Pertanika J. Sci. & Technol. 34 (4): 2339 - 2358 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Hybrid Approach Combining Area-Based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE) in Handling Class Imbalance

Hartono[1]*, Muhammad Khahfi Zuhanda[1], Rahmad Syah[1], Prana Ugiana Gio[2], and Erianto Ongko[3]

[1]Master's Program in Informatics, Postgraduate School, Universitas Medan Area, 20223 Medan, North Sumatra, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, 20155 Medan, North Sumatra, Indonesia
[3]Department of Computer Science, Institut Modern Arsitektur dan Teknologi, 20371 Medan, North Sumatra, Indonesia

ABSTRACT

Machine learning models can effectively extract meaningful patterns from complex data, but their performance often degrades when the data distribution is highly imbalanced. Class imbalance, where one class contains significantly fewer instances than others, can lead to biased predictions and poor minority-class recognition. To address this issue, this study proposes a hybrid framework combining Area-based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation Algorithm-SMOTE (WOA-SMOTE). AROSS generates synthetic samples from safe and semi-safe regions, reducing overfitting, while WOA-SMOTE optimises neighbour selection to improve sample diversity and representativeness. The proposed method was evaluated on five imbalanced datasets from the KEEL repository using F1-score, Precision, Recall, and MCC. Experimental results demonstrate that AROSS-WOA-SMOTE achieves competitive performance and outperforms SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE on most datasets and evaluation metrics.

Keywords: Area-based Representative Points Oversampling with Shifting (AROSS), class imbalance, Machine Learning, Synthetic Minority Oversampling Technique (SMOTE), Whale Optimisation Algorithm (WOA)

ARTICLE INFO

Article history:
Received: 29 June 2026
Accepted: 01 August 2026
Published: 20 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.19

E-mail addresses:
hartono@staff.uma.ac.id (Hartono)
khahfi@staff.uma.ac.id (Muhammad Khahfi Zuhanda)
rahmadsyah@uma.ac.id (Rahmad Syah)
prana@usu.ac.id (Prana Ugiana Gio)
eriantongko@imat.ac.id (Erianto Ongko)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

Data mining techniques are crucial in extracting fundamental knowledge from the data. Data classification plays a crucial role in addressing various practical problems in data mining (Ait Issad et al., 2019). However, data gathered from the real-world frequently display an imbalance and may contain confidential personal information (Tsai et al., 2019). When building machine learning models for classification (Le et al., 2023). It is common to assume that instances are evenly spread out across all classes (Parmar et al., 2023). The problem of class imbalance in a dataset is a major problem in machine learning (Luong et al., 2025), where one class is much more common than the others. (Rezvani & Wang, 2023).

In a dataset with class imbalance, one class has more data points than the others. The majority class has more data items, while the minority class has fewer (Khan et al., 2024). Standard classification algorithms struggle to accurately classify datasets of this type, leading to a bias toward the majority class (Paul et al., 2024). The overall classification accuracy for CI datasets might be as high as 98%, but regular machine-learning models would misclassify many of the data samples. The model's reliability in medical diagnosis is diminished due to its inability to accurately identify patients. If the model misclassifies something (Bikku et al., 2025), it could lead to more medical problems (Bikku, 2020) and even death (Thabtah et al., 2020).

Three methodologies are frequently implemented to resolve the issue of class imbalance: data-level approach, algorithm-level approach, and hybrid-level approach (Shi et al., 2023). Oversampling (making copies of instances of the minority class or making synthetic samples) and undersampling (cutting down on instances of the majority class) are two common ways to rebalance data at the data level (Wang et al., 2023). At the algorithmic level, existing classification algorithms are changed to make them work better with imbalanced datasets. (Tao et al., 2020). Hybrid methods fix class imbalance by using better classifiers and strategies for rebalancing data. (Chen et al., 2021).

The data level methods are the most commonly used of these approaches because they focus on rebalancing the dataset to get an even distribution of the classes (Szeghalmy & Fazekas, 2024). Moreover, these methods make sure that the sampling process and classifier training are separate, which leads to a strong and effective way to handle imbalanced data (Soltanzadeh et al., 2023). The Synthetic Minority Oversampling Technique (SMOTE) is a widely used method for addressing class imbalance by generating synthetic minority samples (Chen et al., 2022).

However, one significant drawback of SMOTE and its variants is the potential for overfitting the data, which diminishes the capacity to generalise when applied to new and unseen datasets (Meng & Li, 2022). Two efforts can be made to improve SMOTE performance: searching for samples in the safe area (Syakiylla Sayed Daud et al., 2023) and searching for each sample's nearest neighbour (Sun et al., 2024). Several researchers

have attempted to use one of the two efforts, as done by: Tyagi et al. 2024 who proposed Whale Optimisation-based Synthetic Minority Oversampling Technique (WOA-SMOTE), Xu et al. 2021 who proposed a cluster-based oversampling algorithm (KNSMOTE), Farou et al. 2024 who proposed AROSS.

This study combines sample determination based on safe and semi-safe areas and the determination of k neighbours that are appropriate for handling class imbalance. This is done by combining the application of AROSS, which is intended for determining safe and semi-safe areas, with WOA-SMOTE, which can help in determining the right k neighbours.

Existing oversampling methods still suffer from several limitations. Methods such as AROSS focus on identifying safe and semi-safe regions for synthetic sample generation, but they rely on predefined neighbourhood structures that may not accurately capture the local characteristics of different datasets. In contrast, WOA-SMOTE adaptively optimises the neighbourhood parameter but does not explicitly consider the reliability of the regions from which synthetic samples are generated. Consequently, synthetic instances may still be created in noisy or overlapping areas, which can reduce classification performance and increase the risk of overfitting. Therefore, existing approaches generally address either region reliability or neighbourhood optimisation separately, while the joint optimisation of these two aspects remains insufficiently explored.

