International Journal of Approximate Reasoning 187 (2025) 109538

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Distribution assessment-based multiple over-sampling with evidence fusion for imbalanced data classification

Hongpeng Tian [a,b], Zuowei Zhang [a,*], Zhunga Liu [a], Jingwei Zuo [c], Caixing Yang [a]

[a] School of Automation, Northwestern Polytechnical University, 710072 Xi'an, China
[b] School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, 450001, China
[c] Technology Innovation Institute, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Distribution assessment
Over-sampling
Imbalanced data
Neighbors
Classification

ABSTRACT

Over-sampling methods concentrate on creating balanced samples and have proven successful in classifying imbalanced data. However, current over-sampling methods fail to consider the uncertainty of produced samples, potentially altering the data distribution and impacting the classification process. To address this issue, we propose a distribution assessment-based multiple over-sampling (DAMO) method for classifying imbalanced data. We first introduce a multiple over-sampling method based on distribution assessment to create different forms of synthetic samples. The core is quantifying the inconsistency of data distribution before and after sampling as a constraint to guide multiple over-sampling, thereby minimizing the data shift and characterizing the uncertainty of produced samples. Then, we quantify the local reliability of the classification results and select several imprecise samples with low local reliability that are indistinguishable between classes. Neighbors serve as additional complementary information to calibrate the results of imprecise samples, thereby reducing the likelihood of misclassification. The calibrated results are combined by the discounting Dempster-Shafer fusion rule to make a final decision. DAMO's efficiency has been demonstrated through comparisons with related methods on various real imbalanced datasets.

1. Introduction

The term "imbalanced learning problems" denotes the uneven allocation of data samples across various classes in the dataset [1–3]. This phenomenon frequently occurs in various applications, including healthcare [2], fault diagnosis [4], and intrusion detection [5], among others. Typically, samples from the minority class contain crucial data, and accurately categorizing them is often essential [6,7]. However, models developed with a focus on precision in their design often show subpar performance in identifying minority classes.

Recently, numerous methods [8,7,9,10] have emerged for classifying unbalanced data sets. The widely favored and optimistic data sampling methods focus on pre-processing the input data to ensure class balance. The benefits of these methods lie in their practicality and efficiency as they function directly on the data. Numerous empirical studies have demonstrated their universal applicability in many cases. Methods for sampling data can be broadly categorized into two types: under-sampling and over-sampling methods. Under-sampling methods, as referenced in various studies [11,6,12], aim to reduce the sample count in the majority class to maintain dataset equilibrium, yet they might face the hazard of overlooking significant majority samples. By contrast, over-sampling methods produce

* Corresponding author.
E-mail address: zhangzuowei@nwpu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.ijar.2025.109538
Received 1 December 2024; Received in revised form 24 June 2025; Accepted 30 July 2025

Available online 6 August 2025

artificial data aimed at augmenting the number of minority samples, thereby preserving crucial minority data. These over-sampling methods can be broadly classified into three groups. 1) Neighbor-based over-sampling methods [13–16] create minority samples based on neighboring samples, thus expanding areas of the minority class and reducing overfitting. 2) Cluster-based over-sampling methods [17–19] produce varied and superior samples across various clusters. 3) Semi-supervised over-sampling methods [20–22] extensively use plentiful unlabeled data to create fresh samples to equilibrate categories. While all three groups have demonstrated effectiveness in addressing class imbalance, neighbor-based over-sampling methods remain widely adopted due to their simplicity, interpretability, and strong theoretical foundation in preserving local data structures. Given these advantages, we focus on neighbor-based over-sampling methods in this study, particularly for their effectiveness in enhancing model robustness and generalization. However, neighbor-based over-sampling methods continue to face certain constraints in both the over-sampling and classification procedures.

- During the over-sampling process, these methods yield unique synthetic samples for the minority class, which fail to accurately characterize the uncertainty in the produced samples. This is due to the unavoidable discrepancy between the produced sample and the real sample.
- During the classification process after sampling, an inaccurate generation could lead to alterations in the initial data distribution, bringing imprecise information to the classification process. In this case, there may be several imprecise samples that are indistinguishable from each other across different classes.

Consequently, it is essential to explore a new over-sampling method that can characterize the uncertainty of produced samples and imprecise information during the classification process.

