficult regions. SMOTE-ENN and SMOTE-Tomek are hybrid methods that combine oversampling with the removal of noisy or overlapping majority samples, resulting in cleaner and more representative datasets.

Applying multiple augmentation strategies enables the model to learn minority class characteristics more effectively without discarding important majority class information. By systematically comparing models trained using each resampling strategy, this study provides a direct assessment of how effectively different approaches address class imbalance. This stage plays a crucial role in ensuring that performance improvements are truly driven by appropriate balancing strategies, particularly in the context of monkeypox detection, where accurate identification of positive cases is essential.

3.6. Training Model

In this study, the training processes for the Random Forest and Gradient Boosting models were conducted separately while following an identical workflow. This design ensures that performance differences can be attributed to the intrinsic characteristics of each algorithm rather than to variations in training procedures. By applying a consistent

 ISSN 2723-6471

training scheme, comparisons between Random Forest and Gradient Boosting can be performed fairly and objectively within the context of clinical data–based monkeypox detection.

Random Forest employs an ensemble of decision trees generated from random sample and feature selections, which enhances its ability to learn intricate clinical patterns and improves generalization. The prediction of the Random Forest model can be mathematically expressed as:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (1)$$

T represents the number of decision trees and h_t (x) denotes the prediction of the t-th decision tree. This ensemble mechanism allows the model to aggregate multiple weak learners into a more robust and stable prediction. Multiple parameter configurations were evaluated to obtain an optimal setup that improves the stability and reliability of monkeypox predictions.

In contrast, Gradient Boosting was trained iteratively, where each new model corrected the errors of its predecessor. The learning process can be formulated as:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (2)$$

F_m (x) is the updated model at iteration m, η is the learning rate, and h_m (x) is the weak learner fitted to the residual errors. This mechanism allows the algorithm to capture non-linear relationships among clinical features and progressively enhance predictive performance. By training Gradient Boosting independently from Random Forest, this study examines the performance of both algorithms under class imbalance conditions and their effectiveness in detecting monkeypox cases.

3.7. Evaluate Model

In the final phase, model performance was assessed using accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC). Precision and recall were calculated for the monkeypox-positive class, while recall corresponds to sensitivity and precision reflects prediction reliability, providing class-specific insight into model behavior under imbalanced conditions. Evaluation was conducted through stratified five-fold cross-validation to avoid dependency on a single train–test split and to better represent the models' generalization capability in capturing clinical patterns related to monkeypox.

Furthermore, Receiver Operating Characteristic (ROC) analysis was utilized to examine the classifiers' ability to discriminate between monkeypox-positive and negative cases. The AUC was used as a summary measure of discriminative performance, particularly under imbalanced class distributions. The integration of conventional performance metrics with ROC-based evaluation offers a comprehensive and reliable assessment of model robustness, especially for minority-class detection.

4. Results and Discussion

This section presents the experimental results and compares model performance for monkeypox detection under several class imbalance handling techniques, including SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek. Random Forest and Gradient Boosting were evaluated using accuracy, precision, recall, F1-score, and AUC. ROC curve analysis was also performed to assess discriminative capability. Before applying class imbalance handling techniques, baseline experiments were first conducted using the original imbalanced dataset, as presented in table 3.

Table 3. Baseline classification performance on the original imbalanced dataset

Ensemble Learning Algorithms	Matrix			
	Accuracy	Precision	Recall	F1-Score
Random Forest	0.7093	0.7154	0.9401	0.8125
Gradient Boosting	0.7036	0.7179	0.9186	0.8059

ISSN 2723-6471

As shown in table 3, both Random Forest and Gradient Boosting demonstrate high recall on the original imbalanced dataset, indicating a tendency to favor the dominant class. Although the recall values appear high, these results are influenced by the dominance of positive cases in the dataset, which may bias the models toward predicting the majority class. Consequently, the observed performance may not fully reflect the models' true discriminative capability. Therefore, class imbalance handling techniques were applied to obtain a more balanced and reliable classification performance.

4.1. Performance Comparison Using Different Augmentation Methods

This subsection presents a performance comparison of the models after applying different class imbalance handling techniques. Each augmentation method was evaluated independently to analyze its impact on the model's ability to detect monkeypox cases, particularly within the minority class. The comparison was carried out using accuracy, precision, recall, and F1-score as evaluation indicators to identify the resampling approach that most effectively improves classification performance, particularly in the presence of class imbalance. Tables 4 and 5 summarize the model performance results following the application of the various imbalance handling techniques.