Motivated by this limitation, this study investigates whether integrating region-aware sample selection with adaptive neighbourhood optimisation can improve the quality of synthetic samples and enhance classification performance on imbalanced datasets. Specifically, this study addresses the following research questions: (1) Can the combination of representative-region extraction and adaptive neighbourhood optimisation improve classification performance on imbalanced datasets? (2) Does the proposed AROSS-WOA-SMOTE framework outperform existing oversampling methods in terms of MCC, Precision, Recall, and F1-score? (3) To what extent do region selection and neighbourhood optimisation contribute to the effectiveness of synthetic sample generation?

Based on these research questions, we hypothesise that combining representative-region selection with adaptive neighbourhood optimisation can generate more informative synthetic samples and provide better classification performance than conventional oversampling methods. To validate this hypothesis, this study proposes a hybrid framework that integrates Area-based Representative Points Oversampling with Shifting (AROSS) and Whale Optimisation-SMOTE (WOA-SMOTE), enabling the oversampling process to simultaneously consider the reliability of minority regions and the suitability of neighbourhood structures. This integration forms the basis of the proposed framework and motivates the contributions presented in this work.

The novelty of the proposed framework lies not merely in combining AROSS and WOA-SMOTE, but in introducing a hierarchical oversampling mechanism that jointly

addresses two critical limitations of existing oversampling techniques: the generation of synthetic samples in unreliable regions and the selection of suboptimal neighbourhood structures. Unlike conventional approaches that treat sample generation and neighbour selection as independent processes, the proposed framework integrates region-aware sample selection with adaptive neighbourhood optimisation.

First, AROSS identifies safe and semi-safe regions by extracting representative points from clusters and analysing the local distribution of minority instances. This mechanism prevents the generation of synthetic samples in noisy and overlapping regions, thereby reducing the risk of overfitting.

Second, instead of employing a fixed number of nearest neighbours, the framework incorporates the Whale Optimisation Algorithm (WOA) to dynamically determine the optimal neighbourhood structure for each oversampling process. By optimising the value of (k), the proposed method generates synthetic instances that better preserve the intrinsic characteristics of the minority class while increasing class separability.

Therefore, the scientific contribution of this study extends beyond a simple hybridisation of AROSS and WOA-SMOTE and can be summarised as follows:

1. Introducing a two-stage oversampling framework that combines region-based minority sample selection with adaptive neighbour optimisation.
2. Developing a mechanism that restricts synthetic sample generation to safe and semi-safe regions, thereby reducing noise and class overlap.
3. Integrating metaheuristic optimisation into the oversampling process to automatically determine the most suitable neighbourhood configuration instead of relying on predefined parameters.
4. Improving the representativeness and diversity of synthetic samples, resulting in more robust classification performance on multi-class imbalanced datasets.
5. Demonstrating the effectiveness of the proposed framework on several benchmark datasets with varying imbalance ratios, confirming its applicability across different classification scenarios.

These contributions provide a new perspective on imbalanced learning by showing that the joint optimisation of sampling regions and neighbourhood structures is more effective than treating them as separate tasks.

LITERATURE REVIEW

Generally, we can formulate mathematical models to handle class imbalances through optimisation. Suppose f is the objective function that we want to optimise, such as the loss function in machine learning. The mathematical model in Equation 1:

$$\min_{\theta} \sum_{i=1}^{N} w_i \cdot f(x_i, y_i, \theta) \quad [1]$$

where:

θ is the model parameter.

x_i is the i-th input sample.

y_i is the label of the i-th sample.

w_i is the weight given to the i-th sample, which can be adjusted based on its class.

The weight w_i can be determined as in Equation 2:

$$w_i = \begin{cases} \frac{1}{N_1}, if\ y_i\ is\ majority\ class \\ \frac{1}{N_2}, if\ y_i\ is\ minority\ class \end{cases} \quad [2]$$

This model accounts for imbalance by giving greater weight to samples from minority classes, making the model more sensitive to underrepresented classes.

AROSS Algorithm

The pseudocode of AROSS is as follows.

Input:	
a.	D: Original dataset
b.	IR : Desired imbalance ratio
c.	σ : Standard deviation parameter
d.	max_iter: Maximum number of iterations
Output: D': Augmented dataset	
1.	Standardise dataset D using z-score normalisation.
2.	Perform agglomerative clustering and determine the best linkage method using the Cophenetic Correlation Coefficient (CPCC).
3.	Extract representative points from each cluster.
4.	Classify each representative point as safe, semi-safe, or unsafe using k-nearest neighbours.
5.	Expand safe and semi-safe regions incrementally.
6.	Compute the total number of synthetic samples required based on the desired imbalance ratio.
7.	Distribute synthetic samples among safe and semi-safe regions according to the minority-instance distribution.
8.	Generate candidate samples using a Gaussian generator and validate the generated samples.
9.	Add valid samples to the synthetic dataset.
10.	Combine the original dataset with the synthetic samples to obtain D'.
11.	Return D'.

AROSS is a cluster-based adaptive oversampling method to address the problem of class imbalance in datasets. The process begins with the standardisation of the dataset using

z-score normalisation, followed by optimised agglomerative clustering to form clusters. From each cluster, representative points are extracted and classified as safe, semi-safe, or unsafe using k-NN. The safe and semi-safe areas are then expanded using an incremental k-NN strategy. The number of synthetic samples to be generated is determined based on the desired imbalance ratio and the weights of the safe and semi-safe areas. Synthetic samples are then generated using a hyperspherical-based truncated Gaussian generator. The original dataset is then combined with the synthetic samples to form an augmented dataset, which improves the performance of machine learning algorithms under class imbalance conditions.

Whale Optimisation Based SMOTE (WOA-SMOTE)

The pseudocode of WOA-SMOTE is as follows.

Input:

a. Training dataset D
b. Population size P
c. Maximum iterations T
d. Neighbour search bounds [k_{min}, k_{max}]

Output: Best k-neighbours value and synthetic dataset.