Interestingly, Dempster-Shafer theory (DST) [23–25], also known as the theory of belief functions, is first introduced in [23]. DST has the advantage of reasoning with uncertain and imprecise information [26,27], which makes it widely used in many fields, including data classification [28–31,27] and decision making fusion [32,26]. Therefore, we propose a distribution assessment-based multiple over-sampling (DAMO) method based on DST to address the limitations of existing over-sampling methods. The contributions of DAMO can be encapsulated in three key dimensions.

1. We defined a multiple over-sampling strategy based on distribution assessment to generate multiple versions of synthetic samples. Doing this can characterize the uncertainty of produced samples and somewhat minimize alterations in data distribution in the over-sampling process.
2. We designed an evidence fusion strategy with neighbor calibration to fuse multiple classification results and make a final decision. In this way, we can utilize the complementary information to improve the classification performance and calibrate the results for imprecise samples, challenging to classify correctly.
3. We conducted numerous experiments with several imbalanced datasets from the KEEL repository. The experimental results demonstrate the effectiveness and superiority of the proposed method in classifying imbalanced datasets compared to other methods.

The organization of this document is as follows. Section 2 provides a concise overview of related works. Section 3 provides the necessary knowledge for the proposed method. Section 4 provides an in-depth explanation of the proposed DAMO method. Section 5 evaluates the efficacy of the DAMO method and compares it with various other standard methods. In summary, we conclude our research in Section 6.

2. Related works

In this section, we review some related neighbor-based over-sampling methods, which are central to our research interest due to their effectiveness in enhancing model robustness and generalization.

Neighbor-based over-sampling methods generate new samples by interpolating between minority samples and their nearest neighbors. The synthetic minority over-sampling technique (SMOTE) [15], a fundamental method for data synthesis-based over-sampling, involves searching for adjacent samples within each minority group and creating artificial samples between these samples and their neighbors. Recently, various modified forms of SMOTE have been introduced to overcome its limitations and broaden its range of applications. A new minority over-sampling method, named Borderline-SMOTE, was proposed in [16], where only the minority samples near the borderline were over-sampled. A Safe-Level-SMOTE method [33] produced synthesized samples of minority groups with higher safe levels calculated by adjacent minority groups. In [34], the introduction of a certainty guided minority over-sampling (CGMOS) method was observed, incorporating minority samples based on the variations in certainty within the dataset. In [35], an innovative data enhancement method named H-SMOTE was introduced, merging neighboring elements and the Manhattan distance to create fresh samples centered on the minority class. A hybrid resampling method named SMOTE-kTLNN [36] merged SMOTE for balanced data augmentation with a two-layer nearest-neighbor classifier and utilized an Iterative-Partitioning Filter to eliminate noisy samples by majority voting. Recently, there is a growing interest in combining neighbor-based over-sampling with deep learning, aiming to better address the challenges of imbalanced data. The literature [37] proposes an over-sampling method that integrates deep learning techniques with the traditional SMOTE algorithm, named the deep synthetic minority over-sampling technique (DeepSMOTE). The literature [38] demonstrates that combining CNN with SMOTE data balancing achieves the highest classification accuracy in hyperspectral image classification. In [39], an improved SMOTE method combined with an LSTM-AdaBoost

ensemble is introduced to enhance credit risk evaluation on imbalanced datasets. In [40], a method combining SMOTE over-sampling, genetic algorithm-based feature selection, and 1D-DCNN classification, is developed to improve deep learning performance on high-dimensional imbalanced datasets.

Although current over-sampling methods effectively mitigate class imbalance and improve model performance, they generate deterministic samples that fail to accurately capture data uncertainty and can distort the original distribution during classification. These limitations highlight the need for more robust methods, motivating our development of the proposed DAMO method.

3. Preliminaries

In this section, we introduce the concepts of density peaks and Dempster–Shafer theory, which provide the necessary preliminaries for the proposed method.

3.1. Density peaks

The concept of density peaks is introduced within the context of density peaks clustering (DPC) [41] method. In DPC, each data point $\mathbf{y}_i$ is characterized by two measures: its local density ρ_i, which reflects how many neighboring samples are close to it, and its relative distance δ_i, indicating how far it is from the nearest sample with a higher local density.

The local density ρ_i is defined as:

$$\rho_i = \sum_{j \neq i} \chi(d_{ij} - d_c), \tag{1}$$

where d_c is a predefined cutoff distance, and the function $\chi(\cdot)$ acts as an indicator: it returns 1 if its argument is less than zero (i.e., if $d_{ij} < d_c$), and 0 otherwise. Therefore, ρ_i counts the number of samples within a distance d_c from $\mathbf{y}_i$, serving as a measure of its local neighborhood density.

