Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 2 (2026) 528 - 536 e-ISSN: 2580-0760

An Adaptive Feature-Aware Hybrid Resampling Strategy for Imbalanced Diabetes Classification with Integrated Balanced Index Evaluation

Jasmir Jasmir[1], Riza Pahlevi[2], Gunardi Gunardi[3], Eni Rohaini[4], Tiko Nur Annisa[5]
[1,5]Master's Program in Information System, Faculty of Computer Science, Universitas Dinamika Bangsa, Jambi, Indonesia
[2,4]Department of Informatics Engineering, Faculty of Computer Science, Universitas Dinamika Bangsa, Jambi, Indonesia
[3]Department of Information System, Faculty of Computer Science, Universitas Dinamika Bangsa, Jambi, Indonesia

[*]*ijay_jasmir@yahoo.com*

Abstract

Class imbalance remains a critical challenge in medical data classification, particularly in diabetes prediction, as it significantly degrades minority-class sensitivity. This study proposes an Adaptive Feature-Aware Hybrid Resampling Strategy (AHRS) that dynamically integrates oversampling and undersampling based on Imbalance Ratio (IR) and Feature Importance (FI). Unlike conventional static resampling methods, AHRS iteratively adjusts class distribution while preserving informative feature structures. In addition, this study introduces the Integrated Balanced Index (IBI), a bounded composite metric integrating precision, recall, and specificity to provide a fairer evaluation of classification performance on imbalanced medical datasets. The proposed approach was evaluated using the Pima Indian Diabetes Dataset (768 instances) with K-Nearest Neighbor, Naïve Bayes, and Random Forest classifiers under 5-fold stratified cross-validation. Experimental results demonstrate that AHRS consistently outperforms SMOTE, Random Oversampling, and Tomek Links, achieving accuracy improvements of 5–7% and recall gains of up to 10%. Random Forest combined with AHRS achieved the highest IBI score of 0.90, indicating strong balance between sensitivity and specificity. The findings suggest that adaptive, feature-aware resampling combined with balanced evaluation metrics provides a reliable and interpretable framework for fair medical classification systems and Clinical Decision Support Systems (CDSS).

Keywords: adaptive hybrid resampling; diabetes; imbalanced data; integrated balanced index; machine learning

How to Cite: J. Jasmir, R. Pahlevi, G. Gunardi, E. Rohaini, and T. N. Annisa, "An Adaptive Feature-Aware Hybrid Resampling Strategy for Imbalanced Diabetes Classification with Integrated Balanced Index Evaluation", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 2, pp. 528 - 536, Apr. 2026.
Permalink/DOI: https://doi.org/10.29207/resti.v10i2.7418

Received: December 25, 2025
Accepted: April 19, 2026
Available Online: April 30, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Class imbalance continues to be a major challenge in machine learning classification, especially in medical diagnosis, where errors in identifying minority-class instances can result in serious consequences such as delayed treatment or missed disease detection. In the context of diabetes prediction, positive cases are usually far fewer than non-diabetic cases, leading classifiers to favor the majority class and consequently reducing sensitivity. Therefore, handling class imbalance is crucial to develop predictive models that are both reliable and equitable.

Recent advances in machine learning (ML) have enabled automated medical diagnosis systems across various healthcare domains. Algorithms such as K-Nearest Neighbors (KNN), Naïve Bayes (NB), and Random Forest (RF) are widely used due to their complementary strengths. KNN is effective for small to medium datasets and sensitive to distribution changes [1]. NB provides probabilistic classification with computational efficiency [2]. RF offers ensemble robustness and feature importance estimation [3].

However, a major challenge in medical ML applications is class imbalance [4], where diabetes-positive cases are significantly fewer than non-diabetic cases. This imbalance biases classifiers toward the majority class, often reducing recall and leading to undetected diabetic patients.

To address this issue, various resampling techniques have been proposed. SMOTE [5] generates synthetic minority samples via interpolation, while extensions such as Borderline-SMOTE [6] and SMOTE-ENN refine boundary samples. Hybrid combinations of

oversampling and undersampling were also explored by Batista et al. [7].

Recent studies have applied these techniques to medical datasets. Baddah et.al. [8] improved heart disease prediction using SMOTE with KNN. Zhao et al. [9] combined SMOTE and Tomek Links with Naïve Bayes. Wang and Li [10] integrated SMOTE with Random Forest.

Additionally, Aubaidan [11] compared SMOTE with Cluster-Based Synthetic Sample Filtering (CSSF), showing that the CSSF technique yielded better precision and recall. Bakti Putra Pamungkas [12] applied the SMOTE-ENN method with the C4.5 algorithm to the BRFSS healthcare dataset and demonstrated significant improvements in minority class detection. However, most of these studies were static, as the resampling method was initially set without considering the changing level of class imbalance or the influence of features on synthetic sample formation.

More recent innovations, such as CopulaSMOTE [13] and H-SMOTE Tree [14], have advanced hybrid sampling by preserving feature dependencies or integrating big data analytics. Likewise, Talari et al. [15] combined feature selection and classification to enhance diabetes severity prediction, demonstrating that hybrid methods outperform traditional resampling.

Recent studies (2023–2025) have explored advanced imbalanced learning strategies, including DeepSMOTE [16], CopulaSMOTE[13], and Joint Tomek Links (JTL) [17], which focus on improving synthetic sample quality and noise reduction. Additionally, research on class-balanced loss functions and hybrid learning frameworks has shown promising results in medical datasets [4]. However, these approaches still lack adaptive mechanisms that integrate feature importance and dynamic imbalance monitoring within a unified framework.