Table 4. Performance Comparison of Random Forest Using Different Data Augmentation Methods

Augmentation Method	Matrix			
	Accuracy	Precision	Recall	F1-Score
SMOTE	0.6994	0.7330	0.8671	0.7945
SMOTE-ENN	0.6949	0.7010	0.9505	0.8067
SMOTE-TOMEK	0.6994	0.7330	0.8671	0.7945
BORDERLINE-SMOTE	0.6255	0.7763	0.6194	0.6887

Table 4 presents the performance comparison of the Random Forest model after applying various data augmentation methods to address class imbalance in the monkeypox dataset. Overall, the evaluation results indicate that different resampling techniques influence performance metrics in different ways, particularly in detecting monkeypox-positive cases as the clinically important class.

SMOTE and SMOTE-Tomek produced identical performance results, with an accuracy of 0.6994, F1-score of 0.7945, and recall of 0.8671, indicating that the model was able to identify most positive monkeypox cases. The identical results suggest that the Tomek Links cleaning step removed few or no borderline samples in this dataset, likely due to the relatively well-separated feature distribution after preprocessing. As a result, the SMOTE-Tomek dataset remained similar to the standard SMOTE output, leading to comparable model performance. SMOTE-ENN achieved the highest recall (0.9505) and highest F1-score (0.8067), indicating near-complete detection of positive cases, albeit with an increase in false positives. This behavior may occur because the Edited Nearest Neighbor (ENN) cleaning step removes samples that are inconsistent with their local neighborhood, which can simplify class boundaries and improve sensitivity for positive cases. However, this cleaning process may also remove informative borderline samples, potentially reducing overall class separability and slightly lowering the AUC value. Conversely, Borderline-SMOTE showed the lowest recall (0.6194), making it less effective at identifying positive cases. The relatively similar AUC values across methods suggest that performance differences are mainly driven by variations in minority class detection.

Overall, SMOTE-ENN is the most effective method for enhancing positive case detection using Random Forest, highlighting the importance of selecting appropriate imbalance handling techniques, especially in medical contexts where sensitivity is prioritized. Table 5 presents the performance evaluation of the Gradient Boosting model after applying various data augmentation techniques to address class imbalance in the monkeypox dataset. The results demonstrate that each method impacts model performance differently, particularly in detecting monkeypox-positive cases, which represent the majority class in the dataset.

SMOTE and SMOTE-Tomek demonstrated comparable performance, achieving F1-scores of approximately 0.78 with recall values around 0.82, indicating that most positive cases were successfully identified. The application of Tomek Links to remove majority-class samples did not result in a notable improvement compared to conventional SMOTE. Among the evaluated methods, SMOTE-ENN delivered the highest recall (0.95) and the strongest F1-score (0.81), reflecting near-complete detection of positive instances, albeit with an increased number of false positives. Conversely,

ISSN 2723-6471

Borderline-SMOTE exhibited substantially lower recall (0.60), suggesting reduced effectiveness in capturing positive cases. The relatively similar AUC scores across all resampling techniques indicate that the observed performance variations are largely attributable to differences in minority-class classification rather than overall discriminative capability.

Table 5. Performance Comparison of Gradient Boosting Using Different Data Augmentation Methods

Augmentation Methods	Matrix			
	Accuracy	Precision	Recall	F1-Score
SMOTE	0.6838	0.7382	0.8182	0.7761
SMOTE-ENN	0.6937	0.6997	0.9519	0.8063
SMOTE- TOMEK	0.6838	0.7382	0.8182	0.7761
BORDERLINE- SMOTE	0.6116	0.7725	0.5956	0.6726

Overall, SMOTE-ENN emerges as the most effective augmentation strategy for enhancing the sensitivity of the Random Forest model, highlighting the need to select imbalance handling techniques according to medical application goals. Results in tables 4 and 5 indicate that class imbalance handling methods have a significant impact on Random Forest's ability to detect positive monkeypox cases. SMOTE-ENN consistently produced the highest recall and F1-score, improving model sensitivity despite a slight reduction in precision. SMOTE and SMOTE-Tomek provided stable performance but were less optimal for minority class detection, while Borderline-SMOTE showed lower recall, risking missed positive cases. These findings emphasize the importance of selecting augmentation methods based on application objectives, particularly when minimizing false negatives and maximizing positive case detection are critical. To further examine class-level prediction behavior under imbalanced conditions, confusion matrices were analyzed for the SMOTE-ENN configuration, as presented in figures 2 and 3.