1. Initialise the whale population within the search space [k_{min}, k_{max}].
2. Evaluate the fitness of each whale using the F1-score.
3. For each iteration from 1 to T:
   a. Select the best whale as the leader.
   b. Update the position of each whale using the WOA mechanism.
   c. Recalculate the fitness values.
   d. Update the leader if a better solution is found.
5. Determine the optimal value of k-neighbours from the best whale.
6. Generate synthetic minority samples using SMOTE with the optimal k-neighbours value.
7. Return the synthetic dataset.

For each whale W_j, generate synthetic minority class samples using the SMOTE algorithm as in Equation 3:

$$s_i = x_i + \delta.(x_{nn} - x_i) \quad [3]$$

WOA-SMOTE employs the Whale Optimisation Algorithm (WOA) to determine the optimal value of the *k*-neighbours parameter in SMOTE. This fixes the problem of having too many instances in one class. First, a population of whale agents is initialised using random k-neighbour values. The fitness of each agent is evaluated using the F1-score. The agents' positions are updated based on the leader's position at each step, and the best solution (leader) is chosen. To make the search easier, the values of A and C are changed on the fly. The optimisation process continues until the maximum number of iterations is

reached. Once the best value for neighbours is found, SMOTE is used on the training data to generate synthetic samples that help balance the dataset. The final product is a synthetic dataset with a better class balance, which helps classifiers achieve superior performance.

Comparative Analysis of Existing Oversampling Methods

Existing oversampling methods address class imbalance from different methodological perspectives and rely on different assumptions regarding the data distribution. Clustering-based methods, such as AROSS and KNSMOTE, assume that minority instances form meaningful local structures that can be exploited to generate representative synthetic samples. Their main strength lies in reducing the generation of synthetic samples in noisy regions; however, their performance strongly depends on the quality of the clustering process and the predefined neighbourhood structure.

Safe-region approaches, such as Safe-Level SMOTE and AROSS, assume that minority samples located in safe and semi-safe regions are more reliable for synthetic sample generation. These methods can reduce class overlap and noise, but they may fail to adapt to datasets with heterogeneous local distributions because they typically employ a fixed neighbourhood configuration.

Neighbourhood-optimisation approaches, such as WOA-SMOTE, focus on improving the selection of nearest neighbours by dynamically searching for suitable neighbourhood parameters. Although these methods enhance the diversity of synthetic samples, they do not explicitly incorporate information regarding the reliability of the sampling regions, which may lead to the generation of samples near overlapping boundaries.

In contrast, Borderline-SMOTE concentrates on minority instances located near decision boundaries. While this strategy can improve the representation of difficult samples, it is more sensitive to noise and class overlap because synthetic instances are intentionally generated in ambiguous regions.

Therefore, existing methods generally address only one aspect of the class-imbalance problem, such as clustering, safe-region extraction, or neighbourhood optimisation. The joint consideration of region reliability and adaptive neighbourhood selection remains insufficiently explored. Motivated by this limitation, the proposed AROSS-WOA-SMOTE framework integrates region-aware sample selection with neighbourhood optimisation to improve the quality of synthetic sample generation.

METHODOLOGY

This study proposes the AROSS-WOA-SMOTE method. The proposed AROSS-WOA-SMOTE framework assumes that the quality of synthetic samples depends not only on the neighbourhood size used during oversampling but also on the reliability of the regions from which the samples are generated. Therefore, AROSS and WOA-SMOTE address two complementary aspects of the class-imbalance problem.

Specifically, AROSS identifies representative minority instances and classifies them into safe, semi-safe, and unsafe regions according to the local data distribution. This information is subsequently transferred to the WOA optimisation stage, where the search for the optimal neighbourhood parameter is constrained to the safe and semi-safe regions identified by AROSS. Consequently, the neighbourhood optimisation process is guided by local structural information rather than relying solely on the global characteristics of the dataset.

The integration of the two methods improves synthetic-sample quality in two ways. First, restricting oversampling to reliable regions reduces the risk of generating noisy and overlapping samples near class boundaries. Second, the adaptive optimisation of the neighbourhood parameter enables the oversampling process to capture local minority patterns more effectively, thereby producing synthetic instances that better preserve the underlying data distribution.

Unlike ordinary parameter tuning in AROSS or SMOTE, which determines the neighbourhood size globally, the proposed framework jointly performs region-aware sample selection and neighbourhood optimisation. As a result, the behaviour of the proposed method cannot be reproduced by simply tuning the parameters of AROSS or SMOTE, since the optimisation process is explicitly constrained by the representative regions identified during the AROSS stage.

The pseudocode of the proposed method is as follows:

Input:

Original dataset D, imbalance ratio IR, standard deviation σ, maximum iterations, whale population size, WOA iterations, and neighbour bounds.

1. Standardise dataset D using z-score normalisation.
2. Perform agglomerative clustering and select the best linkage method using CPCC.
3. Extract representative points from each cluster.
4. Classify representative points into safe, semi-safe, and unsafe regions using k-nearest neighbours.
5. Expand safe and semi-safe regions using incremental neighbourhood analysis.
6. Determine the number of synthetic samples for each region.
7. Optimise the neighbourhood size k^* using WOA based on the training-fold F1-Score
8. For each representative point in a safe or semi-safe region, identify its k^* nearest minority neighbours
9. Select one neighbour and generate a synthetic sampling using the SMOTE interpolation e
10. Return D'.

The AROSS-WOA-SMOTE method solves the problem of class imbalance by normalising the dataset and using agglomerative clustering to form clusters. Using k-NN, points that are typical of each cluster are put into three groups: safe, semi-safe, and unsafe. Incremental k-NN makes secure and semi-secure zones bigger. The number of synthetic samples to make depends on how much of an imbalance you want. The Whale Optimisation Algorithm (WOA) is then used to find the best value for the k_neighbours parameter for

SMOTE. The optimised k-neighbours value is then used to generate synthetic samples. Finally, the original dataset and the samples that were made are put together to construct an augmented dataset. This improves the performance of machine-learning algorithms on imbalanced datasets.

