JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 3, July 2026, pp. 945-963

Hybrid Approach for Class Imbalance Handling using Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling

Hartono[1*], Erianto Ongko[2], Muhammad Khahfi Zuhanda[1]
[1]Department of Magister Program in Informatics, Postgraduate School, Universitas Medan Area, Indonesia
[2]Department of Computer Science, Institut Modern Arsitektur dan Teknologi, Indonesia
hartono@staff.uma.ac.id

Article History:
Received : 05-12-2025
Revised : 17-01-2026
Accepted : 26-01-2026
Online : 01-07-2026

Keywords:
Class Imbalance;
Mult-class classification;
Hybrid Resampling;
Adaptive Weighted Oversampling;
Instance Hardness-Based Undersampling.

ABSTRACT

Class imbalance remains a major challenge in multi-class classification, where existing hybrid resampling methods often combine oversampling and undersampling in a loosely coupled manner, without explicitly coordinating minority enrichment and majority reduction. In this experimental study, we propose a novel hybrid resampling method, Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU), which differs from existing hybrid approaches by explicitly aligning boundary-aware minority oversampling with instance hardness-based majority undersampling. Rather than independently applying oversampling and undersampling, the proposed method integrates both processes through a coordinated design guided by classification difficulty to improve decision boundary quality. Methodologically, AWO-IHU first applies adaptive weighted oversampling to emphasize informative minority instances near class boundaries, followed by instance hardness-based undersampling that selectively removes redundant majority samples using an ensemble-based difficulty estimation. The experimental evaluation is conducted using multiple benchmark datasets with varying numbers of instances, attributes, and classes. Classification performance is evaluated using Accuracy, Precision, Recall, and Cohen's Kappa, enabling a comprehensive assessment of overall correctness, minority sensitivity, and agreement beyond chance under class imbalance. Experimental results show that AWO-IHU consistently outperforms SMOTE, Random Undersampling, and conventional hybrid sampling methods. In particular, the proposed method achieves perfect or near-perfect Recall values up to 1.0, while maintaining high Precision values above 0.89 and producing the highest Cohen's Kappa values up to 0.86. These findings demonstrate that explicitly coordinating minority enrichment with difficulty-aware majority reduction yields more reliable decision boundary learning and improved generalization in imbalanced multi-class classification.

https://doi.org/10.31764/jtam.v10i3.36972

——————————— ◆ ———————————

A. INTRODUCTION

Class imbalance is a persistent challenge in supervised classification, especially in real-world applications such as medical diagnosis (Matharaarachchi et al., 2024), fraud detection (Huang et al., 2024), fault detection (Zhong et al., 2023), and cybersecurity (Alabduallah et al., 2024), where minority class instances are rare but often carry higher decision importance than majority instances (Altalhan et al., 2025). In imbalanced datasets, standard classifiers optimized for overall accuracy tend to prioritize majority classes, which may produce deceptively high accuracy while yielding poor minority detection and elevated false-negative rates (Razali et al., 2025). This imbalance-driven bias limits the practical reliability of classification systems in decision-critical scenarios and motivates the development of data-level strategies that explicitly adjust training distributions to support minority learning (Aymaz, 2025).

To address class imbalance, data-level resampling methods have been widely studied due to their classifier-independent nature (Fachrie et al., 2025). These methods are commonly grouped into oversampling, undersampling, and hybrid strategies, each offering distinct benefits but also exhibiting recurring limitations (Shi et al., 2023). Oversampling approaches increase minority class representation, ranging from Random Oversampling to synthetic generation methods such as SMOTE (Elreedy & Atiya, 2019). While Random Oversampling can cause overfitting through duplication, SMOTE improves generalization by interpolating minority instances (Wang et al., 2023). However, SMOTE and many of its variants still treat minority samples relatively uniformly and may generate samples in regions that are not equally informative for decision boundary learning (Olabisi et al., 2025). Boundary-aware (Li et al., 2022) and density-aware oversampling methods attempt to improve this by prioritizing difficult minority instances (Zhang et al., 2023); for example, MWMOTE assigns higher selection weights to informative minority samples based on their proximity and local density relative to majority samples, encouraging synthetic generation near critical boundaries. Despite these improvements, oversampling methods including MWMOTE share a key limitation: they primarily focus on minority enrichment without explicitly controlling majority class redundancy and dominance. As a consequence, majority bias may persist, class overlap may remain unresolved, and the learning process can still be influenced by abundant, easily classified majority instances (Barua et al., 2014).

Undersampling methods tackle the imbalance from the opposite direction by reducing the majority class size (Hoyos-Osorio et al., 2021). Random undersampling is computationally efficient (Hancock et al., 2022) but may discard informative majority samples and weaken decision boundaries (Taskiran et al., 2025). More selective undersampling strategies attempt to reduce this risk using criteria such as clustering or instance selection (Hassani et al., 2025). In particular, instance hardness-based undersampling removes majority instances that are consistently easy to classify while retaining more informative samples that tend to lie closer to decision boundaries (Lin et al., 2026). This strategy can reduce redundancy and computational cost while preserving challenging majority patterns. However, undersampling methods alone do not address the core problem of minority scarcity, and overly aggressive removal can lead to information loss or unstable boundaries (Xie et al., 2024). Thus, undersampling-only solutions may reduce majority dominance but still leave the minority class insufficiently represented in informative regions of the feature space (Farou et al., 2024).

The limitations above reveal a clear gap in the resampling literature: existing methods often address either minority class scarcity or majority class redundancy, but not both in a coordinated and principled manner (Carvalho et al., 2025). Oversampling methods (including boundary- and density-aware approaches such as MWMOTE) improve minority representation but do not actively refine the majority class distribution (Chen et al., 2024); undersampling methods (including instance hardness-based selection) reduce majority dominance but do not enrich minority information (Chiu & Minku, 2024). Although hybrid resampling strategies have been proposed, many integrate oversampling and undersampling heuristically, treating them as loosely connected steps rather than mechanisms that should be explicitly aligned (Salehi & Khedmati, 2025). In particular, minority sample generation is frequently performed without considering which majority instances are redundant versus boundary-defining (Nouas et al., 2025), and majority reduction is often executed without being guided by where minority enrichment is most needed (Shyalika et al., 2024). This lack of structured coordination motivates the need for a hybrid approach that systematically aligns boundary-aware minority enrichment with difficulty-aware majority reduction, especially for multi-class imbalanced learning where overlap and boundary ambiguity are more severe (M et al., 2025).