The relative distance δ_i represents the minimum distance between $\mathbf{y}_i$ and any other sample $\mathbf{y}_j$ that has a higher local density:

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}). \tag{2}$$

A sample is considered a density peak if it simultaneously exhibits both a significantly high ρ_i and a relatively large δ_i. These samples are typically good candidates for density peaks due to their central location within dense regions and isolation from other clusters. To facilitate interpretation and ranking, γ_i is defined, such that $\gamma_i = \rho_i \cdot \delta_i$, combines both criteria into a single score. Samples with the highest γ_i values are typically selected as density peaks. These representative samples tend to reside in densely populated regions of the data space and are sufficiently far from other clusters, making them ideal candidates for capturing the underlying structure of the dataset. Therefore, density peaks are considered to be highly informative and structurally significant samples that more accurately reflect the true distribution and intrinsic organization of the data.

3.2. Dempster–Shafer theory

Dempster–Shafer theory (DST) [23,24,42], alternatively referred to as the theory of belief functions, was initially proposed in [23]. It offers a robust framework for dealing with uncertainty and has found broad applications across various domains such as classification [28–31,27], data clustering [43–45], and multi-source decision fusion [32,26,46].

Let $\Omega = \{\omega_1, \omega_2, ..., \omega_C\}$ denote the frame of discernment, which contains C mutually exclusive and collectively exhaustive hypotheses or states. DST extends this set to its power-set 2^Ω, defined as $2^\Omega = \{\emptyset, \omega_1, \ldots, \omega_C, \{\omega_1, \omega_2\}, \ldots, \Omega\}$.

A basic belief assignment (BBA), often denoted as $m(A)$, is a function that assigns a degree of belief to each subset A of 2^Ω. It satisfies the following condition:

$$\sum_{A \in 2^\Omega} m(A) = 1, \tag{3}$$

where $m(A) \in [0, 1]$. The value $m(A)$ represents the strength of evidence supporting the proposition that the true state lies within the subset A. Notably, if A consists of more than one singleton element, it can represent uncertain or partial knowledge and is sometimes interpreted as a meta-cluster.

When combining two independent BBAs, say $m_1(A)$ and $m_2(B)$, where $A, B \in 2^\Omega$, the DS rule of combination is applied:

$$m(C) = \frac{1}{1-\kappa} \sum_{A \cap B = C} m_1(A) m_2(B), \tag{4}$$

where the normalization factor κ is given by:

$$\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B). \tag{5}$$

Here, A and B are non-empty subsets of 2^Ω. The parameter κ quantifies the level of conflict between the two pieces of evidence; a higher κ implies greater inconsistency between them. It should be noted that the DS rule performs best when the conflict between the sources is relatively low.

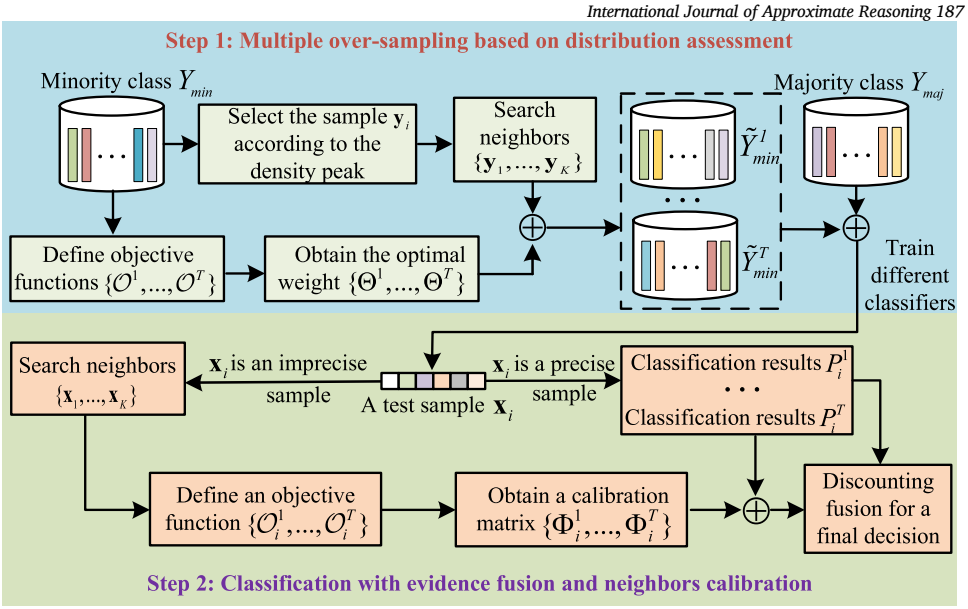

Fig. 1. Workflow of the proposed method. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4. The proposed method