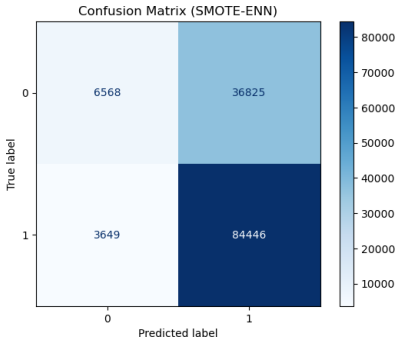

Figure 2. Confusion matrix of Random Forest with SMOTE-ENN

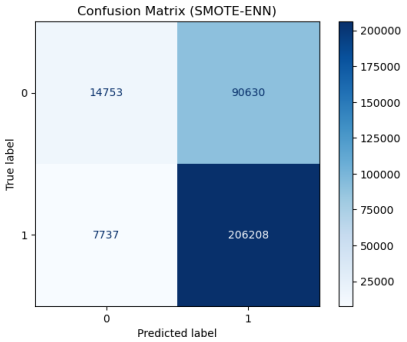

Figure 3. Confusion matrix of Gradient Boosting with SMOTE-ENN

For the Random Forest model, the confusion matrix shows 6,568 true negatives and 84,446 true positives, indicating that the model successfully detects most Monkeypox-positive cases. However, the model also produces 36,825 false positives, resulting in a specificity of approximately 0.15. Similarly, the Gradient Boosting model yields 14,753 true negatives and 206,208 true positives, but also generates 90,630 false positives, corresponding to a specificity of about 0.14. These results indicate that although SMOTE-ENN substantially improves sensitivity toward positive Monkeypox cases, it reduces specificity due to increased false positive predictions, reflecting the typical trade-off between sensitivity and specificity in imbalanced medical datasets.

4.2. ROC Curve and AUC Analysis

ROC curve and AUC analyses were performed to assess the discriminatory capability of the models across varying decision thresholds. These analyses enable a comparative evaluation of different class imbalance handling techniques in terms of their ability to balance true positive and false positive rates, thereby providing a more comprehensive assessment of model performance under class imbalance conditions. The resulting ROC curves are presented in figures 4–11.