To address this gap, this study proposes Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU), a coordinated hybrid resampling framework that integrates the complementary strengths of both directions. Conceptually, the method uses instance hardness to selectively remove redundant majority samples while preserving boundary-relevant majority instances, and employs MWMOTE-inspired adaptive weighting to prioritize informative minority samples for synthetic generation in critical regions. By jointly targeting minority scarcity and majority redundancy within a unified resampling logic, the proposed framework is designed to reduce overlap, improve boundary learning, and enhance minority detection without unnecessary distortion of the training distribution.

The contributions of this research can be summarized as follows: (1) structured hybrid resampling framework that explicitly integrates MWMOTE-style adaptive weighted oversampling with instance hardness-based undersampling to address complementary limitations in prior work; (2) A principled coordination of resampling objectives, linking majority refinement (removing redundant easy majority samples) with minority enrichment (prioritizing informative minority regions), thereby improving decision boundary learning in multi-class imbalance; and (3) Comprehensive empirical validation on multi-class benchmark datasets, demonstrating consistent improvements over standard oversampling, undersampling, and conventional hybrid resampling methods in terms of classification performance and agreement metrics.

B. METHODS

1. Majority Weighted Minority Oversampling Technique (MWMOTE)

The Majority Weighted Minority Oversampling Technique (MWMOTE) is a synthetic oversampling method designed to address class imbalance problems by focusing on difficult-to-learn minority instances. MWMOTE operates in three main phases. The first phase is Noise Removal, where minority instances considered as noise, typically those surrounded entirely by majority class neighbors are excluded from further processing. This step helps prevent the generation of misleading synthetic samples. In the second phase, Informative Instance Weighting, the algorithm identifies a borderline majority set and an informative minority set using k-nearest neighbors. Each informative minority instance is assigned a weight based on two factors: the closeness factor, which measures the proximity to majority class instances, and the density factor, reflecting the local density of majority instances around it. These weights are combined into an information weight for each pair, which are then aggregated to calculate a selection weight for each minority instance. The selection probability is computed by normalizing these weights, ensuring that more critical minority instances have a higher chance of being used for generating synthetic samples. Finally, in the Synthetic Instance Generation phase, the minority class is clustered using average-linkage clustering to preserve local structures. New synthetic samples are created by selecting minority instances based on their calculated selection probability, then interpolating between them and randomly chosen instances from the same cluster. This interpolation ensures that the new samples are generated within meaningful minority class regions, avoiding overlapping with majority class areas. The result is a more balanced dataset containing both real and synthetic minority class instances, aimed at improving classifier performance on imbalanced datasets. The following presents the pseudocode for the Majority Weighted Minority Oversampling Technique (MWMOTE) (Han et al., 2025).

Algorithm 1: Majority Weighted Minority Oversampling Technique (MWSMOTE)

Input: Dataset D with minority class samples S_{min} and majority classs samples S_{maj}
Number of nearest neighbors k
Number of synthetic instances to generate n
Output: Balanced dataset with synthetic minority instances
Step 1: Noise Removal;
$S_{minf} \leftarrow S_{min} - \{x_i \in S_{min}: N(x_i)\ contains\ no\ minority\ instance\}$;
Step 2: Informative Instance Weighting;
Identify borderline majority set S_{bmaj} and informative minority set S_{imin} using k-nearest neighbors;
foreach $x_i \in S_{imin}$ **do**
  foreach $y_i \in S_{bmaj}$ **do**
    Compute closeness factor $C_f(y_i, x_i)$;
    Compute density factor $D_f(y_i, x_i)$;
    Compute information weight $I_w(y_i, x_i) = C_f(y_i, x_i) \,.\, D_f(y_i, x_i)$;
  end
  Compute selection weight $S_w(x_i) = \sum I_w(y_i, x_i)$;
end
Compute selection probability $S_p(x_i) = \frac{s_w(x_i)}{\sum s_w(z_i)}$ for all $x_i \in s_{imin}$;
Step 3: Synthetic Instance Generation;
Perform average-linkage clustering on S_{minf};
for $i \leftarrow 1$ **to** n do
  Select x from s_{imin} based on S_p;
  Randomly select y from same cluster as x;
  Generate synthetic instance $s = x + \alpha.(y - x)$, where $\alpha \in [0,1]$;
  Add s to S_{minf}
end
return Balanced dataset $S_{minf} \cup S_{maj}$

2. MDGP Forest

The MDGP-Forest algorithm is a machine learning method designed to improve classification performance by combining decision forests with genetic programming. The process starts by training the first layer using the original dataset directly. In the next layers, the algorithm separates the dataset into subsets based on each class. For every class-specific subset, genetic programming is used to automatically generate new features that are specifically useful for distinguishing that class. These new features are then combined with the existing data to enrich the input for the next layer. In each layer, several decision forests are trained, and their prediction outputs are collected as additional features for the next iteration. After each layer is completed, the algorithm checks whether the model's classification performance has improved. If the performance increases, the model continues building more layers; if not, the training process stops. The result is a final model that combines both the decision forests and the newly constructed features, making the classifier stronger and more capable of handling multi-class problems, especially in complex or imbalanced datasets. The following presents the pseudocode for MDGP-Forest (Lin et al., 2026).

Algorithm 2: MDGP-Forest: Multi-Class Disassembly and Genetic Programming Enhanced Deep Forest

Input: Dataset D =$\{(x_i, y_u)\}$ with n instances and c classes
Maximum layers T
Number of decision forests per layer K
GP population size P
Max generations for GP G_{max}
Number of constructed features per class p
Instance hardness threshold θ
Output: Trained MDGP-Forest model
Initialize MDGP-Forest$\leftarrow \emptyset$;
$t \leftarrow 1, best\_F1 \leftarrow 0, training \leftarrow True$;
while *training* **do**
  if *t=1* **then**
    $X_t \leftarrow X$;
  else
    Perform Multi-Class Diassembly and Sampling to get
    $\{D_1, D_2, \ldots, D_c\}$;
    for $i \leftarrow 1$ ***to*** c **do**
      Use GP on D_i to obtain feature constructor $FC_i(.)$;
    end
    Combine $\{FC_1(.), \ldots, FC_c(.)\}$ to form $FC_t(.)$;
    $X_t \leftarrow concatenate\ (FC_t(X), E_t)$;
  end
  Initialize $L_t \leftarrow L_t \cup F_i$;
  for $i \leftarrow 1$ **to** K **do**
    Train decision forest F_i on (X_t, Y);

Obtain prediciton vector $V_{t,i}$ from F_i;
$L_t \leftarrow L_t \cup F_i$;
$E_{t+1} \leftarrow concatenate\ (E_{t+1}, V_{t,i})$;
end
end

3. Proposed Method

This research proposed a novel method named Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU) to address the challenges of class imbalance in multi-class classification tasks. The proposed method integrates two complementary strategies: adaptive weighted oversampling and instance hardness-based undersampling, designed to enhance data balance while preserving critical decision boundaries. In the oversampling phase, the method focuses on enriching the minority class through an adaptive weighted mechanism inspired by MWMOTE. Noisy minority instances are first removed to ensure the reliability of synthetic instance generation. Informative minority samples are identified based on their proximity to borderline majority instances using k-nearest neighbors. Selection probabilities are then computed for each informative minority instance, considering closeness and density factors to guide the generation of synthetic samples in strategically important regions.