This part introduces the proposed DAMO method for classifying imbalanced data. It mainly consists of two steps: 1) multiple over-sampling based on distribution assessment; 2) classification with evidence fusion and neighbors calibration, which will be presented in subsections 4.1–4.2. Fig. 1 shows the complete structure of DAMO, designed to clearly demonstrate its function. Assume that a test set $X = \{\mathbf{x}_1,...,\mathbf{x}_N\}$ is classified within the classification framework $\Omega = \{\omega_{min}, \omega_{maj}\}$, based on a training set $Y = \{\mathbf{y}_1,...,\mathbf{y}_M\}$ in H distinct attribute spaces. The symbols Y_{min} and Y_{maj} denote the minority and majority classes in the training dataset, respectively.

4.1. Multiple over-sampling based on distribution assessment

In this subsection, we introduce a distribution assessment-based over-sampling methods to generate samples. By doing this, we can obtain several balanced datasets to train basic classifiers and classify test samples.

Initially, it is essential to choose several samples as benchmarks for generating new ones, ensuring that these samples are varied and accurately reflect the data distribution within the minority class. Specifically, the density peaks introduced in the density peaks clustering (DPC) [41] method aim to identify clustering centers capable of modeling and defining the data distribution in various forms and dimensions. Consequently, we use density peaks to select a set of typical benchmark samples, and this process is introduced in subsection 3.1 in detail. Every benchmark sample conducts searches among K neighbors to create artificial samples, concluding this procedure until the count (*i.e.*, $M'K$) represents the count (*i.e.*, $\left|Y_{maj}\right| - \left|Y_{min}\right|$) of synthetic samples awaiting expansion. The symbol M' denotes the count of benchmark samples required to locate neighboring samples for the creation of synthetic samples.

Assume that a benchmark samples $\mathbf{y}_i$, it is possible to create synthetic samples linking $\mathbf{y}_i$ with adjacent $\mathbf{y}_k$ ($k = 1,...,K$), similar to SMOTE [15]. However, the samples produced are arbitrary and inaccurate, leading to a degree of unpredictability and uncertainty in this procedure. The uncertainty cannot be characterized if we only generate a specific synthetic sample between $\mathbf{y}_i$ and a neighbor $\mathbf{y}_k$. Thus, we introduce a multiple over-sampling strategy for $\mathbf{y}_i$ to generate multiple synthetic samples, characterizing the uncertainty of the synthetic samples. In addition, inaccurate generations of synthetic samples could alter the initial data distribution of the minority class, which is necessary to be considered and addressed in the over-sampling process. Consequently, the t-th new synthetic sample $\mathbf{y}^t_{i\to k}$ ($k = 1,...,K; t = 1,...,T$) is generated between $\mathbf{y}_i$ and its k-th neighbor $\mathbf{y}_k$, denoted as:

$$\mathbf{y}^t_{i\to k} = \mathbf{y}_i + \theta^t_k \alpha^t_k (\mathbf{y}_k - \mathbf{y}_i) \tag{6}$$

where α^t_k is a random value and falls within the range $[0, 1]$, and θ^t_k is a weight parameter used to reduce the change of data distribution of the minority class. Through this process, certain synthetic samples are acquired, and the modified minority class post-over-sampling is labeled as $\tilde{Y}^t_{min}$. The weight vector $\Theta^t = [\theta^t_1,...,\theta^t_K]$ is established to influence neighboring elements during the creation of samples. Optimization of the vector Θ^t is achieved by reducing the disparity in distribution between the modified minority class $\tilde{Y}^t_{min}$ and the initial minority class Y_{min}. In this case, the maximum mean discrepancy (MMD) [47] is utilized to measure the extent of data distribution inconsistency, involving the object function $\mathcal{O}^t$, indicated as:

$$\begin{aligned}\mathcal{O}^t &= \mathbb{D}_{MMD}(\tilde{Y}^t_{min} \parallel Y_{min}) \\ &= \left\| \frac{1}{\left|\tilde{Y}^t_{min}\right|} \sum_{\mathbf{y}_i \in \tilde{Y}^t_{min}} \phi(\mathbf{y}_i) - \frac{1}{|Y_{min}|} \sum_{\mathbf{y}_j \in Y_{min}} \phi(\mathbf{y}_j) \right\|\end{aligned} \tag{7}$$