ISSN 2723-6471

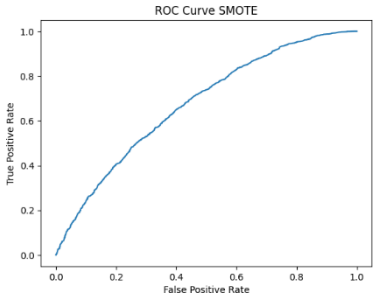

Figure 4. RF SMOTE

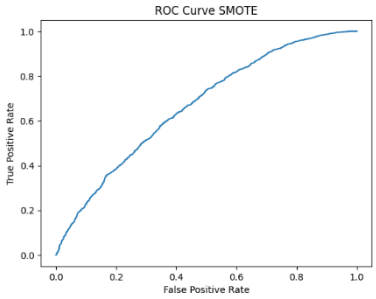

Figure 5. GB SMOTE

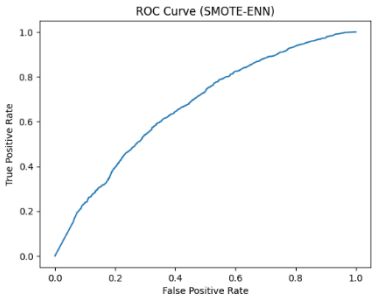

Figure 6. RF SMOTE-ENN

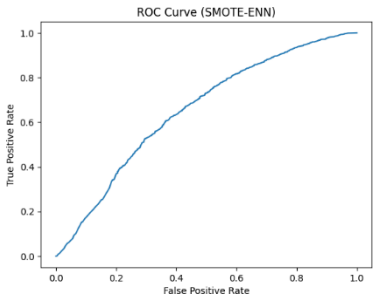

Figure 7. GB SMOTE-ENN

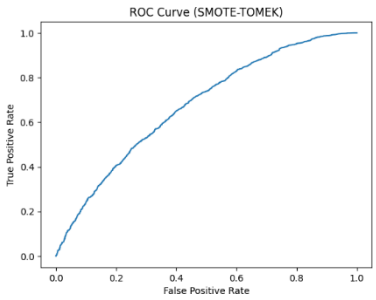

Figure 8. RF SMOTE-Tomek

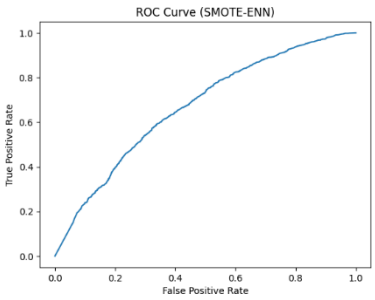

Figure 9. GB SMOTE-Tomek

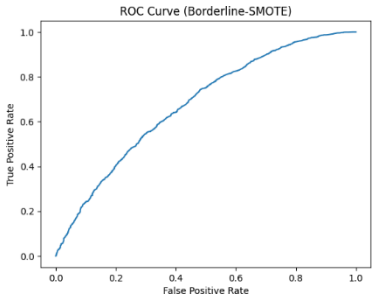

Figure 10. RF Borderline-SMOTE

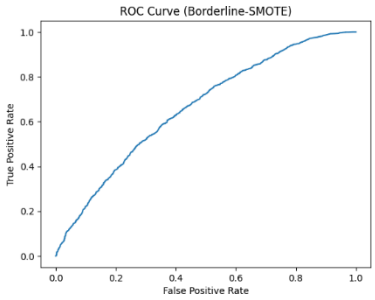

Figure 11. GB Borderline-SMOTE

Figure 4, figure 6, figure 8, and figure 10 present the ROC curves of the Random Forest model using four different class imbalance handling methods. These curves depict the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) at different decision thresholds, providing an overall assessment of the model's discriminative performance. Overall, the ROC curves for all augmentation methods lie above the diagonal reference line, indicating that the Random Forest model is able to distinguish between Monkeypox and non-Monkeypox cases better than random classification. These observations are consistent with the AUC values reported in table 6, where Random Forest achieved AUC scores ranging from 0.6742 to 0.6847. Specifically, Borderline-SMOTE produced the highest AUC value (0.6847), followed by SMOTE-ENN (0.6826), while standard SMOTE achieved 0.6742. These values indicate moderate but consistent discriminative capability across the evaluated augmentation techniques. While the AUC values remain relatively modest for clinical decision support systems, the results demonstrate that class imbalance handling techniques can still improve the model's ability to detect monkeypox cases compared to baseline learning without resampling. This suggests that the application of class imbalance handling techniques contributes positively to classification capability.

Among the evaluated methods, SMOTE-ENN demonstrates the most consistent ROC behavior across a wide range of decision thresholds, indicating improved class separability after the combination of synthetic oversampling and noise removal. However, as shown in table 6, Borderline-SMOTE produced the highest AUC value for Random Forest. This difference occurs because AUC evaluates overall class separability across all thresholds, whereas recall and F1- score emphasize the ability to correctly detect positive Monkeypox cases. Therefore, although Borderline-SMOTE slightly

improves global separability, SMOTE-ENN remains more suitable for this study because it achieves higher recall and F1-score, which are more important for detecting clinically significant cases. This hybrid mechanism helps reduce the influence of ambiguous samples and leads to more stable True Positive Rate improvements. Borderline- SMOTE also shows competitive performance, particularly at relatively low False Positive Rate levels, suggesting that the generation of synthetic samples near minority class regions can assist the Random Forest model in identifying difficult Monkeypox cases.

Standard SMOTE produced relatively consistent ROC curves, although its discriminative performance was slightly lower compared to the hybrid approaches, particularly at lower FPR levels. Meanwhile, SMOTE-Tomek provided competitive results by reducing class overlap through Tomek Links removal, resulting in performance that is comparable but generally not superior to the SMOTE-ENN configuration. Overall, the ROC analysis suggests that hybrid augmentation strategies, particularly SMOTE-ENN, provide the most reliable discriminative capability for the Random Forest model, supporting the improvements observed in the classification metrics.

Similarly, figure 5, figure 7, figure 9, and figure 11 show the ROC curves of the Gradient Boosting model for the four imbalance handling techniques. These curves illustrate the TPR–FPR relationship across varying thresholds to evaluate the model's capacity to distinguish between Monkeypox and non-Monkeypox cases. As with the Random Forest results, the ROC curves generally remain above the diagonal reference line, indicating that Gradient Boosting is capable of effectively separating the two classes when combined with imbalance handling techniques.

SMOTE-ENN again demonstrates the most stable ROC trend across different thresholds, indicating improved robustness when both oversampling and noise filtering are applied. In contrast, the AUC values reported in table 6 indicate that SMOTE and SMOTE-Tomek achieved the highest overall AUC scores for Gradient Boosting (0.6741), suggesting that these methods provide slightly better global class separability. Borderline-SMOTE shows competitive performance at lower to intermediate FPR levels, although its discriminative capability varies across thresholds.

Standard SMOTE produced balanced but slightly lower ROC performance compared to the hybrid techniques, particularly at lower FPR values. Meanwhile, SMOTE-Tomek achieved relatively consistent performance by reducing class overlap, resulting in ROC behavior that is comparable to SMOTE-ENN in several threshold regions but not consistently superior. Overall, the ROC analysis indicates that SMOTE-ENN provides the most stable discriminative performance for the Gradient Boosting model, while SMOTE-Tomek also demonstrates competitive results. These findings further support the effectiveness of hybrid augmentation strategies in improving Monkeypox detection under class imbalance conditions.

To reinforce these findings, AUC (Area Under the Curve) analysis was conducted to evaluate the overall discriminative ability of the Random Forest and Gradient Boosting models without relying on specific decision thresholds. The AUC provides a comprehensive quantitative measure of model stability and consistency for each imbalance handling technique. The AUC evaluation results for each method are summarized in table 6.