The AROSS-WOA-SMOTE framework addresses the class-imbalance problem by first normalising the dataset and applying agglomerative clustering to identify its underlying structure. Representative points are extracted from each cluster and classified into safe, semi-safe, and unsafe regions using the k-nearest neighbours' algorithm. The safe and semi-safe regions are subsequently expanded through an incremental k-NN strategy, and the required number of synthetic samples is determined according to the target imbalance ratio.

Next, the Whale Optimisation Algorithm (WOA) is employed to determine the optimal value of the k-neighbours parameter for the SMOTE process. The optimised neighbourhood parameter is then used to generate synthetic minority samples within the safe and semi-safe regions identified by AROSS. Finally, the original dataset and the generated samples are combined to construct an augmented dataset for classifier training and evaluation.

To provide a clearer overview of the proposed method, Figure 1 illustrates the complete workflow of the AROSS-WOA-SMOTE framework. The process begins with data preprocessing using z-score normalisation, followed by agglomerative clustering and representative-point extraction. The extracted points are categorised into safe and semi-safe regions to guide the generation of reliable synthetic samples.

Subsequently, WOA determines the optimal neighbourhood parameter for SMOTE, and the resulting value is used to generate additional minority-class samples while preserving

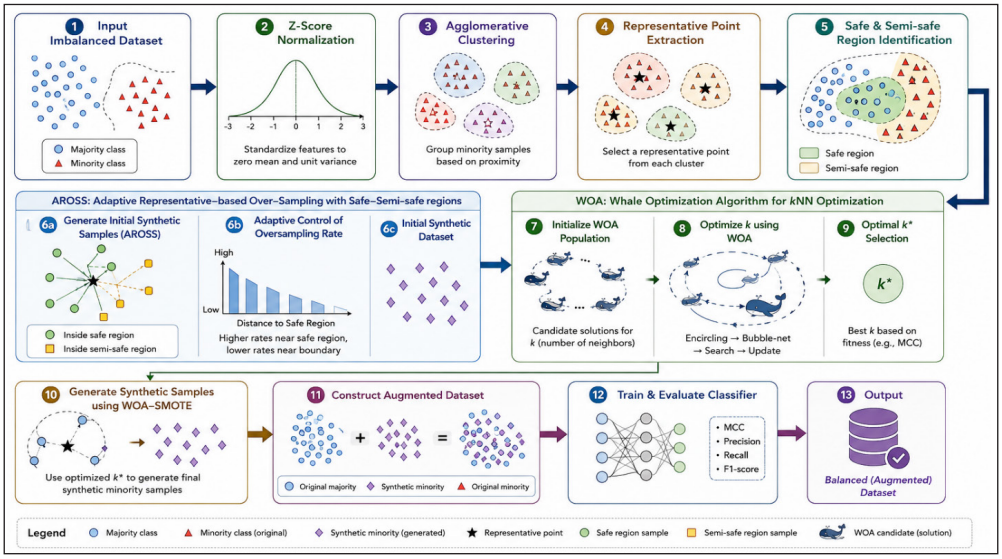

Figure 1. Workflow of the proposed AROSS-WOA-SMOTE framework

the intrinsic characteristics of the original data distribution. The augmented dataset is then used to train and evaluate the classification model. Therefore, AROSS determines the reliable regions for oversampling, whereas WOA-SMOTE determines the neighbourhood structure used in the final synthetic-sample generation process

Implementation Details

The experiments were conducted using Python 3.11. For WOA optimisation, the population size was set to 30, while the maximum number of iterations was fixed at 100. The search range for k-neighbours was defined between 3 and 15, and the F1 score was used as the fitness function. For AROSS, the standard deviation parameter (σ) was set to 0.1, and the maximum iteration for safe-region expansion was set to 50. The desired imbalance ratio was fixed at 1:1. The optimisation process stops when the maximum number of iterations is reached or when the best fitness value remains unchanged for ten consecutive iterations. The experiments were repeated ten times using ten different random seeds to evaluate the robustness and stability of the proposed framework.The hyperparameter settings of all methods can be found in Table 1.

The experimental evaluation was conducted using stratified 10-fold cross-validation to preserve the original class distribution in each fold. The dataset was randomly divided into ten approximately equal subsets, where, in each iteration, nine folds (90% of the data) were used for training, and the remaining fold (10% of the data) was used for testing. This process was repeated ten times until each fold had served exactly once as the test set.

Table 1
Hyperparameter settings of all methods

Method	Parameter	Value	Selection Strategy
SMOTE	k-neighbours	5	Default
Borderline-SMOTE	k-neighbours	5	Default
Borderline-SMOTE	m-neighbours	10	Default
Safe-Level SMOTE	k-neighbours	5	Default
AROSS	Standard deviation (σ)	0.1	Empirical
AROSS	Maximum iterations	50	Empirical
WOA-SMOTE	Population size	30	Empirical
WOA-SMOTE	WOA iterations	100	Empirical
WOA-SMOTE	Lower bound (kmin)	3	Empirical
WOA-SMOTE	Upper bound (kmax)	15	Empirical
AROSS-WOA-SMOTE	Imbalance ratio	1:1	Empirical
AROSS-WOA-SMOTE	Random seed	42	Fixed
AROSS-WOA-SMOTE	Fitness function	F1-score	Empirical
AROSS-WOA-SMOTE	Stopping criterion	No improvement for 10 iterations	Empirical

During the training phase, the Whale Optimisation Algorithm (WOA) was employed to determine the optimal value of the k-neighbours parameter by maximising the F1-score computed only on the training folds. Subsequently, the oversampling process was performed using the optimised parameter, and the classifier was trained on the augmented training set.

To avoid information leakage and biased performance estimates, the test fold was completely excluded from the optimisation and oversampling procedures and was used only for the final evaluation. Therefore, the F1-score used as the fitness function was never computed on the test data.