where the symbol $|.|$ represents the cardinality, and $\phi(.)$ denotes a mapping function used to align $\tilde{Y}^t_{min}$ and Y_{min} in a singular space, thus determining their variance. Here, one can determine the optimal weights $\hat{\Theta}^t$ through the application of this formula:

$$\begin{aligned}\hat{\Theta}^t &= \arg\min_{\Theta^t} \mathcal{O}^t \\ &= \arg\min_{\Theta^t} \left\| \frac{1}{|\tilde{Y}^t_{min}|} \sum_{\mathbf{y}_i \in \tilde{Y}^t_{min}} \phi(\mathbf{y}_i) - \frac{1}{|Y_{min}|} \sum_{\mathbf{y}_j \in Y_{min}} \phi(\mathbf{y}_j) \right\|.\end{aligned} \tag{8}$$

Here, utilizing the interior-point method [48] is applicable in this context to address a standard constrained nonlinear least-squares issue.

After we obtain the optimal weights $\hat{\Theta}^t$, synthetic samples are generated according to Eq. (6). Then, the t-th modified minority class $\tilde{Y}^t_{min}$ is combined with the majority class Y_{maj}, and we obtain the t-th modified balanced training set $\tilde{Y}^t$ $(t = 1, ..., T)$.

Following over-sampling, these training sets are utilized to train the basic classifier to partition test samples. With a given test sample $\mathbf{x}_i$, T classification outcomes P^t_i $(t = 1, ..., T)$ are achievable. Decision making for $\mathbf{x}_i$ can be based on the DS fusion rule. However, different classifiers trained by different training sets have various reliability for classifying each test sample. Despite reducing the alteration in data distribution throughout the over-sampling phase, a slight irregularity in data distribution persists both before and after over-sampling due to sampling uncertainty. In these circumstances, several samples might present challenges in accurate classification, and the following subsection will assist in adjusting the outcomes of these imprecise samples.

4.2. Classification with evidence fusion and neighbors calibration

In this subsection, we fuse different classification outcomes with various reliability to make the final decision. Then, we extract several imprecise samples and use their neighbors as supplementary data to calibrate the classification outcomes.

Different classifiers trained by different balanced training sets are employed to classify test samples. For the test sample $\mathbf{x}_i$, we can obtain T classification results $P^t_i = [p^t_i(\omega_{min}), p^t_i(\omega_{maj})]$, $t = 1, ..., T$. It is worth noting that different classifiers have various abilities in grouping $\mathbf{x}_i$, and the local reliability of T classification results are various. Here, we employ the neighbors $\mathbf{y}_1, ..., \mathbf{y}_K$ of the test sample $\mathbf{x}_i$ to evaluate the local reliability of its t-th classification result P^t_i. If a classifier performs well on $\mathbf{y}_1, ..., \mathbf{y}_K$, it is likely to perform well on $\mathbf{x}_i$. This is because $\mathbf{x}_i$ shares a comparable data configuration and distribution with these neighbors, tending to exhibit comparable classification behaviors. Therefore, enhanced efficacy of the classifier in grouping $\mathbf{y}_1, ..., \mathbf{y}_K$ correlates with increased reliability in classifying $\mathbf{x}_i$. Utilizing the preceding analysis, we calculate the reliability β^t_i of the t-th classification result P^t_i of $\mathbf{x}_i$, as indicated:

$$\beta^t_i = \exp(-\sum_{k=1}^{K} \xi_k \left\| P^t_k - L^t_k \right\|) \tag{9}$$

with

$$\xi_k = \frac{\exp(-d_{ik})}{\sum_{\chi=1}^{K} \exp(-d_{i\chi})} \tag{10}$$

where the symbol ξ_k denotes the weight assigned to the neighbor $\mathbf{y}_k$. It's evident that as the distance d_{ik} between the neighbor $\mathbf{y}_k$ and $\mathbf{x}_i$ decreases, so does the weight ξ_k. $P^t_k = [p^t_k(\omega_{min}), p^t_k(\omega_{maj})]$ represents the classification result of $\mathbf{y}_k$. $L^t_k = [l^t_k(\omega_{min}), l^t_k(\omega_{maj})]$, a binary vector, establishes the ground truth of $\mathbf{y}_k$. $l^t_k(\omega_{min}) = 1$ in cases where ω_{min} signifies the true class of $\mathbf{y}_k$. If not, the magnitude of $l^t_k(\omega_{min})$ becomes zero, and $l^t_k(\omega_{maj})$ is equal to 1 in this case.