In parallel, the undersampling phase targets the reduction of redundant majority class instances. Instance hardness is computed using an ensemble of decision forests to evaluate the classification difficulty of each majority instance. Majority instances with low hardness (i.e., those consistently classified correctly) are removed, ensuring that only critical and informative majority samples are retained. Overall, this research proposed method balances the dataset by simultaneously increasing minority class representation and reducing unnecessary majority class samples. By combining adaptive weighted oversampling and instance hardness-based undersampling, AWO-IHU offers a robust solution that mitigates the effects of class imbalance, improves classifier learning, and enhances predictive performance in complex, multi-class datasets.

Algorithm 3: Robust Class Imbalance Handling Using Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU)

Input: Dataset $D = \{(x_i, y_u)\}$ with n instances and c classes
Number of nearest neighbors k
Number of synthetic minority instances n_{syn}
Instance hardness threshold θ
Number of decision forests K per layer
Maximum layers T
Output: Balanced dataset $D_{balanced}$
Phase 1: Adaptive Weighted Oversampling;
Identify minority class samples S_{min} and majority class samples S_{maj};
Step 1.1: Noise Removal;

$S_{minf} \leftarrow S_{min} - \{x_i \in S_{min}: N(x_i) contains\ no\ minority\ instance\}$;
Step 1.2: Informative Instance Weighting;
Identify informative minority set S_{imin} and boderline majority set S_{bmaj} using k-nearest neighbors;
foreach $x_i \in S_{imin}$ **do**
  foreach $\boldsymbol{y}_i \in S_{bmaj}$ **do**
    Compute closeness factor $C_f(y_i, x_i)$;
    Compute density factor $D_f(y_i, x_i)$;
    Compute information weight $I_w(y_i, x_i) = C_f(y_i, x_i) \, . \, D_f(y_i, x_i)$;
  e**nd**
  Compute selection weight $S_w(x_i) = \sum I_w(y_i, x_i)$;
end
Compute selection probability $S_p(x_i) = \frac{s_w(x_i)}{\sum s_w(z_i)}$ for all $x_i \in S_{imin}$;
Step 1.3: Synthetic Instance Generation:
Cluster S_{minf} using average-linkage clustering;
For $i \leftarrow 1$ ***to*** n_{syn} ***do***
  Select x from s_{imin} based on S_p;
  Select y randomly from same cluster as x;
  Generate synthetic instance $s = x + \alpha \, . \, (y - x), \alpha \in [0,1]$;
  Add s to s_{minf}
end
Phase 2: Instance Hardness-Based Undersampling;
Train a temporary decision forest using MDGP forests;
For each instance $x_i \in S_{maj}$, compute instance hardness $IH(x_i)$ based on misclassification rates across forests;
Remove x_i from S_{maj} if $IH(x_i) < \theta$
Phase 3: Final Dataset Construction;
$D_{balanced} \leftarrow S_{minf} \cup retained\ S_{maj}$;
Return $D_{balanced}$

Figure 1 illustrates the workflow of the proposed method.

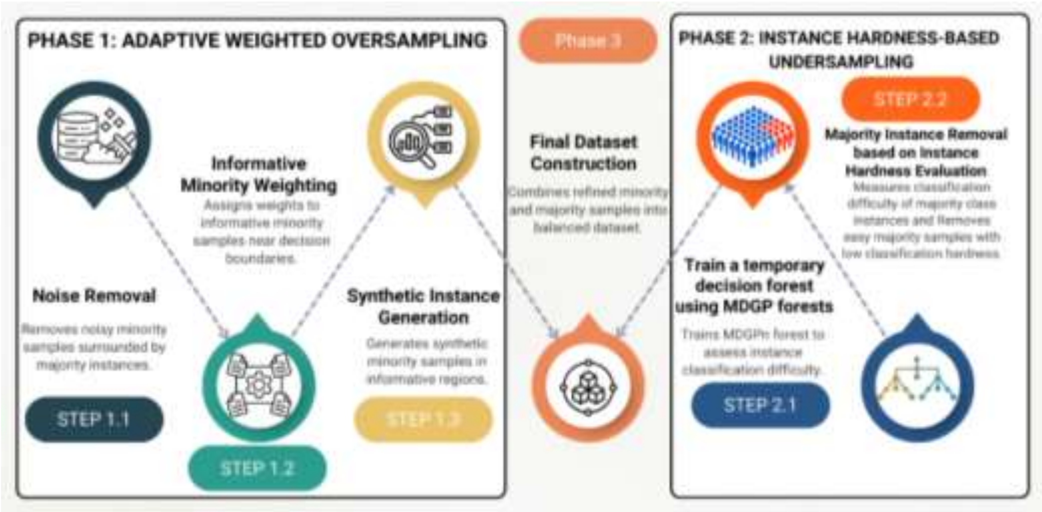

Figure 1. Workflow of the Proposed Method

Based on Figure 1, it can be observed that the proposed Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU) method is structured as a coordinated hybrid resampling framework consisting of three sequential phases: adaptive weighted oversampling, instance hardness-based undersampling, and final dataset construction. The workflow illustrates how minority class enrichment and majority class reduction are systematically aligned to improve class balance and decision boundary learning.

In Phase 1, adaptive weighted oversampling is applied to enhance minority class representation in strategically important regions of the feature space. The process begins with a noise removal step, where minority samples that are entirely surrounded by majority class neighbors are identified and excluded. This step prevents the generation of misleading synthetic samples that could negatively affect classifier performance. Next, informative minority instances are identified based on their proximity to borderline majority samples using a k-nearest neighbors strategy. Minority samples located near decision boundaries are assigned higher selection importance through closeness and density considerations. Synthetic minority instances are then generated within locally coherent minority regions, ensuring that the synthetic data remain representative and informative.

Phase 2 focuses on instance hardness-based undersampling to reduce majority class dominance while preserving boundary-relevant information. As shown in Figure 1, a temporary MDGP-Forest model is trained to assess the classification difficulty of majority class instances. Majority samples that are consistently easy to classify are considered redundant and are selectively removed, while instances exhibiting higher classification difficulty are retained to maintain meaningful decision boundaries. Finally, Phase 3 constructs the resampled training dataset by combining the refined minority samples from the oversampling phase with the selected majority samples from the undersampling phase. This coordinated integration, as illustrated in Figure 1, results in a balanced dataset that improves minority representation, reduces majority bias, and supports more effective classifier training in multi-class imbalanced learning scenarios.