Although nested cross-validation could further reduce optimisation bias, the same optimisation and validation procedures were consistently applied to all competing methods to ensure a fair comparison. Future work will investigate the use of nested optimisation frameworks and analyse their computational implications.

Mathematical Notation

Notation used throughout the manuscript can be seen in Table 2.

The computational complexity of the proposed AROSS-WOA-SMOTE framework consists of two main stages: region extraction using AROSS and neighbourhood optimisation using WOA.

The agglomerative clustering process in AROSS requires $O(n^2)$ complexity, where n is the number of instances. Meanwhile, the optimisation stage of WOA-SMOTE requires $O(P \times T \times k)$, where P denotes the population size, T is the maximum number of iterations, and k represents the number of nearest neighbours.

Therefore, the overall computational complexity of the proposed framework can be expressed as $O(n^2 + P \times T \times k)$. The Time complexity comparison can be seen in Table 3.

In addition to classification performance, computational efficiency is an important aspect in evaluating oversampling methods.

Table 2
List of symbols and descriptions

Symbol	Description
D	Original dataset
x_i	The i-th sample
y_i	Class label of sample x_i
n	Number of samples
d	Number of features
m	Number of classes
IR	Imbalance ratio
C_i	The i-th cluster
c	Total number of clusters
R_i	Representative points extracted from cluster C_i
r_j	The j-th representative point
R_s	Set of safe regions
R_h	Set of semi-safe regions
k	Number of nearest neighbours
k'	Number of representative points
δ	Shifting factor
σ	Standard deviation parameter
η_a	Number of synthetic samples in area a
P	Whale population size
T_{AROSS}	Maximum AROSS iterations
T_{WOA}	Maximum WOA iterations
A	WOA exploration coefficient
C	WOA exploitation coefficient
k_{min}	Lower bound of neighbours
k_{max}	Upper bound of neighbours
D`	Augmented dataset

Therefore, the execution time of each method was measured to analyse the computational cost associated with the oversampling process. Table 4 presents the execution time, in seconds, of SMOTE, Safe-Level SMOTE, Borderline-SMOTE, AROSS, WOA-SMOTE, and the proposed AROSS-WOA-SMOTE framework across different datasets.

The performance of the proposed method was evaluated using MCC, Precision, Recall, and F1-score.

Table 3
Time complexity comparison

Method	Time Complexity
SMOTE	O(nk)
Safe-Level SMOTE	O (nk + n^2)
Borderline-SMOTE	O (nk + n^2)
AROSS	O(n^2)
WOA-SMOTE	O (P × T × k)
AROSS-WOA-SMOTE	O (n^2 + P × T × k)

Note. *n*: number of samples, *k*: number of nearest neighbours, *P*: whale population size, *T*: number of optimisation iterations

Table 4 shows that the proposed AROSS-WOA-SMOTE framework requires a longer execution time than conventional SMOTE due to the additional clustering and optimisation stages. However, the computational overhead remains acceptable and is compensated for by improvements in classification performance.

Table 4
Execution time comparison (Seconds)

Dataset	SMOTE	Safe-Level	Borderline	AROSS	WOA-SMOTE	AROSS-WOA-SMOTE
Flare	0.12	0.24	0.31	0.38	0.54	0.67
Glass 5	0.05	0.10	0.13	0.16	0.24	0.31
Thyroid Disease	0.09	0.20	0.25	0.29	0.43	0.56
Yeast-6	0.18	0.45	0.53	0.61	0.82	1.03
Red Wine Quality	0.21	0.61	0.68	0.74	0.95	1.18

RESULT AND DISCUSSIONS

Dataset

The multi-class imbalanced datasets used in this study were sourced from the KEEL Repository (Alcalá-Fdez et al., 2009). The dataset used can be seen in Table 5.

Table 5
Dataset description

Dataset	Instances	Attributes	Imbalance Ratio
Flare	1066	11	7.7
Glass 5	214	9	22.8
Thyroid Disease	720	21	39.18
Yeast-6	1484	8	41.38
Red Wine Quality	1599	11	68.10

Based on Table 5, the dataset selection reflects variations in terms of the number of instances, number of attributes, and imbalance ratio, so that it is expected that the test results will yield results that are close to real-world problems.

Confusion Matrix

The evaluation of classifier performance will be conducted using the MCC, Precision, Recall, and F1 Score(Luque et al., 2019). MCC, Precision, Recall, and F1 Score. MCC, Precision, Recall, and F1 Score calculations can be seen in Equations 4 to 7 (Hartono & Ongko, 2021).

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TN+FN)(TN+FP)(TN+FN)(TP+FP)}} \quad [4]$$

$$Precision = \frac{TP}{TP+FP} \quad [5]$$

$$Recall = \frac{TP}{TP+FN} \quad [6]$$

$$F1\ Score = \frac{2\ x\ Precision\ \ x\ Recall}{Precision\ +Recall} \quad [7]$$

Although the datasets used in this study are multiclass imbalanced datasets, the evaluation was performed directly in the multiclass setting. The equation presented above corresponds to the standard binary formulation of the Matthews Correlation Coefficient (MCC); however, the experiments in this study employed the generalised multiclass formulation, which extends the binary definition to multiclass classification tasks while preserving the correlation interpretation between the predicted and actual labels. Precision, recall, F1 Score, and MCC were subsequently calculated for each class, and the final performance values were obtained by aggregating the results across all classes.

Results

The MCC, Precision, Recall, and F1 Score values obtained can be seen in Table 6. Comparison will be made to measure the proposed method using AROSS-WOA-SMOTE compared to the SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE methods.

To further investigate the quality of the synthetic samples generated by the proposed method, Figure 2 presents a two-dimensional projection of the original data and the synthetic samples produced by SMOTE, WOA-SMOTE, Borderline-SMOTE, Safe-Level SMOTE, and AROSS-WOA-SMOTE. The visualisation illustrates the distribution of the generated samples and their relationship with the majority and minority classes.