$$\beta_{i,\max} = \max\{\beta^1_i, ..., \beta^T_i\}. \tag{11}$$

With the rise in $\beta_{i,\max}$'s value, the likelihood of accurately categorizing $\mathbf{x}_i$ escalates. Here, we use the K-means clustering [49] method to partition the different $\beta_{i,\max}$ $(i = 1, ..., N)$ into two classes. According to the cluster centers, one class has higher β_i values, indicating that for $\mathbf{x}_i$, at least one classification result has high reliability. In this case, these samples are considered precise samples with a higher $\beta_{i,\max}$, and they have a high reliability of being correctly classified.

- **Classification of precise samples**

Then, the classification results of these precise samples are combined by the discounting DS fusion rule [24] to make decisions. Assume that the sample $\mathbf{x}_i$ is a precise sample, and the discounted masses of belief are obtained by:

$$\begin{cases} m_i^t(A) = \hat{\beta}_i^t p_i^t(A), A \subset \Omega; \\ m_i^t(\Omega) = 1 - \hat{\beta}_i^t + \hat{\beta}_i^t p_i^t(\Omega) \end{cases} \tag{12}$$

with

$$\hat{\beta}_i^t = \frac{\beta_i^t}{\sum_{\tau=1}^{T} \beta_i^\tau} \tag{13}$$

where β_i^t represents the local reliability of the t-th classification result P_i^t of $\mathbf{x}_i$. $\hat{\beta}_i^t$ $(t = 1, ..., T)$ represents the discounting factors (weights) used to balance the contributions of different classifiers. In the context of classifiers based on DST, such as EK-NN [28], $p_i^t(\Omega)$ represents the probability that $\mathbf{x}_i$ belongs to the completely unknown class Ω. However, for traditional probabilistic classifiers like Support Vector Machin (SVM) [50] and Multi-Layer Perceptron (MLP) [51], which operate under a classical probability framework, $p_i^t(\Omega)$ is not defined and thus set to 0. A classifier that achieves a higher β_i^t is assigned a higher discounting factor $\hat{\beta}_i^t$, reflecting its greater reliability for classifying $\mathbf{x}_i$ correctly. In this case, the less information of P_i^t obtained by this classifier is discounted, the greater its contribution to the fusion process and decision-making. Conversely, a classifier with a lower $\hat{\beta}_i^t$ receives a lower weight, reducing its influence on the final decision. By dynamically adjusting the weights based on local reliability, we ensure that more reliable classifiers contribute more significantly to the fused decision, thereby optimizing the overall performance.

Through this process, information that has been devalued can be submitted to the total unknown class (Ω), thereby reducing the degree of discord among pieces of evidence. This permits the application of the DS fusion rule introduced in Eqs. (4) and (5) for combining discounted results and making a decision.

- **Classification of imprecise samples**

In contrast, samples in the other class have lower $\beta_{i,\max}$, which means they have a higher risk of misclassification and should be carefully considered as imprecise samples waiting for further partitioning.

Here, assume that the sample $\mathbf{x}_i$ is an imprecise sample. We calibrate the results of $\mathbf{x}_i$ according to its neighbors, as they have similar spatial structures. The calibration matrix Φ_i^t for $\mathbf{x}_i$ is defined by:

$$\Phi_i^t = \begin{bmatrix} \phi_{min,min}^t & \phi_{min,maj}^t \\ \phi_{maj,min}^t & \phi_{maj,maj}^t \end{bmatrix} \tag{14}$$

where $\phi_{uv}^t \in [0, 1]$, such that $u, v \in \{\omega_{min}, \omega_{maj}\}$. The symbol ϕ_{uv}^t denotes the conditional probability that the sample $\mathbf{x}_i$ is part of class ω_v, contingent on its classification to ω_u based on the classifier's outcome.