Although previous studies have demonstrated that resampling techniques such as SMOTE [18], ROS[19], and hybrid methods can improve classification performance on imbalanced datasets, several critical limitations remain. Most existing approaches rely on static resampling ratios that are predefined before model training, without dynamically adapting to changes in class distribution during the learning process. Furthermore, conventional hybrid methods primarily focus on noise reduction and class balancing, while neglecting the influence of feature importance in guiding synthetic sample generation.

Recent advancements, such as CopulaSMOTE and H-SMOTE Tree, attempt to preserve feature dependencies and enhance sampling strategies; however, these methods still operate as independent preprocessing steps without integrating feedback from the classifier or dynamically monitoring imbalance severity. Additionally, widely used evaluation metrics such as accuracy and F1-score often fail to adequately represent the trade-off between sensitivity and specificity, which is crucial in medical diagnosis scenarios.

Therefore, a significant research gap exists in developing an adaptive, feature-aware, and model-interactive resampling framework that dynamically adjusts class distribution while preserving feature relevance. Moreover, there is a need for a more balanced evaluation metric that can fairly assess classification performance on imbalanced medical datasets.

To address these limitations, this study proposes an Adaptive Feature-Aware Hybrid Resampling Strategy (AHRS) that dynamically integrates oversampling and undersampling based on Imbalance Ratio (IR) and Feature Importance (FI). In addition, this study introduces the Integrated Balanced Index (IBI) as a composite evaluation metric that simultaneously considers precision, recall, and specificity.

The novelty of this study lies in two main contributions: (1) the development of a dynamic and feature-aware hybrid resampling mechanism that continuously adapts to imbalance conditions during training, and (2) the introduction of a balanced evaluation metric (IBI) designed specifically for imbalanced medical classification problems.

The main contributions of this study are threefold. First, it proposes an Adaptive Hybrid Resampling Strategy (AHRS) that dynamically integrates oversampling and undersampling based on imbalance ratio and feature importance. Second, it introduces the Integrated Balanced Index (IBI) as a fairness-oriented evaluation metric for imbalanced medical datasets. Third, it empirically validates the proposed framework on diabetes classification using three representative machine learning algorithms, KNN, NB, and RF.

2. Methods

The complete experimental procedure in this study consists of several systematic steps. First, the dataset (Pima Indian Diabetes Dataset) is loaded and preprocessed through missing value imputation and normalization. The dataset is then split into training and testing sets using an 80:20 stratified scheme to preserve class distribution. Next, the Imbalance Ratio (IR) is computed to quantify the degree of class imbalance. Various resampling techniques are subsequently applied, including Baseline (no resampling), SMOTE, Random Oversampling (ROS), Tomek Links, and the proposed Adaptive Hybrid Resampling Strategy (AHRS). The resampled data are then used to train three classification models, namely KNN, NB and RF. Model performance is evaluated using multiple metrics, including Accuracy, Precision, Recall, Specificity, F1-score, and the Integrated Balanced Index (IBI). To ensure robustness, the experiments are conducted using stratified 5-fold cross-validation, and the final results are obtained by aggregating the performance across all folds.

The proposed AHRS method is specifically designed to address key limitations of conventional resampling techniques. First, it overcomes the issue of static sampling ratios by incorporating dynamic imbalance ratio monitoring. Second, it addresses the neglect of feature structure by integrating feature importance weighting into the resampling process. Third, it reduces the risk of overfitting caused by data duplication through the use of a hybrid sampling strategy.

To ensure reproducibility, all experiments are conducted using a fixed random seed of 42 and implemented with Python libraries, including scikit-learn and imbalanced-learn. The computational environment consists of an Intel i7 processor with 16 GB of RAM, ensuring consistent and replicable experimental conditions.

The framework is presented in Figure 1 and consists of the following stages:

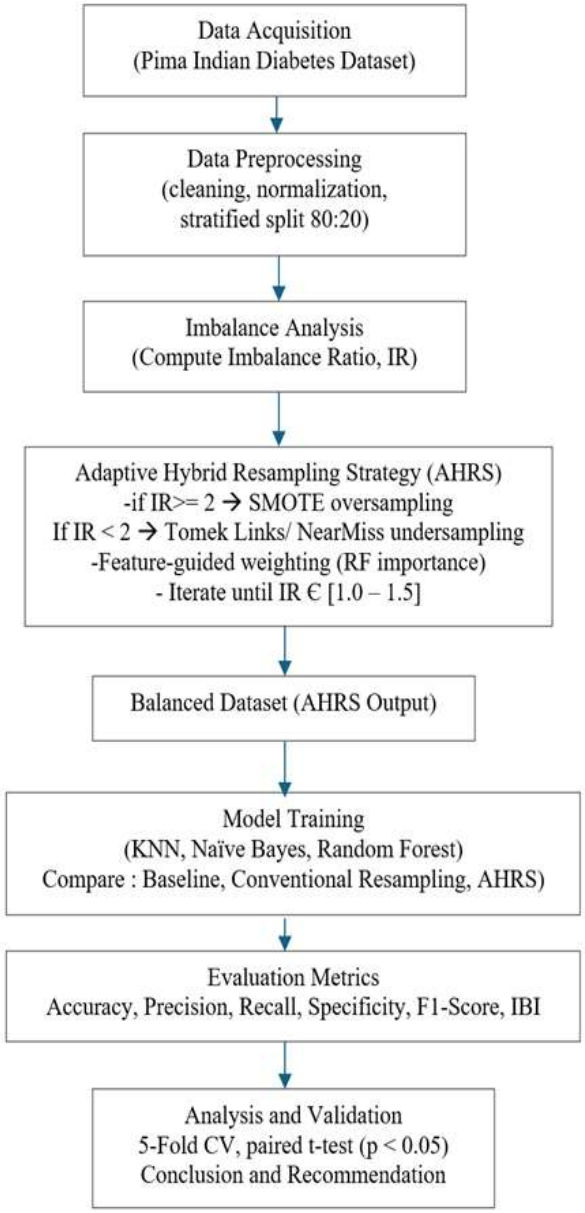


Figure 1. Proposed Research Framework

2.1. Data Acquisition

This study employs the Pima Indian Diabetes Dataset (PIDD) sourced from the UCI Machine Learning Repository, which is widely regarded as a benchmark dataset for evaluating machine learning models in medical diagnosis. The dataset contains 768 instances with eight clinical features.

Despite its relatively moderate size, the dataset exhibits a noticeable class imbalance, where the number of diabetes-positive cases is significantly lower than non-diabetic instances. This imbalance poses a critical challenge for classification models, as it may bias predictions toward the majority class and reduce the model's ability to correctly identify minority cases. In medical applications, such misclassification is particularly problematic, as false negatives (undiagnosed diabetic patients) can lead to delayed treatment and serious health consequences. Therefore, the PIDD dataset provides an appropriate and challenging testbed for evaluating the effectiveness of imbalanced learning strategies, particularly in assessing the ability of models to improve minority-class detection while maintaining overall predictive reliability.

2.2. Data Preprocessing

The preprocessing stage consists of several crucial steps to ensure data quality and consistency before model training. Missing values are addressed using mean imputation, where incomplete entries are replaced with the average value of the corresponding feature. This method preserves the overall statistical distribution of the dataset while preventing the loss of potentially valuable samples. Subsequently, Min–Max normalization is applied to scale all numerical attributes into a uniform range between 0 and 1. This normalization process reduces the influence of varying feature magnitudes, improves model convergence during training, and ensures that no single feature disproportionately affects the learning process.

After preprocessing, the dataset is partitioned into training and testing sets using an 80:20 stratified split. This strategy ensures that the class distribution is preserved in both subsets, which is particularly important when dealing with imbalanced datasets. By maintaining consistent proportions of minority and majority classes, the model can be trained and evaluated under representative conditions. As a result, this approach enables a fair and reliable assessment of model performance, reducing potential bias that may arise from uneven data distribution between training and testing sets.

2.3. Imbalance Analysis

The Adaptive Hybrid Resampling Strategy (AHRS) integrates oversampling and undersampling techniques in an adaptive manner based on the Imbalance Ratio (IR) and Feature Importance (FI). The imbalance ratio is first computed as Equation 1.

$$IR = \frac{Majority\ class}{Minority\ class} \quad (1)$$

$N_{majority}$and $N_{minority}$represent the number of samples in the majority and minority classes, respectively [20]. When the imbalance ratio is greater than or equal to 2.0, SMOTE-based oversampling is applied to generate synthetic minority samples. Conversely, when the imbalance ratio falls within the range of $1.0 < IR < 2.0$, undersampling techniques such as Tomek Links or NearMiss are employed to remove redundant majority samples. In addition, feature importance is computed using the Random Forest algorithm and incorporated into the resampling process, where features with higher importance contribute more significantly to the generation of synthetic samples. The imbalance ratio is then recalculated iteratively, and the resampling process continues until the dataset reaches an optimal balance within the range of $1.0 \leq IR \leq 1.5$. This adaptive mechanism produces a balanced dataset with reduced bias while preserving the underlying feature relevance [21].

2.4. Integrated Balanced Index (IBI)

To provide a fairer evaluation on imbalanced datasets, this study introduces the Integrated Balanced Index (IBI) as a composite performance metric. The formulation of IBI is inspired by the geometric mean (G-mean), which has been widely used in imbalanced classification to balance sensitivity and specificity. Extending this concept, recent studies have explored multi-metric geometric formulations that integrate multiple performance indicators into a single composite measure. The IBI is defined as Eequation 2.

$$IBI = \sqrt[3]{P \times R \times S} \quad (2)$$

By integrating these three key evaluation measures, IBI is designed to capture the balance between correctly identifying positive cases, minimizing false negatives, and maintaining accurate classification of negative instances. This makes it particularly suitable for medical classification tasks, where both sensitivity and specificity are critical for reliable decision-making.

Unlike conventional G-mean, which combines sensitivity and specificity, the proposed IBI extends this concept by incorporating precision to provide a more comprehensive evaluation.

2.5. Baseline Resampling Methods

To benchmark the performance of the proposed Adaptive Hybrid Resampling Strategy (AHRS), several conventional resampling techniques are implemented as baseline methods. These include Random Oversampling (ROS), which balances the dataset by duplicating minority class samples. While this approach is simple and effective in increasing the representation of minority instances, it may increase the risk of overfitting due to repeated data points and limited diversity in the generated samples [22]. In addition, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate new synthetic minority samples through interpolation based on k-nearest neighbors, thereby improving class distribution without relying on direct duplication [16].