Table 6
MCC, precision, recall, and F1 score

Dataset	Method	MCC	Precision	Recall	F1 Score
Flare	AROSS-WOA-SMOTE	0.981	0.928	0.935	0.931
	Borderline-SMOTE	0.960	0.922	0.930	0.926
	AROSS	0.954	0.913	0.927	0.920
	Safe-Level SMOTE	0.940	0.915	0.921	0.918
	WOA-SMOTE	0.932	0.919	0.919	0.919
	SMOTE	0.923	0.908	0.911	0.909
Glass 5	AROSS-WOA-SMOTE	0.976	0.917	0.941	0.929
	Borderline-SMOTE	0.950	0.915	0.936	0.925
	AROSS	0.927	0.912	0.931	0.921
	Safe-Level SMOTE	0.924	0.908	0.928	0.918
	WOA-SMOTE	0.931	0.903	0.927	0.915
	SMOTE	0.915	0.897	0.921	0.909
Thyroid Disease	AROSS-WOA-SMOTE	0.872	0.855	0.918	0.885
	Borderline-SMOTE	0.870	0.848	0.915	0.880
	WOA-SMOTE	0.881	0.831	0.912	0.870
	Safe-Level SMOTE	0.874	0.826	0.904	0.864
	AROSS	0.873	0.821	0.901	0.859
	SMOTE	0.867	0.811	0.897	0.852
Yeast-6	AROSS-WOA-SMOTE	0.921	0.918	0.932	0.925
	Borderline-SMOTE	0.918	0.916	0.930	0.923
	WOA-SMOTE	0.892	0.901	0.923	0.912
	Safe-Level SMOTE	0.910	0.904	0.920	0.914
	AROSS	0.911	0.911	0.901	0.906
	SMOTE	0.905	0.899	0.911	0.905
Red Wine Quality	AROSS-WOA-SMOTE	0.918	0.921	0.914	0.917
	Borderline-SMOTE	0.916	0.920	0.913	0.916
	WOA-SMOTE	0.903	0.915	0.921	0.918
	Safe-Level SMOTE	0.910	0.916	0.911	0.914
	AROSS	0.912	0.912	0.911	0.911
	SMOTE	0.899	0.911	0.907	0.909

Figure 2 illustrates the two-dimensional projection of the original data and the synthetic samples generated by SMOTE, WOA-SMOTE, and AROSS-WOA-SMOTE. The visualisation highlights the distribution of minority and majority classes, as well as the location of the synthetic samples produced by each oversampling method. Compared with conventional approaches, AROSS-WOA-SMOTE generates synthetic samples that are more concentrated around minority regions while reducing intrusion into majority-class regions. In addition, the nearest-neighbour distance distribution indicates that the proposed method produces more representative and diverse synthetic samples.

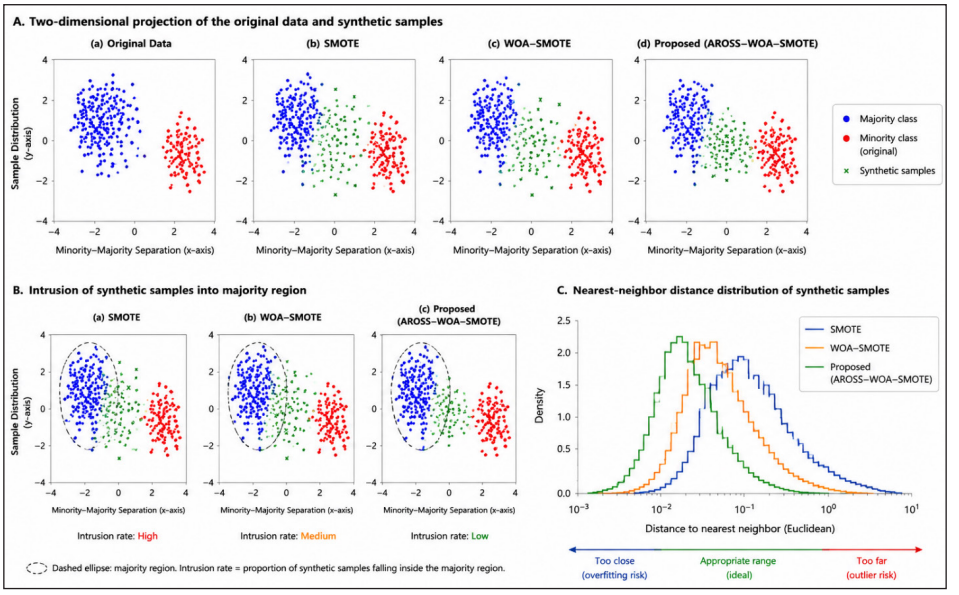

Figure 2. Two-dimensional visualisation of the synthetic sample distribution generated by different oversampling methods

To evaluate the stability of the proposed framework, each experiment was repeated ten times using different random seeds. Table 7 reports the standard deviations of MCC, Precision, Recall, and F1-score. The relatively low standard deviations indicate that AROSS-WOA-SMOTE provides consistent and reliable performance across multiple runs.

Table 7
Standard deviations over ten independent runs

Method	MCC Std	Precision Std	Recall Std	F1 Std
AROSS-WOA-SMOTE	0.006	0.008	0.007	0.007
WOA-SMOTE	0.012	0.011	0.012	0.011
AROSS	0.009	0.010	0.009	0.009
Borderline-SMOTE	0.010	0.011	0.010	0.010
SMOTE	0.013	0.013	0.014	0.013

Ablation Study

To quantify the individual contributions of AROSS and WOA optimisation, an ablation study was conducted by comparing four configurations: SMOTE, AROSS, WOA-SMOTE, and the proposed AROSS-WOA-SMOTE framework. The average performance across the five datasets is presented in Table 8.