Table 6. Area Under the Curve (AUC) Comparison Using Different Class Imbalance Handling Methods

Ensemble Learning Algorithms	Augmentation Methods			
	SMOTE	SMOTE-ENN	SMOTE-TOMEK	BORDERLINE-SMOT
Random Forest	0.6843	0.6742	0.6843	0.6847
Gradient Boosting	0.6741	0.6614	0.6741	0.6679

Table 6 presents a comparison of the AUC values obtained by the Random Forest and Gradient Boosting models under different class imbalance handling methods. Overall, Random Forest yields higher and more stable AUC scores than Gradient Boosting. For Random Forest, SMOTE, SMOTE-Tomek, and Borderline-SMOTE produce relatively similar AUC values, with Borderline-SMOTE achieving the highest score (0.6847). In contrast, SMOTE-ENN yields a slightly lower AUC, even though it performs better in terms of sensitivity.

For Gradient Boosting, the AUC scores tend to be lower than those of Random Forest, with SMOTE and SMOTE-Tomek achieving the best performance (0.6741). SMOTE-ENN shows the lowest AUC for both models, indicating

that aggressive sample cleaning may reduce the model's overall discriminative ability. These findings suggest that although SMOTE-ENN can improve sensitivity toward Monkeypox cases, it does not necessarily enhance AUC performance. Therefore, the choice of augmentation technique should be aligned with the specific evaluation objectives of the model. The obtained AUC values indicate a moderate level of discriminative ability, suggesting that although the models perform better than random classification, further methodological improvements are required before considering practical clinical deployment.

4.3. Impact of Class Imbalance Handling on Minority Class Detection

Class imbalance handling is shown to improve the model's ability to detect positive monkeypox cases, which represent the majority class in the dataset. Among the evaluated methods, SMOTE-ENN produces the highest recall in both Random Forest and Gradient Boosting, while SMOTE-Tomek shows performance comparable to standard SMOTE.

However, this improvement in sensitivity comes with a reduction in precision due to higher false positives, reflecting a trade-off between sensitivity and specificity. Overall, these results indicate that SMOTE-ENN provides the most effective improvement in recall for both Random Forest and Gradient Boosting. In contrast, Borderline-SMOTE produces substantially lower recall values, suggesting reduced effectiveness in detecting Monkeypox-positive cases under imbalanced conditions.

4.4. Discussion and Practical Implications

The findings of this study confirm that class imbalance handling plays a crucial role in improving machine learning performance for monkeypox detection. Hybrid methods such as SMOTE-ENN and SMOTE-Tomek effectively enhance sensitivity toward positive cases an essential aspect in medical contexts to reduce the risk of false negatives without compromising overall class separation, as demonstrated by the ROC and AUC analyses.

From a practical standpoint, these results are relevant for the development of early detection systems for monkeypox, as appropriate data augmentation strategies can substantially improve predictive reliability in imbalanced clinical datasets. The combination of Random Forest and Gradient Boosting with hybrid augmentation methods has the potential to be integrated into medical decision support systems to help healthcare professionals identify positive cases more quickly and accurately.

Despite the promising results, several limitations remain. First, the study did not examine model calibration, which is important for evaluating probability reliability in clinical decision support. Second, feature importance analysis was not explored, limiting the interpretability of the identified symptom patterns. Finally, the robustness of the proposed approach under different imbalance ratios was not investigated, which may influence model stability in real-world datasets.

Furthermore, this study highlights that the selection of imbalance handling techniques should be aligned with the goals of the application: hybrid methods are better suited for maximizing positive case detection, whereas standard approaches such as SMOTE may be preferable when minimizing false positives is the primary objective. Statistical significance testing was not conducted in this study and is left for future work, given the synthetic nature of the dataset and its role as a benchmarking resource rather than a clinically validated dataset.

4.5. Limitations and Future Work

Although this study demonstrates that hybrid augmentation methods such as SMOTE-ENN and SMOTE-Tomek are effective for improving monkeypox detection, several limitations remain. The experiments were limited to independently trained Random Forest and Gradient Boosting models, leaving the potential of combined ensemble strategies unexplored. In addition, the dataset used in this study was synthetically generated based on reported clinical characteristics of monkeypox cases. Moreover, the dataset used in this study is synthetically generated from clinical reports, which may not fully represent real clinical variability. Synthetic generation may influence symptom distributions and correlations between features, potentially affecting model evaluation outcomes. Therefore, further validation using real-world clinical datasets is required before practical deployment.

For future work, it is recommended to explore hybrid augmentation methods within more advanced ensemble approaches, such as stacking that combines Random Forest and Gradient Boosting. Integrating class balancing

techniques with sophisticated feature selection and hyperparameter tuning strategies may also enhance both sensitivity and precision, leading to more accurate, reliable, and practically deployable monkeypox detection models.