C. RESULT AND DISCUSSION

1. Dataset

This research employs imbalanced multi-class datasets sourced from the UCI Machine Learning Repository, as shown in Table 1.

Table 1. Dataset

Dataset	Number of Instances	Number of Attributes	Number of Classes
Balance	625	4	3
Car	1728	6	4
Contraceptive	1473	9	3
Ecoli	336	7	8
Yeast	1484	8	10

2. Perfromance Metrics

The performance of the classifier will be evaluated using Accuracy, Precision, Recall, and Cohen's Kappa metrics. These evaluations are conducted based on the confusion matrix, as presented in Table 2.

Table 2. Confusion Matrix

	Predictive Positive Class	Predictive Negative Class
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

The Accuracy, Precision, Recall, and Cohen's Kappa calculations can be seen in the following equation.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Kappa = \frac{P_0-P_c}{1-P_c} \quad (4)$$

3. Results

The performance of the classifier will be evaluated using Accuracy, Precision, Recall, and Cohen's Kappa metrics. The performance testing was compared the proposed method with SMOTE, Random Undersampling, and Hybrid Sampling. The performance using Balance Dataset can be seen in Figure 2 and Table 3.

Table 3. Performance using Balance Dataset

Method	Accuracy	Precision	Recall	Kappa
Proposed Method	0.9312	0.8950	1.0	0.8604
SMOTE	0.8944	0.8661	0.9705	0.7874
Random Undersampling	0.8480	0.8471	0.9122	0.6964
Hybrid Sampling	0.8048	0.8264	0.8571	0.6094

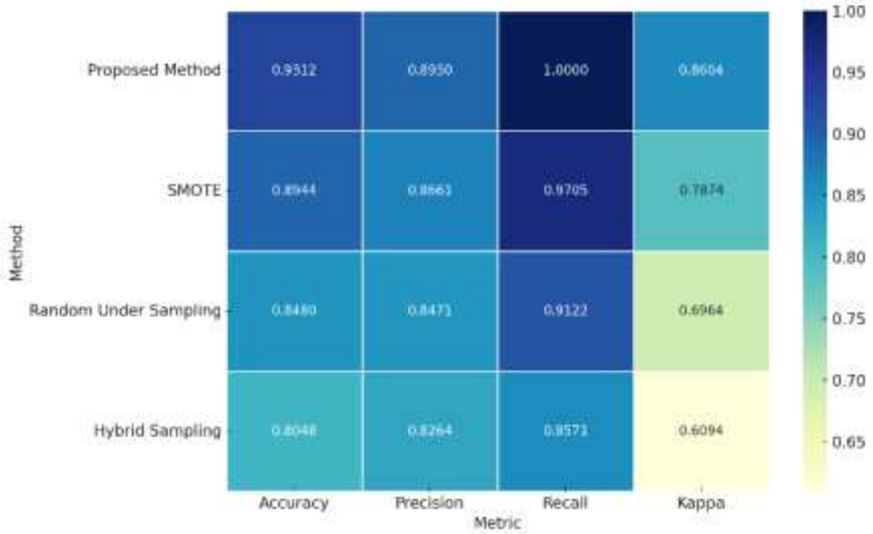

Figure 2. Performance using Balance Dataset

For the Balance dataset, Table 3 and Figure 2 shows that the proposed method provides the most reliable improvement across all evaluation metrics. Although the dataset has relatively low dimensionality, multi-class imbalance can still cause biased decision boundaries because minority patterns may occupy narrow regions in the feature space. The proposed method achieves the highest accuracy (0.9312) and precision (0.8950) while attaining perfect recall (1.0), indicating that the model successfully captures all minority-class instances without substantially increasing false positives. This is important because recall gains in imbalanced settings often come with a precision penalty; however, the high precision here suggests that synthetic minority generation is concentrated in informative regions rather than being overly aggressive or noisy.

A more critical indicator is Cohen's Kappa, which accounts for chance agreement and is more informative than accuracy under imbalance. The proposed method reaches the highest Kappa (0.8604), implying that the observed performance gain reflects genuine improvement in class discrimination rather than an artifact of class frequencies. In comparison, SMOTE improves recall (0.9705) and accuracy (0.8944) but yields a lower Kappa (0.7874), suggesting that uniform interpolation may still introduce overlap or ambiguity, even in a low-attribute dataset. Random undersampling further reduces Kappa (0.6964), which is consistent with the risk of removing informative majority instances and weakening boundary definition, leading to less stable multi-class separation. The conventional hybrid approach produces the lowest Kappa (0.6094), indicating that loosely coupled oversampling–undersampling can distort the training distribution without effectively strengthening discriminative structure. The performance using Car Dataset can be seen in Figure 3 and Table 4.

Table 4. Performance using Car Dataset

Method	Accuracy	Precision	Recall	Kappa
Proposed Method	0.9306	0.894	1.0	0.8533
SMOTE	0.8947	0.866	0.9738	0.7748
Random Undersampling	0.8495	0.8467	0.9152	0.6798
Hybrid Sampling	0.8044	0.8254	0.8585	0.5862

Figure 3. Performance using Car Dataset

For the Car dataset, the results in Table 4 and Figure 3 indicate that the proposed method consistently outperforms the baseline sampling techniques across all evaluation metrics. Compared to the Balance dataset, the Car dataset exhibits a larger sample size and higher class complexity, which increases the risk of class overlap and boundary ambiguity under imbalance. The proposed method achieves the highest accuracy (0.9306) and precision (0.894) while attaining perfect recall (1.0), demonstrating its ability to identify all minority class instances without sacrificing overall classification quality. This indicates that the adaptive oversampling component effectively enriches informative minority regions, while the instance hardness-based undersampling prevents excessive false positives that often arise in multi-class settings with overlapping decision regions.

Cohen's Kappa further highlights the robustness of the proposed method, with a value of 0.8533 that substantially exceeds those of SMOTE (0.7748), Random Undersampling (0.6798), and Hybrid Sampling (0.5862). Given the increased number of classes, Kappa is particularly important because accuracy alone may mask biased predictions toward majority classes. The higher Kappa value suggests that the proposed method improves true class agreement beyond chance and yields more stable decision boundaries. In contrast, SMOTE improves recall and accuracy but still exhibits a lower Kappa, indicating that uniform synthetic generation may not sufficiently account for class-specific boundary difficulty in a four-class problem. Random Undersampling further degrades Kappa, reflecting the loss of informative majority instances that are crucial for defining multi-class boundaries. The conventional hybrid method performs the worst, suggesting that loosely coordinated oversampling and undersampling strategies are insufficient to handle the increased structural complexity of the Car dataset.