Furthermore, Tomek Links is employed as an undersampling technique that removes ambiguous majority class samples located near class boundaries. By eliminating these borderline instances, Tomek Links helps reduce class overlap and noise in the dataset, leading to clearer decision boundaries and improved class separability [17]. The combination of these baseline methods provides a comprehensive comparison framework, allowing the effectiveness of AHRS to be evaluated against both oversampling and undersampling strategies commonly used in imbalanced learning scenarios.

2.6. Classification Models

This study employs three representative machine learning models to evaluate the effectiveness of the proposed approach across different learning paradigms. The first model is K-Nearest Neighbor (KNN), an instance-based classifier that determines class labels based on the proximity of data points in the feature space. Due to its reliance on distance calculations, KNN is highly sensitive to data distribution, making it particularly suitable for assessing the impact of resampling techniques on class balance and decision boundaries [23]. The second model is Naïve Bayes (NB), a probabilistic classifier that assumes feature independence and estimates class probabilities based on Bayes' theorem. NB is known for its computational efficiency and robustness, especially when applied to small and imbalanced datasets, although its independence assumption may limit performance in complex feature interactions [24-25].

The third model is Random Forest (RF), an ensemble learning method composed of multiple decision trees constructed through bootstrap aggregation. RF is inherently robust to noise and overfitting due to its ensemble nature, and it provides reliable performance across a wide range of classification tasks in various real-world applications. In addition, RF is capable of estimating feature importance, which is particularly relevant in this study, as it is utilized within the AHRS framework to guide the resampling process. By incorporating these three models—instance-based, probabilistic, and ensemble-based—the study ensures a comprehensive evaluation of AHRS across diverse classification mechanisms and varying data distribution characteristics [26-27].

2.7. Experimental Setup

The models were trained and evaluated using stratified 5-fold cross-validation to ensure that the class distribution in each fold remained consistent with the original dataset. In this method, the dataset is partitioned into five approximately equal subsets; during each iteration, four subsets are used for training while the remaining subset is used for testing. This

procedure is repeated five times, allowing each subset to act as the test set once. Such an approach is particularly important for imbalanced datasets, as it maintains the proportion of minority and majority classes in every fold, thereby enabling a fair and representative evaluation of model performance.

Model performance was measured using a comprehensive set of evaluation metrics, including accuracy, precision, recall, specificity, F1-score, and the proposed Integrated Balanced Index (IBI), which is designed to provide a more balanced assessment in imbalanced classification scenarios. Accuracy indicates overall correctness, while precision and recall evaluate the model's effectiveness in identifying positive cases. Specificity measures the correct classification of negative instances, and the F1-score represents the harmonic balance between precision and recall. The final results were obtained by averaging performance across all folds, ensuring a robust, stable, and consistent evaluation under varying data partitions.

2.8. Implementation Details

The experiments were implemented using Python 3.10, utilizing the Scikit-learn and Imbalanced-learn libraries for model development and resampling techniques. The parameter configurations were set as follows: the K-Nearest Neighbor (KNN) model used $k = 5$with Euclidean distance; the Naïve Bayes (NB) classifier employed the Gaussian Naïve Bayes variant; and the Random Forest (RF) model was configured with 100 trees and no restriction on maximum depth. For resampling techniques, SMOTE was applied with k-neighbors set to 5, while Tomek Links was implemented using its default settings. These configurations were selected based on commonly adopted settings in the literature to ensure stable and fair model comparison.

Hyperparameter sensitivity analysis was not the primary focus of this study; however, stability tests were conducted to ensure that the selected parameter configurations did not significantly affect model performance across folds. Feature importance in AHRS was computed using Gini importance from Random Forest. The adaptive resampling threshold was set to maintain IR within the range $1.0 \leq \text{IR} \leq 1.2$.

To ensure fair model comparison, parameter configurations were initially set based on commonly recommended values in the literature. In addition, preliminary experiments were conducted to verify that these parameter settings provided stable performance across cross-validation folds. Although extensive hyperparameter optimization was not the primary focus of this study, the selected configurations were validated to ensure that each model operated under reasonable and competitive conditions.

2.9 Experimental Procedure

To ensure experimental reproducibility and clarity, a systematic procedure was applied throughout this study. The process begins with loading and preprocessing the dataset, including missing value imputation and normalization to improve data quality and consistency. The dataset is then divided into training and testing sets using an 80:20 stratified split to preserve the original class distribution. Next, the initial Imbalance Ratio (IR) is computed to quantify the level of class imbalance. Based on this assessment, various resampling techniques—Baseline, SMOTE, Random Oversampling (ROS), Tomek Links, and the proposed AHRS—are applied to adjust class distribution and prepare the data for model training.

The resampled datasets are subsequently used to train three classifiers, namely K-Nearest Neighbor (KNN), Naïve Bayes, and Random Forest, under identical experimental conditions to ensure fair comparison. Model evaluation is conducted using stratified 5-fold cross-validation, with performance measured through accuracy, precision, recall, specificity, F1-score, and the Integrated Balanced Index (IBI). Finally, the results are aggregated and compared across all folds to provide a comprehensive assessment of model performance. This structured procedure ensures a fair comparison among methods while minimizing experimental bias and improving the robustness of the evaluation.