5. Conclusion

This work investigates the impact of several class imbalance handling techniques namely SMOTE, Borderline-SMOTE, SMOTE-ENN, and SMOTE-Tomek on monkeypox detection performance using Random Forest and Gradient Boosting models. The experimental results demonstrate that hybrid resampling strategies particularly SMOTE-ENN consistently enhance recall and F1-score, thereby improving the identification of clinically important monkeypox-positive cases. This improvement is especially important in medical screening scenarios, where reducing false negatives is a critical priority. Nevertheless, the observed increase in sensitivity is accompanied by a reduction in precision and minor variations in AUC, indicating a higher incidence of false positive predictions. In contrast, SMOTE and SMOTE-Tomek yield more balanced and stable performance across evaluation metrics, whereas Borderline-SMOTE exhibits comparatively lower recall.

Overall, these findings highlight that the selection of class imbalance handling techniques should be guided by the specific objectives of the application. When maximizing sensitivity and reliable case detection is the primary concern, SMOTE-ENN emerges as the most appropriate approach. However, further validation using real-world clinical datasets is required before practical deployment in medical decision support systems. In addition, several methodological aspects were not explored in this study. Future research may incorporate model calibration analysis, feature importance evaluation, and robustness testing under different imbalance ratios to further strengthen the reliability and interpretability of the proposed approach.

6. Declarations

6.1. Author Contributions

Conceptualization: F.R., W., and C.E.W.; Methodology: F.R.; Software: F.R.; Validation: F.R., W., and C.E.W.; Formal Analysis: F.R., W., and C.E.W.; Investigation: F.R.; Resources: W.; Data Curation: W.; Writing Original Draft Preparation: F.R., W., and C.E.W.; Writing Review and Editing: W., F.R., and C.E.W.; Visualization: F.R.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The following resources used in this study are publicly accessible:

- The dataset underpinning the findings of this research is publicly accessible at https://www.kaggle.com/datasets/muhammad4hmed/monkeypox-patients-dataset.

- The Python code used in this study, which includes data preprocessing, model training, and performance evaluation for comparing Random Forest and Gradient Boosting algorithms in Monkeypox diagnosis, is publicly available at https://colab.research.google.com/drive/123aW0SqHsYD81rhizPe7UjbLnKyjhyno?usp=sharing.

6.3. Funding

This research was funded by the Indonesian Education Scholarship (BPI). The study was also supported by the High-Performance Computing Laboratory and the Research Cluster for Mathematical Modeling and Optimization, Diponegoro University, Indonesia.

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] J. Al-Dabbagh, E. Mohammad Deeb, R. Younis, and R. Eissa, "The dermatological manifestations and differential diagnosis of monkeypox: a narrative review," *Medicine*, vol. 103, no. Nov., pp. 1–12, 2024, doi: 10.1097/MD.0000000000040359.

[2] E. Lansiaux, N. Jain, S. Laivacuma, and A. Reinis, "The virology of human monkeypox virus (hMPXV): a brief overview," *Virus Res.,* vol. 322, no. Dec., pp. 1–12, 2022, doi: 10.1016/j.virusres.2022.198932.

[3] B. Cabanillas, "A compilation answering 50 questions on monkeypox virus and the current monkeypox outbreak," *Allergy,* vol. 78, no. Mar., pp. 639–662, 2023, doi: 10.1111/all.15633.

[4] N. L. Bragazzi, "Knowing the unknown: the underestimation of monkeypox cases. insights and implications from an integrative review of the literature," *Front. Microbiol.*, vol. 13, no. Sep., pp. 1–12, 2022, doi: 10.3389/fmicb.2022.1011049.

[5] A. S. Sathwik, B. Naseeba, J. C. Kiran, K. Lokesh, V. S. D. Ch, and N. P. Challa, "Early detection of monkeypox skin disease using patch based DL model and transfer learning techniques," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 9, no. May, pp. 1–9, 2023, doi: 10.4108/eetpht.9.4313.

[6] S. Nagro, "A stacked ensemble approach for symptom-based monkeypox diagnosis," *Comput. Biol. Med.*, vol. 191, no. Jun., pp. 1–12, 2025, doi: 10.1016/j.compbiomed.2025.110140.

[7] M. Meysami, "Utilizing logistic regression to compare risk factors in disease modeling with imbalanced data: a case study in vitamin D and cancer incidence," *Front. Oncol.,* vol. 13, no. Sep., pp. 1–14, 2023, doi: 10.3389/fonc.2023.1227842.

[8] S. Ul Sabha, A. Assad, N. M. U. Din, and M. R. Bhat, "Comparative analysis of oversampling techniques on small and imbalanced datasets using deep learning," *IEEE Access,* vol. 2023, no. Mar., pp. 1–10, 2023, doi: 10.1109/AISP57993.2023.10134981.

[9] E. Gentili, "Machine learning from real data: a mental health registry case study," *Comput. Methods Programs Biomed. Update*, vol. 5, no. Jan., pp. 1–12, 2024, doi: 10.1016/j.cmpbup.2023.100132.