Overall, the Car dataset results emphasize an important trade-off in imbalanced multi-class learning: increasing sensitivity to minority classes while preserving reliable class separation. The proposed method effectively mitigates this trade-off by coordinating boundary-aware minority enrichment with difficulty-aware majority reduction, resulting in perfect recall, high precision, and the strongest agreement with ground truth. This demonstrates the method's ability to generalize effectively in datasets with higher sample size and moderate class complexity. The performance using Contraceptive Dataset can be seen in Figure 4 and Table 5.

Table 5. Performance using Contraceptive Dataset

Method	Accuracy	Precision	Recall	Kappa
Proposed Method	0.9301	0.8933	1.0	0.8523
SMOTE	0.8941	0.8661	0.9727	0.7734
Random Undersampling	0.8493	0.8464	0.9153	0.6792
Hybrid Sampling	0.8038	0.8252	0.8576	0.5849

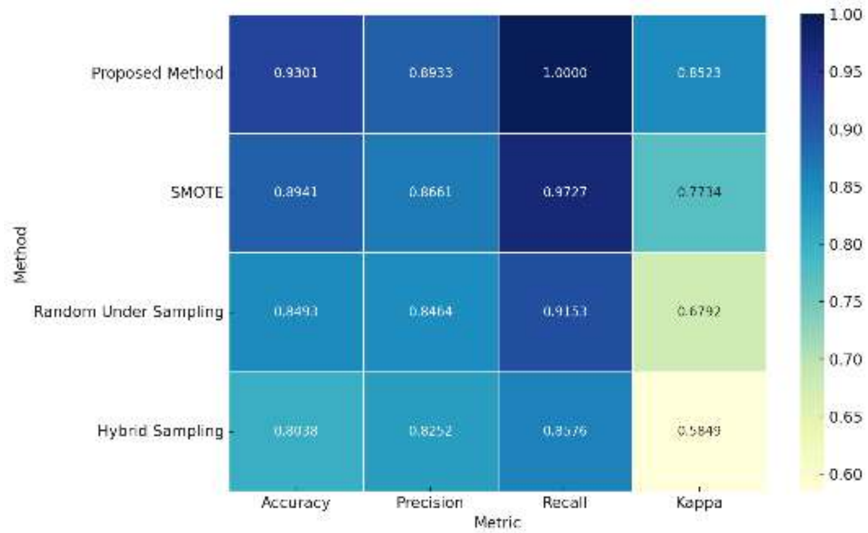

Figure 4. Performance using Contraceptive Dataset

For the Contraceptive dataset, the results in Table 5 and Figure 4 demonstrate that the proposed method consistently delivers the strongest overall performance compared to the baseline sampling techniques. The relatively higher number of attributes increases feature-space complexity, which can amplify class overlap and make minority patterns harder to distinguish under imbalance. In this context, the proposed method achieves the highest accuracy (0.9301) and precision (0.8933), while attaining perfect recall (1.0). This indicates that the method successfully captures all minority class instances without introducing excessive false positives, suggesting that the adaptive weighted oversampling strategy effectively focuses on informative regions rather than uniformly expanding the minority class.

Cohen's Kappa further provides a critical assessment of classification reliability in this dataset. The proposed method records the highest Kappa value (0.8523), substantially outperforming SMOTE (0.7734), Random Undersampling (0.6792), and Hybrid Sampling (0.5849). Given the increased dimensionality, Kappa is particularly important because accuracy alone may obscure misclassification patterns driven by dominant classes. The higher Kappa value indicates that the proposed method improves agreement with true labels beyond chance and produces more stable decision boundaries. In contrast, although SMOTE achieves relatively high recall and accuracy, its lower Kappa suggests that uniform synthetic generation may still introduce ambiguity in high-dimensional regions. Random Undersampling further degrades Kappa, reflecting the loss of boundary-relevant majority samples, while the conventional hybrid method performs the worst, indicating that loosely integrated resampling strategies struggle to preserve discriminative structure in higher-dimensional feature spaces.

Overall, the Contraceptive dataset results highlight the trade-off between sensitivity and reliability in imbalanced learning with moderate class complexity and higher dimensionality. The proposed method effectively mitigates this trade-off by coordinating boundary-aware minority enrichment with instance hardness-based majority reduction, resulting in perfect sensitivity, high precision, and the strongest agreement with ground truth. This confirms the method's robustness and generalization capability in imbalanced datasets with more complex feature representations. The performance using Ecoli Dataset can be seen in Figure 5 and Table 6.

Table 6. Performance using Ecoli Dataset

Method	Accuracy	Precision	Recall	Kappa
Proposed Method	0.9315	0.8959	1.0	0.8549
SMOTE	0.8899	0.8661	0.9652	0.7643
Random Undersampling	0.8452	0.8479	0.9064	0.6705
Hybrid Sampling	0.8006	0.8286	0.8488	0.578

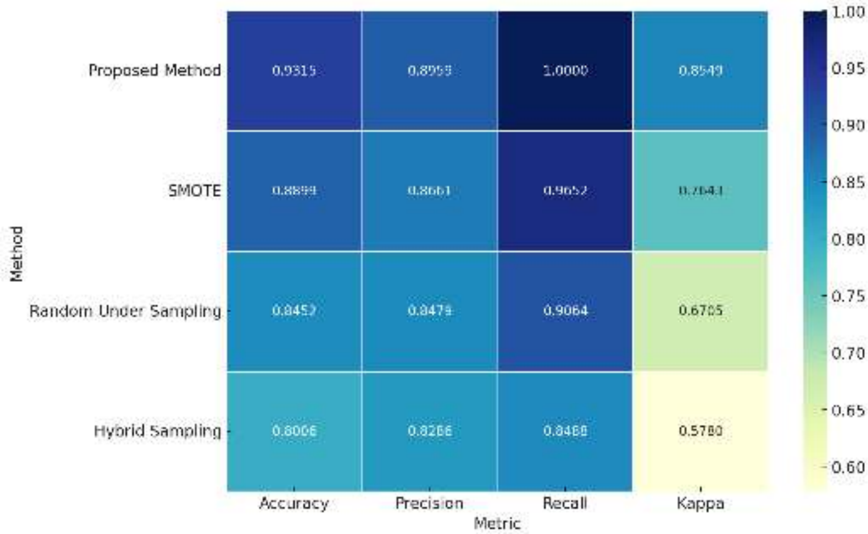

Figure 5. Performance using Ecoli Dataset

For the Ecoli dataset, the results in Table 6 and Figure 5 highlight the effectiveness of the proposed method in a highly challenging imbalanced multi-class scenario. Compared to the other datasets, Ecoli is characterized by a relatively small number of instances combined with a large number of classes, which increases class sparsity and exacerbates class overlap. Under such conditions, classifiers are prone to biased learning toward majority classes and unstable decision boundaries. The proposed method achieves the highest accuracy (0.9315) and precision (0.8959) while attaining perfect recall (1.0), indicating its strong ability to correctly identify all minority class instances despite limited data availability and increased class complexity. The maintenance of high precision alongside perfect recall suggests that minority enrichment is carefully controlled and does not lead to excessive false positives, which is a common risk in oversampling for small datasets.