To derive the calibration matrix $\Phi_i^t = [\phi_{uv}^t]_{2\times 2}$, one can minimize the discrepancy between the prediction and the ground truths. Our design involves a global objective function $\mathcal{O}_i^t$, characterized by:

$$\mathcal{O}_i^t = \sum_{k=1}^{K} \xi_k \sqrt{\sum_{\omega_c \in \Omega} \left[\hat{p}_k^t(\omega_c) - l_k^t(\omega_c)\right]^2} \tag{15}$$

subject to

$$\hat{p}_k^t(\omega_c) = \begin{cases} p_k^t(\omega_{min}) \cdot \phi_{min,min}^t + p_k^t(\omega_{maj}) \cdot \phi_{maj,min}^t, & \omega_c = \omega_{min} \\ p_k^t(\omega_{min}) \cdot \phi_{min,maj}^t + p_k^t(\omega_{maj}) \cdot \phi_{maj,maj}^t, & \omega_c = \omega_{maj} \end{cases} \tag{16}$$

where $P_k^t = [p_k^t(\omega_{min}), p_k^t(\omega_{maj})]$ and $L_k^t = [l_k^t(\omega_{min}), l_k^t(\omega_{maj})]$ represent the classification result and ground truth of $\mathbf{y}_k$, respectively. K denotes the count of neighbors. $l_k^t(\omega_c) = 1$ indicates that the ground truth of $\mathbf{y}_k$ is ω_c. If not, the value of $l_k^t(\omega_c)$ equals zero. The symbol ξ_k denotes the weight assigned to the k-th neighbor. As the distance between the $\mathbf{x}_i$ and the neighbor becomes smaller, the neighbor plays a greater role in calibrating $\mathbf{x}_i$, resulting in a greater weight value in this case.

Subsequently, we can optimize the calibration matrix Φ_i^t by applying this formula:

$$\begin{aligned} \hat{\Phi}_i^t &= \arg\min_{\Phi_i^t} \mathcal{O}_i^t \\ &= \arg\min_{\Phi_i^t} \sum_{k=1}^{K} \xi_k \sqrt{\sum_{\omega_c \in \Omega} \left[\hat{p}_k^t(\omega_c) - l_k^t(\omega_c)\right]^2} \end{aligned} \tag{17}$$

where $\hat{\Phi}_i^t$ represents the optimal calibration matrix, such that $\hat{\Phi}_i^t = [\hat{\phi}_{uv}^t]_{2\times 2}$ and $u, v \in \{\omega_{min}, \omega_{maj}\}$. The interior-point method, similar to Eq. (14), is used here to solve a normal constrained nonlinear least-squares problem.

Table 1
Basic information of the benchmark datasets.

Data	#Size.	#Min.	#Maj.	#Attr.	#IR.
glass1	214	76	138	9	1.82
vehicle2	846	218	628	18	2.88
glass0123vs456	214	51	163	9	3.20
ecoli1	336	77	259	7	3.36
ecoli2	336	52	284	7	5.46
segment0	2308	329	1979	19	6.02
ecoli034vs5	200	20	180	7	9.00
ecoli0234vs5	202	20	182	7	9.10
vowel0	988	90	898	13	9.98
glass06vs5	108	9	99	9	11.00
page-blocks13vs4	472	28	444	10	15.86
poker8vs6	1477	17	1460	10	85.88

The calibrated classification result $\hat{P}_i^t = [\hat{p}_i^t(\omega_{min}), \hat{p}_i^t(\omega_{maj})]$ is defined by:

$$\begin{cases} \hat{p}_i^t(\omega_{min}) = p_i^t(\omega_{min}) \cdot \hat{\phi}_{min,min}^t + p_i^t(\omega_{maj}) \cdot \hat{\phi}_{maj,min}^t \\ \hat{p}_i^t(\omega_{maj}) = p_i^t(\omega_{min}) \cdot \hat{\phi}_{min,maj}^t + p_i^t(\omega_{maj}) \cdot \hat{\phi}_{maj,maj}^t \end{cases}. \tag{18}$$

Subsequently, a conclusive determination of $\mathbf{x}_i$ can be reached, guided by the calibration outcome. For $\mathbf{x}_i$, we can obtain T calibration matrices Φ_i^t $(t = 1, ..., T)$ thereby calibrating the classification result P_i^t accordingly. Then, we can obtain T calibrated results $\hat{P}_i^t$ $(t = 1, ..., T)$, and they are fused by the discounting fusion rule reported in Eqs. (4)−(5) to make final decisions.