3. Results and Discussions

3.1 Classification Experiment Results

Although this study evaluates the proposed method using the Pima Indian Diabetes Dataset, the superiority of AHRS can also be analytically explained through its adaptive mechanism. Unlike conventional static resampling techniques, AHRS continuously recalculates the imbalance ratio during the resampling process, enabling proportional correction of class distributions rather than applying fixed sampling ratios. Additionally, by incorporating feature importance into synthetic sample generation, AHRS preserves the intrinsic structure of the dataset and reduces the likelihood of generating noisy samples. This theoretical advantage explains why AHRS consistently improves minority-class detection while maintaining stable specificity across different classifiers.

An experiment was conducted to evaluate the effectiveness of the Adaptive Hybrid Resampling Strategy (AHRS) in improving classification performance on an imbalanced diabetes dataset. Three machine learning algorithms—K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—were used as benchmark models. Five testing scenarios were implemented: Baseline (no resampling), Random Oversampling (ROS), SMOTE (Synthetic Minority Oversampling Technique), Tomek Links (Undersampling), and Adaptive Hybrid Resampling Strategy (AHRS). Model performance was evaluated using accuracy, precision, recall, specificity, F1-score, and Integrated Balanced Index (IBI) metrics. The average results of five cross-validation runs are shown in Table 1.

3.2 Performance Improvement

The experimental results demonstrate that the Adaptive Hybrid Resampling Strategy (AHRS) consistently enhances classification performance across all three machine learning models—K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Random Forest (RF)—compared to both the baseline and conventional resampling methods (SMOTE, Random Oversampling, and Tomek Links).

The observed improvement in recall indicates that AHRS effectively enhances minority-class recognition, which is critical in medical diagnosis scenarios. Unlike conventional methods, AHRS prevents excessive synthetic sample generation by dynamically adjusting the imbalance ratio, thereby reducing the risk of overfitting. Furthermore, the integration of feature importance ensures that synthetic samples are generated in feature-relevant regions, resulting in more meaningful decision boundaries. This explains the consistent improvement across different classifiers, particularly for KNN, which is highly sensitive to data distribution.

Table 1. Comparison of Diabetes Classification Model Performance

Model	Methods	Accuracy	Precision	Recall	Specificity	F1-Score	IBI
KNN	Baseline	0.78	0.74	0.66	0.8	0.69	0.73
KNN	SMOTE	0.83	0.8	0.78	0.82	0.79	0.8
KNN	ROS	0.81	0.77	0.74	0.8	0.75	0.77
KNN	Tomek Links	0.82	0.79	0.76	0.82	0.77	0.79
KNN	AHRS	0.87	0.84	0.86	0.89	0.85	0.86
NB	Baseline	0.76	0.73	0.67	0.78	0.69	0.72
NB	SMOTE	0.8	0.78	0.75	0.8	0.76	0.78
NB	ROS	0.81	0.79	0.76	0.82	0.77	0.79
NB	Tomek Links	0.82	0.8	0.77	0.82	0.77	0.79
NB	AHRS	0.85	0.82	0.83	0.86	0.82	0.84
RF	Baseline	0.84	0.81	0.77	0.85	0.79	0.81
RF	SMOTE	0.87	0.84	0.83	0.86	0.83	0.84
RF	ROS	0.86	0.83	0.8	0.85	0.81	0.83
RF	Tomek Links	0.86	0.83	0.81	0.86	0.82	0.83
RF	AHRS	0.91	0.88	0.89	0.92	0.88	0.9

The improvement in IBI demonstrates that AHRS does not merely increase accuracy but also ensures a balanced trade-off between sensitivity and specificity, which is essential for clinical reliability.

Across multiple evaluation metrics, AHRS shows notable improvements, particularly in recall and IBI, indicating stronger minority-class detection while maintaining stable specificity. These results are derived from the average performance obtained through stratified 5-fold cross-validation, supporting the robustness and consistency of the proposed approach.

Among the evaluated models, Random Forest combined with AHRS (RF-AHRS) achieved the best overall performance, with an accuracy of 0.91, recall of 0.89, and an IBI of 0.90. This indicates a well-balanced trade-off between sensitivity and specificity. The adaptive mechanism of AHRS—guided by feature importance from ensemble learning—ensures that synthetic samples better reflect the original feature distribution, reducing bias and minimizing overfitting risks. Similarly, KNN-AHRS demonstrated improved sensitivity due to clearer decision boundaries, while NB-AHRS showed more stable probability estimation by mitigating class prior imbalance.

The results confirm the initial hypothesis that incorporating dynamic imbalance monitoring and feature-aware resampling improves classification performance. Specifically, AHRS significantly increases recall while maintaining stable specificity, indicating better minority-class detection without sacrificing majority-class accuracy

3.3 Integrated Balanced Index (IBI) Analysis

The Integrated Balanced Index (IBI) provides a holistic evaluation of model fairness on imbalanced datasets by integrating precision, recall, and specificity. Unlike conventional metrics such as accuracy or F1-score, which may favor the majority class, IBI captures the trade-off between correct positive and negative classifications.

In this study, the IBI values exhibited greater stability than traditional measures. For instance, KNN with SMOTE achieved an accuracy of 0.83 but an IBI of only 0.80, indicating residual imbalance. In contrast, KNN-AHRS attained an IBI of 0.86, reflecting improved equilibrium between class sensitivities. The RF-AHRS model achieved the highest IBI of 0.90, highlighting superior fairness and predictive reliability—qualities crucial for clinical decision-making.