Cohen's Kappa provides particularly important insight for this dataset, as accuracy alone can be misleading when class distributions are highly uneven across many classes. The proposed method attains the highest Kappa value (0.8549), substantially outperforming SMOTE (0.7643), Random Undersampling (0.6705), and Hybrid Sampling (0.578). This indicates a strong agreement between predicted and true labels beyond chance, reflecting more reliable and stable decision boundary learning. In contrast, SMOTE improves recall and accuracy but yields a lower Kappa, suggesting that uniform synthetic generation may still introduce ambiguity in sparse class regions. Random Undersampling further degrades Kappa, highlighting the risk of discarding informative majority samples in datasets where each class already has limited representation. The conventional hybrid approach performs the worst, indicating that loosely coordinated resampling strategies struggle to handle the combined challenges of small sample size and high class cardinality.

Overall, the Ecoli dataset results emphasize the critical trade-off between sensitivity and reliability in imbalanced learning with many classes and limited data. The proposed method effectively mitigates this trade-off by focusing oversampling on informative minority regions while selectively removing redundant majority instances based on instance hardness. This coordinated strategy leads to perfect sensitivity, high precision, and the strongest agreement with ground truth, demonstrating robust generalization even in highly sparse and complex multi-class settings. The performance using Yeast Dataset can be seen in Figure 6 and Table 7.

Table 7. Performance using Yeast Dataset

Method	Accuracy	Precision	Recall	Kappa
Proposed Method	0.9071	0.901	0.9817	0.8551
SMOTE	0.8867	0.8701	0.9598	0.7351
Random Undersampling	0.8413	0.8267	0.8976	06892
Hybrid Sampling	0.8102	0.8311	0.8491	0.6012

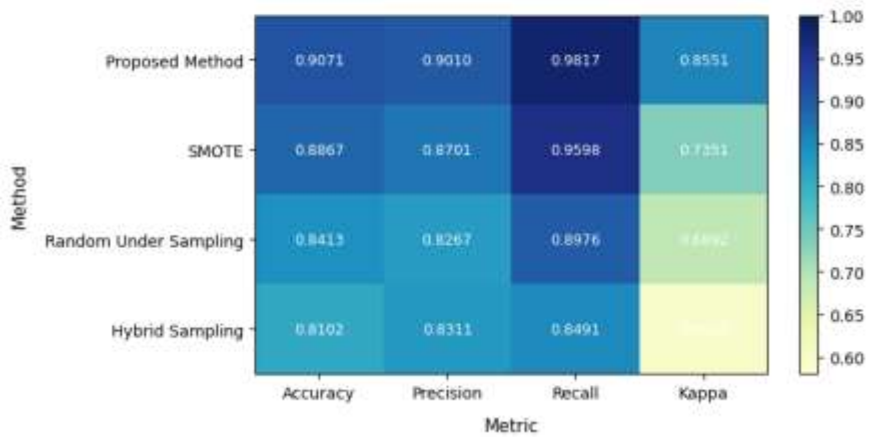

Figure 6. Performance using Yeast Dataset

For the Yeast dataset, the results in Table 7 and Figure 6 demonstrate the robustness of the proposed method in a highly complex multi-class imbalanced setting. Compared to other datasets, Yeast exhibits the highest number of classes combined with moderate dimensionality, which substantially increases class overlap and boundary ambiguity. In such scenarios, achieving high recall without degrading precision becomes particularly challenging. The proposed method attains the highest accuracy (0.9071) and precision (0.901), while maintaining a very high recall (0.9817), indicating that the majority of minority class instances are correctly identified without excessive false positives. Unlike simpler datasets where perfect recall is achieved, the slight reduction in recall reflects the inherent difficulty of separating ten classes with overlapping feature distributions rather than a weakness of the method.

Cohen's Kappa provides a more reliable indicator of performance under these conditions. The proposed method achieves a Kappa value of 0.8551, which is substantially higher than those of SMOTE (0.7351), Random Undersampling (0.6892), and Hybrid Sampling (0.6012). This suggests that the proposed approach yields more consistent and meaningful agreement between predicted and true labels beyond chance, even as class cardinality increases. In contrast, although SMOTE improves recall and accuracy, its lower Kappa indicates that uniform synthetic oversampling may still amplify overlap in densely populated regions. Random Undersampling further reduces Kappa, highlighting the risk of removing informative majority samples in datasets where class boundaries are already fragmented across many classes. The conventional hybrid approach performs the worst, suggesting that loosely integrated resampling strategies are insufficient to handle the compounded challenges of high class count and imbalance.

Overall, the Yeast dataset results highlight an important performance trade-off in imbalanced learning with high class cardinality: maintaining sensitivity to minority classes while preserving reliable class discrimination. The proposed method effectively mitigates this trade-off by concentrating minority enrichment on informative regions and selectively reducing redundant majority instances based on instance hardness. This leads to strong recall, the highest precision, and the most stable agreement level among all compared methods, demonstrating robust generalization in highly complex multi-class imbalanced datasets.

The experimental results obtained in this study are largely consistent with findings reported in previous research on resampling for imbalanced learning. Prior studies have shown that oversampling methods such as SMOTE improve minority class recall but may introduce class overlap and limited gains in classification reliability when applied uniformly. Boundary-aware and density-aware approaches, including MWMOTE, were introduced to address this limitation by prioritizing informative minority samples near decision boundaries, and our results support these findings by demonstrating improved recall and precision when minority enrichment is focused on difficult regions. However, previous works have also reported that oversampling alone is insufficient to fully mitigate majority class dominance, particularly in multi-class settings with overlapping distributions. The results of this study further confirm this observation, as methods relying solely on oversampling or undersampling exhibit lower Cohen's Kappa values, indicating reduced agreement beyond chance. In contrast, the proposed AWO-IHU method extends existing hybrid resampling approaches by explicitly coordinating boundary-aware minority enrichment with instance hardness-based majority reduction. The consistently higher Kappa and balanced precision–recall performance across datasets suggest that this coordinated design provides a more stable decision boundary than methods reported in earlier studies. Overall, the results support and extend previous findings by demonstrating that aligning oversampling and undersampling mechanisms in a principled manner leads to more reliable multi-class classification under class imbalance.