[10] E. F. Agyemang, "Addressing class imbalance problem in health data classification: practical application from an oversampling viewpoint," *Appl. Comput. Intell. Soft Comput.,* vol. 2025, no. Feb., pp. 1–12, 2025, doi: 10.1155/acis/1013769.

[11] M. Yadav Mamilla, R. Al-Haddad, and S. Chowdhury, "Resampling imbalanced healthcare data for predictive modelling," *Int. J. Adv. Comput. Sci. Appl.,* vol. 16, no. Feb., pp. 36–44, 2025, doi: 10.14569/IJACSA.2025.0160204.

[12] R. Bounab, K. Zarour, B. Guelib, and N. Khlifa, "Enhancing medicare fraud detection through machine learning: addressing class imbalance with SMOTE-ENN," *IEEE Access*, vol. 12, no. Apr., pp. 54382–54396, 2024, doi: 10.1109/ACCESS.2024.3385781.

[13] S. Korkmaz, "Hybridization of DEBOHID with ENN algorithm for highly imbalanced datasets," *Eng. Sci. Technol. Int. J.,* vol. 63, no. Mar., pp. 1–12, 2025, doi: 10.1016/j.jestch.2025.101976.

[14] R. Bounab, B. Guelib, and K. Zarour, "A novel machine learning approach for handling imbalanced data: leveraging SMOTE-ENN and XGBoost," *IEEE Access,* vol. 2024, no. Apr., pp. 1–10, 2024, doi: 10.1109/PAIS62114.2024.10541220.

[15] M. S. N. M. Danuri, R. A. Rahman, I. Mohamed, and A. Amin, "The improvement of stress level detection in Twitter: imbalance classification using SMOTE," *IEEE Access,* vol. 2022, no. Nov., pp. 294–298, 2022, doi: 10.1109/ICOCO56118.2022.10031684.

[16] Q. Dong and W. Lu, "Imbalance data classification method based on improved SMOTE algorithm and granular computing," *IEEE Access*, vol. 2022, no. Jul., pp. 3196–3201, 2022, doi: 10.23919/CCC55666.2022.9902406.

[17] H. Hairani and D. Priyanto, "A new approach of hybrid sampling SMOTE and ENN to the accuracy of machine learning methods on unbalanced diabetes disease data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. Aug., pp. 585–590, 2023, doi: 10.14569/IJACSA.2023.0140864.

[18] A. H. Zuhairi, F. Yakub, M. Omar, M. Sharifuddin, K. A. Razak, and A. Faruq, "Imbalanced flood forecast dataset resampling using SMOTE-Tomek link," *arXiv,* vol. 2024, no. Oct., pp. 845–850, 2024, doi: 10.5109/7323359.

[19] A. M. A. Babai, O. A. Ehinola, O. I. M. F. A. Gebbayin, and M. A. E. Ibrahim, "Clastic facies classification using machine learning-based algorithms: a case study from Rawat Basin, Sudan," *Energy Geosci.,* vol. 6, no. Jan., pp. 1–12, 2025, doi: 10.1016/j.engeos.2024.100353.

[20] D. Dhabliya, M. R. Borade, N. F. Rizvi, A. Dhablia, V. D. Ghonge, and S. D. Sirsat, "Utilizing data mining techniques to develop accurate predictive models for cardiovascular disease risk assessment and early detection," *IEEE Access,* vol. 2023, no. Nov., pp. 1–6, 2023, doi: 10.1145/3647444.3647927.

[21] M. Maindola, "Utilizing random forests for high-accuracy classification in medical diagnostics," *IEEE Access,* vol. 2024, no. Sep., pp. 1679–1685, 2024, doi: 10.1109/IC3I61595.2024.10828609.

[22] N. Sadhasivam, J. Harish, and M. Bharanidharan, "Diabetes diagnosis using machine learning," *J. Trends Comput. Sci. Smart Technol.,* vol. 5, no. Dec., pp. 388–400, 2023, doi: 10.36548/jtcsst.2023.4.002.

[23] S. Rampogu, "A review on the use of machine learning techniques in monkeypox disease prediction," *Sci. One Health*, vol. 2, no. Jan., pp. 1–12, 2023, doi: 10.1016/j.soh.2023.100040.

[24] R. Dhanalakshmi, N. Vijayaraghavan, S. Narasimhan, and R. Rajesh, "A demonstrative study on prediction of disease using Bayesian machine learning techniques," *IEEE Access,* vol. 2023, no. Nov., pp. 1–10, 2023, doi: 10.1109/ICSCAN58655.2023.10395684.