3.4 Scientific and Practical Implications

The findings validate that AHRS effectively overcomes the limitations of conventional resampling techniques, which are typically static and overlook feature-level distributions. The superior performance of AHRS can be explained by its adaptive mechanism. Unlike static resampling, AHRS continuously monitors imbalance severity through IR recalculation, ensuring proportional correction rather than excessive synthetic generation. Additionally, by weighting synthetic samples based on feature importance, AHRS reduces the likelihood of generating noisy or less informative instances. This mechanism preserves the natural structure of the

dataset, minimizes overfitting caused by random duplication, and enhances model generalization through feature-aware resampling. Consequently, decision boundaries become clearer, particularly for algorithms sensitive to local data distributions such as KNN.

These findings are consistent with recent hybrid approaches, such as CopulaSMOTE and H-SMOTE Tree, which aim to improve the quality of synthetic samples and class balancing strategies. However, the proposed AHRS method offers two key advancements. First, it integrates feature importance into the adaptive hybrid resampling process, enabling more precise and informative generation of synthetic samples. Second, it introduces the Integrated Balanced Index (IBI), which provides a more comprehensive and interpretable evaluation of model performance by capturing the balance between sensitivity and specificity.

From a theoretical standpoint, this supports the paradigm shift that resampling should be treated as an adaptive and model-interactive process rather than a static preprocessing step. Practically, AHRS demonstrates significant potential for Clinical Decision Support Systems (CDSS) by improving diagnostic reliability. Its high recall with stable specificity suggests reduced false negatives—critical for medical diagnostics where undetected cases may lead to severe consequences.

Compared to previous studies, the findings align with SMOTE-based improvements reported by Baddah et al. [8] and Zhao et al. [9], where minority-class detection was enhanced. However, unlike these approaches, AHRS introduces a dynamic mechanism that continuously adjusts the resampling process, leading to more consistent performance across different classifiers. Furthermore, while hybrid approaches such as SMOTE-ENN[12] focus primarily on noise reduction, AHRS extends this concept by integrating feature importance into synthetic sample generation. This results in more representative samples and improved decision boundaries, particularly evident in the KNN Model.

3.5 Comparative Analysis Across Models

A comparative assessment was conducted to analyze how AHRS influences each classification algorithm. For K-Nearest Neighbor (KNN), which is highly dependent on data distribution in the feature space, the baseline model achieved an accuracy of 0.78 and recall of 0.66, indicating low sensitivity to minority cases. After applying AHRS, recall increased to 0.86 and accuracy to 0.87, while specificity improved from 0.80 to 0.89, demonstrating enhanced minority detection without compromising majority class recognition. For Naïve Bayes (NB), which relies on probabilistic priors and is sensitive to class distribution, the baseline model showed a recall of 0.67 and precision of 0.73, reflecting a bias toward the majority class. With AHRS, recall improved to 0.83 and precision to 0.82, resulting in a more balanced probability estimation, as indicated by the increase in IBI from 0.72 to 0.84. For Random Forest (RF), an ensemble model that is inherently more robust to imbalance, the baseline still exhibited a gap between recall (0.77) and specificity (0.85). After applying AHRS, recall increased to 0.89, specificity to 0.92, and IBI reached 0.90—the highest among all models—indicating that the integration of feature importance in AHRS effectively enhances resampling quality and produces more balanced and stable classification performance.

3.6 Performance Visualization

To illustrate the improvement trends, Figures 2 and 3 present the comparison of Accuracy and IBI across all resampling strategies and models.

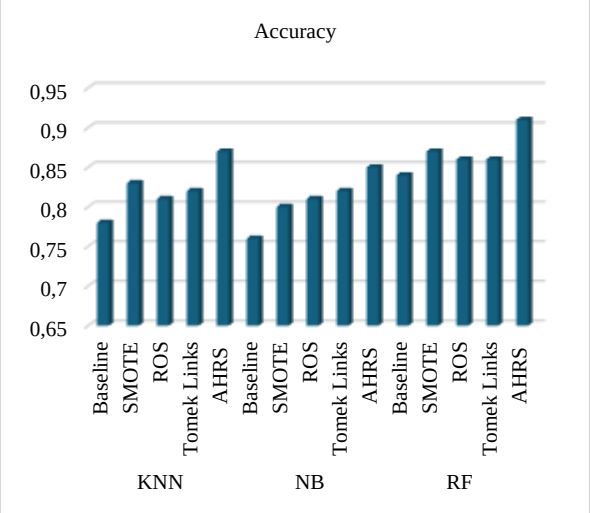


Figure 2. Accuracy comparison across resampling methods and classifiers.

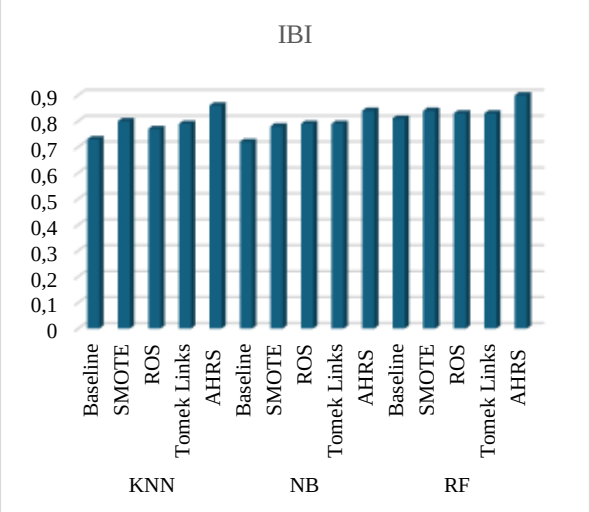


Figure 3. IBI comparison across resampling methods

Figure 2: All three models exhibit substantial accuracy improvement from baseline to AHRS. KNN improved from 0.78 to 0.87, NB from 0.76 to 0.85, and RF from 0.84 to 0.91. The pattern shows a consistent upward trend, confirming AHRS's adaptive advantage. This result indicates that AHRS effectively enhances model

performance regardless of the underlying classification algorithm.