4. Discussion

The experimental results across all datasets demonstrate that the proposed method consistently outperforms SMOTE, Random Undersampling, and conventional Hybrid Sampling in terms of classification reliability and minority class sensitivity. These improvements can be directly linked to the coordinated mechanisms embedded in the proposed framework, rather than merely to changes in class distribution.

From a methodological perspective, the superior performance of proposed method is largely attributed to the complementary interaction between adaptive weighted oversampling and instance hardness-based undersampling. The oversampling component prioritizes informative minority instances located near decision boundaries, which enhances the classifier's ability to learn discriminative patterns in regions that are most prone to misclassification. Unlike uniform oversampling approaches such as SMOTE, this targeted strategy reduces the risk of generating synthetic samples in uninformative or noisy regions. Simultaneously, the undersampling component removes redundant majority instances that are consistently easy to classify, thereby reducing majority dominance while preserving boundary-defining samples. This coordinated design explains why the proposed method achieves high recall without a corresponding drop in precision across datasets with varying characteristics.

Comparative analysis further highlights the limitations of existing approaches. SMOTE improves recall by increasing minority representation, but its lower Cohen's Kappa values across datasets indicate that uniform synthetic generation may still introduce class overlap and reduce agreement beyond chance. Random Undersampling exhibits an even greater decline in Kappa, reflecting the loss of informative majority samples and weakened decision boundaries, especially in multi-class settings. Conventional hybrid sampling methods, which combine oversampling and undersampling without explicit coordination, consistently produce the weakest performance, suggesting that loosely integrated strategies may distort data distributions without effectively improving discriminative learning. In contrast, proposed method explicitly aligns minority enrichment with majority refinement, leading to more stable and reliable classification outcomes.

Despite these advantages, several limitations of the proposed approach should be acknowledged. First, the computation of instance hardness relies on an ensemble-based model, which introduces additional computational overhead compared to simpler resampling techniques. Second, the effectiveness of the method may be influenced by parameter choices such as the number of nearest neighbors, hardness thresholds, and ensemble size, which were fixed in this study. Although consistent performance was observed across datasets with different sizes and class cardinalities, further investigation is needed to assess sensitivity to these parameters. Additionally, the current evaluation focuses on static benchmark datasets;

the behavior of the method in streaming or highly dynamic environments remains an open question.

In terms of generalization, the results suggest that proposed method is particularly well-suited for imbalanced multi-class problems characterized by class overlap, sparsity, and boundary ambiguity. The consistent improvements observed across datasets with varying numbers of instances, attributes, and classes indicate that the proposed framework is not tailored to a specific dataset structure. Moreover, because the method operates at the data level and is classifier-independent, it has the potential to be integrated with other learning models, including deep learning architectures and domain-specific classifiers. Future work may explore adaptive parameter tuning, scalability to high-dimensional data, and application to real-world domains such as medical diagnosis, bioinformatics, and fault detection.

D. CONCLUSION AND SUGGESTIONS

This study introduces a hybrid resampling method, termed Adaptive Weighted Oversampling and Instance Hardness-Based Undersampling (AWO-IHU), to address class imbalance in multi-class classification. In the context of existing literature on hybrid resampling methods, the main scientific contribution of this work lies in the explicit coordination between boundary-aware minority oversampling and instance hardness-based majority undersampling. Unlike conventional hybrid approaches that apply oversampling and undersampling independently, the proposed method integrates both processes through a unified design that prioritizes decision boundary quality. This contributes to the literature by demonstrating that improving the structural properties of the training data is more effective than solely rebalancing class distributions.

Experimental results on multiple benchmark datasets show that the proposed method consistently outperforms widely used resampling techniques such as SMOTE, Random Undersampling, and conventional hybrid methods across accuracy, precision, recall, and Cohen's Kappa. The high recall values indicate strong minority sensitivity, while the simultaneously high precision and Kappa values suggest that these gains are not achieved through overly aggressive sampling or chance agreement. Nevertheless, the results also reveal that performance varies with dataset characteristics, such as class cardinality and sample size, indicating that the method is particularly effective in imbalanced multi-class settings with class overlap and boundary ambiguity rather than in trivial imbalance scenarios.

Despite its effectiveness, several limitations of the proposed method should be acknowledged. The instance hardness estimation relies on ensemble-based models, which introduces additional computational cost compared to simpler resampling strategies. Moreover, the method's performance may depend on parameter choices, including neighborhood size and hardness thresholds, which were fixed in this study. These factors may influence scalability and efficiency when applied to very large or high-dimensional datasets, suggesting that careful parameter selection is necessary in practical applications.

Future research directions are directly motivated by these findings and limitations. Promising extensions include adaptive parameter tuning to reduce sensitivity to user-defined settings, investigation of more efficient hardness estimation techniques to lower computational overhead, and integration of the proposed method with deep learning classifiers for high-dimensional data. In addition, applying the method to streaming or evolving data environments and evaluating its performance in real-world domains such as medical diagnosis, bioinformatics, and fault detection represent important avenues for further study. Overall, this work provides a principled and extensible hybrid resampling method that advances current approaches for imbalanced multi-class learning.

ACKNOWLEDGEMENT

This research was funded by the Ministry of Higher Education, Science, and Technology based on a contract between the LLDIKTI Region I and Universitas Medan Area, under Contract Number 7/SPK/LL1/AL.04.03/PL/2025.

REFERENCES

Alabduallah, B., Maray, M., Alruwais, N., Alabdan, R., Darem, A. A., Alallah, F. S., Alsini, R., & Yafoz, A. (2024). Class imbalanced data handling with cyberattack classification using Hybrid Salp Swarm Algorithm with deep learning approach. *Alexandria Engineering Journal*, *106*, 654–663. https://doi.org/10.1016/j.aej.2024.08.061

Altalhan, M., Algarni, A., & Turki-Hadj Alouane, M. (2025). Imbalanced Data Problem in Machine Learning: A Review. *IEEE Access*, *13*, 13686–13699. https://doi.org/10.1109/ACCESS.2025.3531662

Aymaz, S. (2025). Unlocking the power of optimized data balancing ratios: A new frontier in tackling imbalanced datasets. *The Journal of Supercomputing*, *81*(2), 443. https://doi.org/10.1007/s11227-025-06919-2