[25] C. Xu, L. Y. Zhao, C. S. Ye, K. C. Xu, and K. Y. Xu, "The application of machine learning in clinical microbiology and infectious diseases," *Front. Cell. Infect. Microbiol*., vol. 15, no. May, pp. 1–12, 2025, doi: 10.3389/fcimb.2025.1545646.

[26] J. Aalam, S. N. Ahmad Shah, and R. Parveen, "An extensive review on infectious disease diagnosis using machine learning techniques and next generation sequencing: state-of-the-art and perspectives," *Comput. Biol. Med.*, vol. 189, no. Mar., pp. 1–12, 2025, doi: 10.1016/j.compbiomed.2025.109962.

[27] S. Paul, P. Ranjan, S. Kumar, and A. Kumar, "Disease predictor using random forest classifier," *IEEE Access*, vol. 2022, no. Jan., pp. 1–4, 2022, doi: 10.1109/ICONAT53423.2022.9726023.

[28] S. N. Turhan, "Leveraging graph databases for enhanced healthcare data management: a performance comparison study," *IEEE Access*, vol. 2023, no. Dec., pp. 5007–5013, 2023, doi: 10.1109/BigData59044.2023.10386297.

[29] M. Moradi and J. Hamidzadeh, "Handling class imbalance and overlap with a hesitation-based instance selection method," *Knowl. Based Syst*., vol. 294, no. Jun., pp. 1–12, 2024, doi: 10.1016/j.knosys.2024.111745.

[30] J. P. van Zyl and A. P. Engelbrecht, "Analysis of classification metric behaviour under class imbalance," *Egypt. Inform. J.,* vol. 31, no. Sep., pp. 1–12, 2025, doi: 10.1016/j.eij.2025.100711.

[31] J. Fieggen, "Navigating severe class imbalance in population cohort data," *IEEE Access,* vol. 2025, no. Jul., pp. 1–6, 2025, doi: 10.1109/EMBC58623.2025.11254293.

[32] G. Alhyasat and B. Alhijawi, "Experimental evaluation of resampling techniques for imbalanced stroke data," *IEEE Access,* vol. 2025, no. Apr., pp. 233–238, 2025, doi: 10.1109/ICTCS65341.2025.10989344.

[33] N. Meenakshisundaram and G. Sajiv, "Evaluating oversampling strategies for imbalanced cervical cancer risk prediction: a comparative analysis of SMOTE, borderline-SMOTE and ADASYN," *IEEE Access,* vol. 2025, no. Aug., pp. 1–10, 2025, doi: 10.1109/NMITCON65824.2025.11188220.

[34] N. I. Fardana, R. R. Isnanto, and O. D. Nurhayati, "Pneumothorax detection system in thoracic radiography images using CNN method," *Sci. J. Inform.,* vol. 11, no. Jan., pp. 981–990, 2025, doi: 10.15294/sji.v11i4.16635.

[35] H. Hairani, T. Widiyaningtyas, and D. Dwi Prasetya, "Addressing class imbalance of health data: a systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies," *J. Inform. Vis*., vol. 8, no. Sep., pp. 1310–1318, 2024, doi: 10.62527/joiv.8.3.2283.

[36] A. Gocer and M. O. Cingiz, "Improving breast cancer detection with upsampling and resizing algorithms in image processing," *IEEE Access,* vol. 2024, no. Jun., pp. 1–10, 2024, doi: 10.1109/ECAI61503.2024.10607503.

[37] Z. Xu, D. Shen, Y. Kou, and T. Nie, "A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification," *IEEE Trans. Neural Netw. Learn. Syst*., vol. 35, no. Mar., pp. 3740–3753, 2024, doi: 10.1109/TNNLS.2022.3197156.

[38] H. Insan, S. S. Prasetiyowati, and Y. Sibaroni, "SMOTE-LOF and borderline-SMOTE performance to overcome imbalanced data and outliers on classification," *IEEE Access,* vol. 2023, no. Dec., pp. 136–141, 2023, doi: 10.1109/ICICyTA60173.2023.10428902.

[39] K. Omari and A. Oukhatar, "Advanced phishing website detection with SMOTETomek-XGB: addressing class imbalance for optimal results," *Procedia Comput. Sci.*, vol. 252, no. Jan., pp. 289–295, 2025, doi: 10.1016/j.procs.2024.12.031.

[40] N. I. Fardana, R. R. Isnanto, and O. D. Nurhayati, "Handling class imbalance in health datasets: a comparative study of SMOTE and SMOTEENN with TabNet," *IEEE Access,* vol. 2025, no. Oct., pp. 305–310, 2025, doi: 10.1109/ICICoS68590.2025.11329876.

[41] D. Fania, I. Waspada, and H. A. Wibawa, "Addressing data limitations in cardiovascular disease prediction: integration of public databases and clinical records," *IEEE Access,* vol. 2026, no. Jan., pp. 293–298, 2026, doi: 10.1109/icicos68590.2025.11329869.