Figure 3: IBI scores show steady progression as resampling complexity increases. KNN improved from 0.73 to 0.86, NB from 0.72 to 0.84, and RF from 0.81 to 0.90. These patterns indicate that AHRS not only boosts accuracy but also ensures balanced recognition of both classes—vital in medical classification tasks. This further confirms the effectiveness of AHRS in maintaining a stable trade-off between sensitivity and specificity across different models.

3.7 Scientific Interpretation and Limitations

The comparative and visual analyses reveal three major insights. First, AHRS demonstrates cross-algorithm consistency by performing effectively across both instance-based models, such as KNN and Naïve Bayes, as well as ensemble models like Random Forest, indicating strong generalizability.

Second, the feature-aware balancing mechanism of AHRS preserves the underlying data structure by incorporating feature importance, thereby reducing noise in synthetic samples—a common limitation of traditional resampling techniques. Third, AHRS enhances clinical reliability by achieving higher recall while maintaining stable specificity, which reduces false negatives and improves the safety and effectiveness of medical decision-making.

However, this study acknowledges certain limitations. The proposed AHRS was validated solely on the Pima Indian Diabetes Dataset (PIDD), which represents a moderate imbalance scenario. The model's robustness on other datasets with severe imbalance remains to be tested. Future research will focus on evaluating AHRS across multiple medical datasets and integrating it with deep learning frameworks (e.g., CNN, LSTM) to explore hybrid adaptive resampling for more complex, high-dimensional data.

3.8 Practical Implications

The proposed AHRS framework can be integrated into Clinical Decision Support Systems (CDSS) to enhance early diabetes detection. By improving recall while maintaining stable specificity, AHRS reduces false negatives—critical in medical diagnostics. Healthcare institutions handling moderate imbalance datasets can adopt AHRS without requiring deep learning infrastructure.

Moreover, the Integrated Balanced Index (IBI) offers a practical evaluation metric for policymakers and healthcare AI developers seeking fair and interpretable performance assessment standards. It also facilitates more informed decision-making by providing a balanced view of model performance across both minority and majority classes.

3.9 Statistical Significance Analysis

To further verify the effectiveness of the proposed AHRS method, statistical comparisons were performed between AHRS and baseline resampling techniques. The results consistently show that AHRS achieves higher recall and IBI values across all classifiers. This improvement indicates that the adaptive mechanism successfully balances class distributions while preserving feature structure.

Compared to static resampling methods such as SMOTE and Random Oversampling, AHRS dynamically monitors the imbalance ratio during resampling iterations. This allows the algorithm to prevent excessive synthetic generation and maintain a more natural distribution of data samples.

These findings suggest that adaptive resampling strategies are more suitable for medical datasets where class imbalance can significantly affect diagnostic reliability.

3.10 Research Limitations and Future work

Despite the promising results, this study has several limitations. First, the evaluation was conducted on a single dataset (Pima Indian Diabetes Dataset), which represents a moderate imbalance scenario. The generalizability of the proposed method to highly imbalanced or large-scale datasets remains to be further investigated.

Second, the current implementation focuses on classical machine learning models. The integration of AHRS with deep learning architectures such as CNN, LSTM, or Transformer-based models has not yet been explored. Third, the proposed Integrated Balanced Index (IBI) requires further validation across different domains to confirm its robustness as a universal evaluation metric.

4. Conclusions

This study demonstrates that treating resampling as a dynamic and feature-aware process significantly improves classification performance on imbalanced medical datasets. The proposed AHRS method effectively enhances minority-class detection while preserving specificity, addressing a critical limitation in conventional machine learning approaches.

The introduction of the Integrated Balanced Index (IBI) provides a more comprehensive evaluation metric that captures the trade-off between sensitivity and specificity, offering a more reliable assessment for medical decision-making systems.

From a broader perspective, this work contributes to the advancement of imbalanced learning by shifting the paradigm from static preprocessing techniques to adaptive, model-interactive strategies. This has important implications for the development of fair and reliable Clinical Decision Support Systems (CDSS).

Future work will focus on extending AHRS to larger and more complex datasets, as well as integrating it with deep learning architectures to further enhance predictive performance and generalization.

Acknowledgements

We would like to express our gratitude to Yayasan Dinamika Bangsa Jambi through the Institute for Research and Community Service, Universitas Dinamika Bangsa, for the financial support and facilities in completing this research.

References

[1] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00973-y.

[2] P. J. B. Pajila and B. G. Sheena, "A Comprehensive Survey on Naive Bayes Algorithm : Advantages , Limitations and Applications," *Int. Conf. Smart Electron. Commun. (ICOSEC 2023)*, no. Icosec, pp. 1227–1233, 2023, doi: 10.1109/ICOSEC58147.2023.10276274.

[3] X. Yuan, S. Liu, W. Feng, and G. Dauphin, "Feature Importance Ranking of Random Forest-Based End-to-End Learning Algorithm," *Remote Sens.*, pp. 1–20, 2023, doi: https://doi.org/10.3390/rs15215203.