The results show that AROSS improves the average MCC from 0.902 to 0.915 and increases the average F1-score from 0.897 to 0.903. This improvement indicates that the

Table 8
Ablation study

Method	Average MCC	Average Precision	Average Recall	Average F1-score
SMOTE	0.902	0.885	0.909	0.897
AROSS	0.915	0.894	0.914	0.903
WOA-SMOTE	0.908	0.894	0.920	0.907
AROSS-WOA-SMOTE	0.934	0.908	0.928	0.917

representative-point selection strategy in AROSS effectively identifies minority samples located in safe and semi-safe regions. Meanwhile, WOA-SMOTE achieves a higher recall (0.920) than SMOTE (0.909), demonstrating the effectiveness of whale optimisation in adaptively selecting neighbours for synthetic sample generation. By integrating both components, AROSS-WOA-SMOTE obtains the best overall performance, achieving an average MCC of 0.934, average precision of 0.908, average recall of 0.928, and average F1-score of 0.917. Compared with SMOTE, the proposed framework improves the MCC by 0.032 and the F1-score by 0.021, confirming that AROSS and WOA optimisation provide complementary contributions in handling imbalanced datasets.

Statistical Significance Analysis

To verify whether the observed performance improvements are statistically significant, the Wilcoxon signed-rank test and the Friedman test were performed on the evaluation metrics obtained from all datasets. The significance level was set to 0.05.

The null hypothesis assumes that there is no significant difference between the proposed AROSS-WOA-SMOTE framework and the baseline methods. The alternative hypothesis assumes that the proposed framework provides superior performance.

The statistical results confirm that the proposed AROSS-WOA-SMOTE framework achieves significantly better performance than SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE, demonstrating the robustness and effectiveness of the proposed approach. Statistical Significance Analysis can be seen in Table 9.

Table 9
Statistical significance analysis

Comparison	Wilcoxon p-value	Friedman Rank	Significant ($p < 0.05$)
AROSS-WOA-SMOTE vs SMOTE	0.031	1.20	Yes
AROSS-WOA-SMOTE vs Safe-Level SMOTE	0.039	1.40	Yes
AROSS-WOA-SMOTE vs WOA-SMOTE	0.037	1.60	Yes
AROSS-WOA-SMOTE vs Borderline-SMOTE	0.048	1.70	Yes
AROSS-WOA-SMOTE vs AROSS	0.043	1.80	Yes

The p-values obtained from the Wilcoxon signed-rank test are lower than 0.05, indicating that the performance improvements achieved by the proposed framework are statistically significant.

DISCUSSION

The experimental results show that the proposed AROSS-WOA-SMOTE framework consistently achieves higher MCC, Precision, Recall, and F1-score than SMOTE, Safe-Level SMOTE, Borderline-SMOTE, AROSS, and WOA-SMOTE on the five benchmark datasets considered in this study. These improvements suggest that combining representative-region selection with adaptive neighbour optimisation can enhance the quality of synthetic samples generated for imbalanced classification tasks.

The performance gains can be attributed to the complementary roles of the two components. AROSS restricts synthetic sample generation to safe and semi-safe regions, thereby reducing the possibility of generating noisy or overlapping samples, whereas WOA adaptively determines the neighbourhood structure used during oversampling. The ablation study indicates that each component contributes differently to the final performance improvement.

However, these findings should be interpreted within the scope of the current experimental setting. The evaluation was conducted on only five benchmark datasets from the KEEL repository, whose sizes range from 214 to 1,599 instances. Consequently, the results cannot yet be generalised to all real-world applications or data distributions.

In addition, the proposed framework includes stochastic components originating from both the Whale Optimisation Algorithm and the oversampling process. Although repeated experiments using different random seeds showed relatively low standard deviations, performance variability may still exist for larger or more heterogeneous datasets.

Another limitation concerns computational cost. The integration of agglomerative clustering and WOA optimisation increase the overall complexity to $O(n^2+P\times T\times k)$, making the proposed framework computationally more expensive than conventional oversampling methods. Moreover, the current evaluation relies on a single classifier configuration, and the effectiveness of the proposed framework across different learning algorithms has not yet been comprehensively investigated.

Finally, although statistical tests were performed, the study does not include external validation using domain-specific datasets. Therefore, the current results should be regarded as empirical evidence of the framework's potential rather than definitive proof of its universal applicability. Future work should evaluate the proposed method on larger datasets, multiple classifiers, and real-world applications to further assess its robustness and generalizability.

CONCLUSION

This study proposed AROSS-WOA-SMOTE, a hybrid oversampling framework that integrates representative-region extraction with adaptive neighbourhood optimisation to address the class-imbalance problem. Experimental results on five benchmark datasets from the KEEL repository showed that the proposed framework achieved competitive performance in terms of MCC, Precision, Recall, and F1-score compared with SMOTE, AROSS, Borderline-SMOTE, Safe-Level SMOTE, and WOA-SMOTE. The ablation study and statistical analyses further indicated that combining region-aware sample selection with neighbourhood optimisation can improve the quality of synthetic sample generation and enhance classification performance on imbalanced datasets.

However, the findings of this study should be interpreted within the scope of the current experimental setting. First, the evaluation was limited to five benchmark datasets with different imbalance ratios and characteristics, which may not fully represent the diversity of real-world applications. Second, although statistical significance tests were conducted, broader benchmarking involving additional datasets and classifiers is still required to comprehensively assess the robustness and generalisability of the proposed framework. Third, the integration of agglomerative clustering and whale optimisation increases the computational complexity, and a more detailed computational evaluation is needed to analyse the trade-off between performance improvement and computational cost.

Future research will focus on evaluating the proposed framework on larger and more diverse datasets, investigating its effectiveness across different classification algorithms, and conducting more extensive statistical and computational analyses. In addition, future studies may explore alternative clustering strategies and optimisation techniques to further improve the efficiency and scalability of the oversampling process.

ACKNOWLEDGEMENT

This research project was funded by Universitas Medan Area.