Barua, S., Islam, Md. M., Yao, X., & Murase, K. (2014). MWMOTE–Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning. *IEEE Transactions on Knowledge and Data Engineering*, *26*(2), 405–425. https://doi.org/10.1109/TKDE.2012.232

Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: A review from a data perspective. *Journal of Big Data*, *12*(1), 71. https://doi.org/10.1186/s40537-025-01119-4

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. *Artificial Intelligence Review*, *57*(6), 137. https://doi.org/10.1007/s10462-024-10759-6

Chiu, C. W., & Minku, L. L. (2024). Smoclust: Synthetic minority oversampling based on stream clustering for evolving data streams. *Machine Learning*, *113*(7), 4671–4721. https://doi.org/10.1007/s10994-023-06420-y

Elreedy, D., & Atiya, A. F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Information Sciences*, *505*, 32–64. https://doi.org/10.1016/j.ins.2019.07.070

Fachrie, M., Musdholifah, A., & Pulungan, R. (2025). Effectiveness of data resampling and ensemble learning in multiclass imbalance learning. *Artificial Intelligence Review*, *58*(12), 368. https://doi.org/10.1007/s10462-025-11357-w

Farou, Z., Wang, Y., & Horváth, T. (2024). Cluster-based oversampling with area extraction from representative points for class imbalance learning. *Intelligent Systems with Applications*, *22*, 200357. https://doi.org/10.1016/j.iswa.2024.200357

Han, F., Wang, C., Ling, Q., & Han, H. (2025). A denoising majority weighted minority oversampling technique for imbalanced classification. *Expert Systems with Applications*, *288*, 128199. https://doi.org/10.1016/j.eswa.2025.128199

Hancock, J., Khoshgoftaar, T. M., & Johnson, J. M. (2022). The Effects of Random Undersampling for Big Data Medicare Fraud Detection. *2022 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, 141–146. https://doi.org/10.1109/SOSE55356.2022.00023

Hassani, H., Entezarian, M. R., Zaeimzadeh, S., Marvian, L., & Komendantova, N. (2025). An oversampling-undersampling strategy for large-scale data linkage. *Frontiers in Big Data*, *8*. https://doi.org/10.3389/fdata.2025.1542483

Hoyos-Osorio, J., Alvarez-Meza, A., Daza-Santacoloma, G., Orozco-Gutierrez, A., & Castellanos-Dominguez, G. (2021). Relevant information undersampling to support imbalanced data classification. *Neurocomputing*, *436*, 136–146. https://doi.org/10.1016/j.neucom.2021.01.033

Huang, H., Liu, B., Xue, X., Cao, J., & Chen, X. (2024). Imbalanced credit card fraud detection data: A solution based on hybrid neural network and clustering-based undersampling technique. *Applied Soft Computing*, *154*, 111368. https://doi.org/10.1016/j.asoc.2024.111368

Li, D.-C., Shi, Q.-S., Lin, Y.-S., & Lin, L.-S. (2022). A Boundary-Information-Based Oversampling Approach to Improve Learning Performance for Imbalanced Datasets. *Entropy*, *24*(3), 322. https://doi.org/10.3390/e24030322

Lin, Z., Xu, Y., Liu, K., & Chen, L. (2026). MDGP-forest: A novel deep forest for multi-class imbalanced learning based on multi-class disassembly and feature construction enhanced by genetic programming. *Pattern Recognition*, *170*, 112070. https://doi.org/10.1016/j.patcog.2025.112070

M, P., Tyagi, B., R, N. P., & B, M. (2025). Hybrid Synthetic Minority Over-sampling Technique (HSMOTE) and Ensemble Deep Dynamic Classifier Model (EDDCM) for big data analytics. *Scientific Reports*, *15*(1), 39495. https://doi.org/10.1038/s41598-025-23062-3

Matharaarachchi, S., Domaratzki, M., & Muthukumarana, S. (2024). Enhancing SMOTE for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, *18*, 100597. https://doi.org/10.1016/j.mlwa.2024.100597

Nouas, S., Oukid, L., & Boumahdi, F. (2025). Syngo: Synthetic genetic oversampling technique for textual data. *Social Network Analysis and Mining*, *15*(1), 9. https://doi.org/10.1007/s13278-025-01423-0

Olabisi, O., Maurya, L., & Bader-El-Den, M. (2025). Localised ensemble learning (LEL) – a localised approach to class imbalance. *Pattern Analysis and Applications*, *28*(4), 177. https://doi.org/10.1007/s10044-025-01545-3

Razali, M. N., Arbaiy, N., Lin, P.-C., & Ismail, S. (2025). Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets. *Electronics*, *14*(4), Article 4. https://doi.org/10.3390/electronics14040705

Salehi, A., & Khedmati, M. (2025). Hybrid clustering strategies for effective oversampling and undersampling in multiclass classification. *Scientific Reports*, *15*(1), 3460. https://doi.org/10.1038/s41598-024-84786-2

Shi, S., Li, J., Zhu, D., Yang, F., & Xu, Y. (2023). A hybrid imbalanced classification model based on data density. *Information Sciences*, *624*, 50–67. https://doi.org/10.1016/j.ins.2022.12.046

Shyalika, C., Wickramarachchi, R., El Kalach, F., Harik, R., & Sheth, A. (2024). Evaluating the Role of Data Enrichment Approaches towards Rare Event Analysis in Manufacturing. *Sensors (Basel, Switzerland)*, *24*(15), 5009. https://doi.org/10.3390/s24155009

Taskiran, S. F., Turkoglu, B., Kaya, E., & Asuroglu, T. (2025). A comprehensive evaluation of oversampling techniques for enhancing text classification performance. *Scientific Reports*, *15*(1), 21631. https://doi.org/10.1038/s41598-025-05791-7

Wang, A. X., Chukova, S. S., & Nguyen, B. P. (2023). Synthetic minority oversampling using edited displacement-based k-nearest neighbors. *Applied Soft Computing*, *148*, 110895. https://doi.org/10.1016/j.asoc.2023.110895

Xie, Y., Huang, X., Qin, F., Li, F., & Ding, X. (2024). A majority affiliation based under-sampling method for class imbalance problem. *Information Sciences*, *662*, 120263. https://doi.org/10.1016/j.ins.2024.120263

Zhang, R., Lu, S., Yan, B., Yu, P., & Tang, X. (2023). A density-based oversampling approach for class imbalance and data overlap. *Computers & Industrial Engineering*, *186*, 109747. https://doi.org/10.1016/j.cie.2023.109747

Zhong, X., Zhang, L., & Ban, H. (2023). Deep reinforcement learning for class imbalance fault diagnosis of equipment in nuclear power plants. *Annals of Nuclear Energy*, *184*, 109685. https://doi.org/10.1016/j.anucene.2023.109685