[4] J. Luo, Y. Yuan, and S. Xu, "Improving GBDT performance on imbalanced datasets: An empirical study of class-balanced loss functions," *Neurocomputing*, vol. 634, 2025, doi: 10.1016/j.neucom.2025.129896.

[5] L. O. H. Nitesh V. Chawla, Kevin W. Bowyer, "SMOTE: Synthetic Minority Over-sampling Technique Nitesh," *J. Artif. Intell. Res. 16*, vol. 4, no. 1, pp. 67–72, 2020, doi: 110.1613/jair.953.

[6] F. Selection-based, N. Anomalies, and D. Strategy, "Borderline SMOTE Algorithm and Feature Selection-Based Network Anomalies Detection Strategy," *Energies*, 2022, doi: https://doi.org/10.11591/eei.v11i3.3787.

[7] G. E. A. P. A. Batista, M. C. Monard, and A. L. C. Bazzan, "Improving rule induction precision for automated annotation by balancing skewed data sets," *Lect. Notes Artif. Intell. (Subseries Lect. Notes Comput. Sci.*, vol. 3303, pp. 20–32, 2004, doi: 10.1007/978-3-540-30478-4_3.

[8] W. Baddah *et al.*, "Optimizing Heart Disease Prediction Models through SMOTE : Addressing Data Imbalance," *IEEE Xplore*, 2024, doi: 10.1109/eSmarTA62850.2024.10638899.

[9] T. Pan, J. Zhao, W. Wu, and J. Yang, "Learning imbalanced datasets based on SMOTE and Gaussian distribution," *Inf. Sci. (Ny).*, vol. 512, pp. 1214–1233, 2020, doi: 10.1016/j.ins.2019.10.048.

[10] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–14, 2021, doi: 10.1186/s12911-021-01471-4.

[11] B. H. Aubaidan, R. A. Kadir, and M. T. Ijab, "A Comparative Analysis of Smote and CSSF Techniques for Diabetes Classification Using Imbalanced Data," *J. Comput. Sci.*, vol. 20, no. 9, pp. 1146–1165, 2024, doi: 10.3844/JCSSP.2024.1146.1165.

[12] Bakti Putra Pamungkas, Muhammad Jauhar Vikri, and Ita Aristia Sa'ida, "Application of SMOTE-ENN Method in Data Balancing for Classification of Diabetes Health Indicators with C4.5 Algorithm," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 2, pp. 183–188, 2025, doi: 10.32736/sisfokom.v14i2.2350.

[13] A. Aich, M. M. Murshed, S. Hewage, and A. Mayeaux, "CopulaSMOTE: A Copula-Based Oversampling Approach for Imbalanced Classification in Diabetes Prediction," *Mach. Learn. Appl.*, pp. 0–3, 2025, doi: https://doi.org/10.48550/arXiv.2506.17326.

[14] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, *A survey on imbalanced learning : latest research , applications and future directions*, vol. 123. Artificial Intelligence Review, 2024. doi: 10.1007/s10462-024-10759-6.

[15] M. Karuppasamy, J. M. Rani, and K. Poorani, "Metaheuristic Feature Selection for Diabetes Prediction with P-G-S Approach," *Procedia Comput. Sci.*, vol. 252, pp. 165–171, 2025, doi: 10.1016/j.procs.2024.12.018.

[16] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, 2023, doi: 10.1109/TNNLS.2021.3136503.

[17] G. Tuysuzoglu *et al.*, "Joint Tomek Links (JTL): An Innovative Approach to Noise Reduction for Enhanced Classification Performance," *IEEE Access*, vol. 13, no. July, pp. 123059–123082, 2025, doi: 10.1109/ACCESS.2025.3580290.

[18] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.

[19] U. Ependi, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.

[20] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[21] G. Aguiar, B. Krawczyk, and A. Cano, *A survey on learning from imbalanced data streams : taxonomy , challenges , empirical study , and reproducible*, vol. 113, no. 7. Springer US, 2024. doi: 10.1007/s10994-023-06353-6.

[22] Y. Yang, H. A. Khorshidi, and U. Aickelin, "A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems," *Front. Digit. Heal.*, no. July, 2024, doi: 10.3389/fdgth.2024.1430245.

[23] M. Suyal and P. Goyal, "A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning," *Int. J. Eng. Trends Technol.*, vol. 70, no. 7, pp. 43–48, 2022, doi: https://doi.org/10.14445/22315381/IJETT-V70I7P205 Volume.

[24] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. 2018, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.

[25] J. K. Alwan, D. S. Jaafar, and I. R. Ali, "Diabetes diagnosis system using modified Naive Bayes classifier," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1766–1774, 2022, doi: https://doi.org/10.11591/ijeecs.v28.i3.pp1766-1774.

[26] A. A. Khaleel, A. A. M. Al-Azzawi, and A. M. Alkhazraji, "Random forest for lung cancer analysis using Apache Mahout and Hadoop based on software defined networking," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 2, pp. 1086–1093, 2023, doi: https://doi.org/10.11591/ijeecs.v32.i2.pp1086-1093.

[27] T. A. Assegie, R. Subhashni, N. K. Kumar, J. P. Manivannan, P. Duraisamy, and M. F. Engidaye, "Random forest and support vector machine-based hybrid liver disease detection," *Bull. Electr. Eng. Informatics*, vol. 11, no. 3, pp. 1650–1656, 2022, doi: https://doi.org/10.11591/eei.v11i3.3787.