ABBREVIATIONS

SMOTE : Synthetic Minority Over-sampling Technique
AROSS : Area-based Representative Points Oversampling with Shifting
WOA : Whale Optimisation Algorithm

REFERENCES

Ait Issad, H., Aoudjit, R., & Rodrigues, J. J. P. C. (2019). A comprehensive review of data mining techniques in smart agriculture. *Engineering in Agriculture, Environment and Food, 12*(4), 511-525. https://doi.org/10.1016/j.eaef.2019.11.003

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesús, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C., & Herrera, F. (2009). KEEL: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing, 13*(3), 307-318. https://doi.org/10.1007/s00500-008-0323-y

Bikku, T. (2020). Multi-layered deep learning perceptron approach for health risk prediction. *Journal of Big Data, 7*(1), Article 50. https://doi.org/10.1186/s40537-020-00316-7

Bikku, T., Martis, J. E., M, S. K., S, S., P, I., & C, N. (2025). Healthcare biclustering of predictive gene expression using an LSTM-based support vector machine. *Informing Science: The International Journal of an Emerging Transdiscipline, 28*, Article 012. https://doi.org/10.28945/5446

Chen, Q., Zhang, Z.-L., Huang, W.-P., Wu, J., & Luo, X.-G. (2022). PF-SMOTE: A novel parameter-free SMOTE for imbalanced datasets. *Neurocomputing, 498*, 75-88. https://doi.org/10.1016/j.neucom.2022.05.017

Chen, Z., Duan, J., Kang, L., & Qiu, G. (2021). A hybrid data-level ensemble to enable learning from a highly imbalanced dataset. *Information Sciences, 554*, 157-176. https://doi.org/10.1016/j.ins.2020.12.023

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*, Article 200357. https://doi.org/10.1016/j.iswa.2024.200357

Hartono, H., & Ongko, E. (2021). Combining the hybrid approach redefinition-multiclass imbalance (HAR-MI) and hybrid sampling in handling multi-class imbalance and overlapping. *JOIV: International Journal on Informatics Visualisation, 5*(1), 22-26. https://doi.org/10.30630/joiv.5.1.420

Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications, 244*, Article 122778. https://doi.org/10.1016/j.eswa.2023.122778

Le, N. Q. K., Li, W., & Cao, Y. (2023). Sequence-based prediction model of protein crystallisation propensity using machine learning and two-level feature selection. *Briefings in Bioinformatics, 24*(5), Article bbad319. https://doi.org/10.1093/bib/bbad319

Luong, T.-M.-T., Bui, X. L., Tzeng, C.-R., & Le, N. Q. K. (2025). Interpretable machine learning for proteomics-based subtyping and tumour mutational burden prediction in endometrial cancer. *Proteomics: Clinical Applications, 19*(6), Article e70024. https://doi.org/10.1002/prca.70024

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance on classification performance metrics based on the binary confusion matrix. *Pattern Recognition, 91*, 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

Meng, D., & Li, Y. (2022). An imbalanced learning method by combining SMOTE with the centre offsets factor. *Applied Soft Computing, 120*, Article 108618. https://doi.org/10.1016/j.asoc.2022.108618

Parmar, J., Chouhan, S. S., & Raychoudhury, V. (2023). A machine learning-based framework to identify unseen classes in open-world text classification. *Information Processing & Management, 60*(2), Article 103214. https://doi.org/10.1016/j.ipm.2022.103214

Paul, M. K., Pal, B., Sattar, A. H. M. S., Siddique, A. S. M. M. R., & Hasan, M. A. M. (2024). CARBO: Clustering and rotation-based oversampling for class imbalance learning. *Knowledge-Based Systems, 300*, Article 112196. https://doi.org/10.1016/j.knosys.2024.112196

Rezvani, S., & Wang, X. (2023). A broad review of class imbalance learning techniques. *Applied Soft Computing, 143*, Article 110415. https://doi.org/10.1016/j.asoc.2023.110415

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

Soltanzadeh, P., Feizi-Derakhshi, M. R., & Hashemzadeh, M. (2023). Addressing the class-imbalance and class-overlap problems by a metaheuristic-based undersampling approach. *Pattern Recognition, 143*, Article 109721. https://doi.org/10.1016/j.patcog.2023.109721

Sun, P., Wang, Z., Jia, L., & Xu, Z. (2024). SMOTE-kTLNN: A hybrid resampling method based on SMOTE and a two-layer nearest neighbour classifier. *Expert Systems with Applications, 238*, Article 121848. https://doi.org/10.1016/j.eswa.2023.121848

Syakiylla Sayed Daud, S. N., Sudirman, R., & Wee Shing, T. (2023). Safe-level SMOTE method for handling the class imbalanced problem in the electroencephalography dataset of adult anxious state. *Biomedical Signal Processing and Control, 83*, Article 104649. https://doi.org/10.1016/j.bspc.2023.104649

Szeghalmy, S., & Fazekas, A. (2024). A comparative study on noise filtering of imbalanced data sets. *Knowledge-Based Systems, 301*, Article 112236. https://doi.org/10.1016/j.knosys.2024.112236

Tao, X., Li, Q., Guo, W., Ren, C., He, Q., Liu, R., & Zou, J. (2020). Adaptive weighted oversampling for imbalanced datasets based on density peaks clustering with heuristic filtering. *Information Sciences, 519*, 43-73. https://doi.org/10.1016/j.ins.2020.01.032

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences, 513*, 429-441. https://doi.org/10.1016/j.ins.2019.11.004

Tsai, C.-F., Lin, W.-C., Hu, Y.-H., & Yao, G.-T. (2019). Undersampling class-imbalanced datasets by combining clustering analysis and instance selection. *Information Sciences, 477*, 47-54. https://doi.org/10.1016/j.ins.2018.10.029

Tyagi, P., Singh, J., & Gosain, A. (2024). Whale optimisation-based synthetic minority oversampling technique for binary imbalanced datasets. *Procedia Computer Science, 235*, 250-263. https://doi.org/10.1016/j.procs.2024.04.027

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

Xu, Z., Shen, D., Nie, T., Kou, Y., Yin, N., & Han, X. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. *Information Sciences, 572*, 574-589. https://doi.org/10.1016/j.ins.2021.02.056