5. Experiment applications

Here, we assess and scrutinize the efficacy of the proposed DAMO method compared to other methods. Section 5.1 reports the evaluation indices used in experiments. Section 5.2 presents an introduction to the benchmark datasets and the process of them in experiments. Subsequently, the comparison methods are detailed in Section 5.3. Then, Section 5.4 details the experimental results and analysis. Furthermore, Section 5.5 reports extensive discussions to examine and analyze the proposed DAMO method.

5.1. Evaluation indices

For evaluating different methods, performance indices [52] such as F-measure (FM), G-mean (GM), Area Under Curve (AUC), and Average Precision (AP), frequently used in classifying imbalanced data, are utilized in experiments. An increase in the values of these indices correlates with improved method performance.

These indices are chosen for their ability to capture complementary aspects of classifier performance:

- FM balances precision and recall, providing a harmonic mean that ensures the model does not bias toward the majority class in imbalanced scenarios.
- GM explicitly emphasizes minority-class performance by taking the geometric mean of class-wise accuracies, penalizing models that neglect underrepresented classes.
- AUC quantifies the overall discriminative power across all classification thresholds, where the ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR).
- AP directly measures the area under the Precision-Recall (PR) curve, which is critical for imbalanced datasets because the PR curve focuses on precision and recall, better reflecting the performance of minority classes.

Consequently, these indices collectively provide a comprehensive evaluation of the classification performance on imbalanced datasets. By combining these indices, we achieve a reliable and robust performance analysis that avoids over-reliance on any single measure, enhancing the reliability and validity of experiment results.

5.2. Benchmark datasets

Twelve real-world imbalanced datasets from the KEEL repository (http://www.keel.es/) are employed to evaluate and validate the effectiveness of different methods for classifying imbalanced data. Table 1 provides essential details for these datasets, including the total number of samples (#Size.), the number of samples in the majority class (#Maj.), the number of samples in the minority class (#Min), attributes (#Attr.), and the imbalance ratio ($\#IR.$). For subsequent experiments, each dataset is partitioned using a five-fold stratified cross-validation approach. The simulations yield statistical results that are the average of ten runs, minimizing randomness and ensuring more reliable outcomes.

Table 2
FM with standard deviation of different methods in classifying various datasets by SVM.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	53.06±4.87	53.93±3.99	52.22±4.76	50.90±5.06	52.94±3.91	50.24±4.09	53.60±1.58	16.38±2.98	**54.72±7.86**
vehicle2	73.95±2.37	74.20±1.34	74.09±2.38	74.12±2.48	72.98±2.72	73.47±2.48	60.98±0.62	57.40±0.25	**87.81±2.50**
glass0123vs456	82.68±5.60	79.48±4.00	81.86±4.25	83.01±3.63	82.87±5.24	80.43±4.35	77.86±0.17	84.53±0.37	**86.10±2.56**
ecoli1	74.88±2.40	74.60±1.87	74.30±1.87	74.06±2.48	74.72±1.69	67.09±1.94	75.94±0.59	**77.65±0.13**	76.45±4.20
ecoli2	68.44±3.83	62.16±3.63	68.04±4.28	68.46±3.77	68.73±3.12	56.67±5.64	**86.03±0.63**	82.53±0.62	81.75±4.86
segment0	96.56±1.26	78.55±1.32	96.58±1.23	96.60±1.25	95.72±1.40	96.29±1.08	48.42±3.72	41.02±3.18	**98.20±0.29**
ecoli034vs5	72.23±3.00	62.65±1.66	78.36±3.27	78.46±3.18	70.29±1.77	71.61±1.54	82.79±1.14	74.25±0.66	**85.20±3.24**
ecoli0234vs5	70.47±5.26	59.58±1.89	78.49±2.75	78.54±2.28	65.15±2.98	65.46±2.26	82.21±0.31	73.12±0.39	**82.40±4.08**
vowel0	68.93±1.74	76.67±1.57	68.98±1.72	68.16±1.80	67.89±2.00	68.87±1.16	81.55±0.28	74.86±0.26	**97.96±0.89**
glass06vs5	65.52±3.60	62.91±4.18	64.81±8.41	66.18±2.36	64.28±4.19	69.12±5.92	36.27±1.62	41.63±0.96	**81.24±5.17**
page-blocks13vs4	51.30±2.83	50.55±3.05	51.96±2.39	50.41±3.18	50.27±2.01	46.73±1.46	43.10±0.91	32.60±0.69	**92.57±2.90**
poker8vs6	2.50±0.31	2.36±0.74	2.83±0.30	3.06±0.87	2.47±0.49	2.59±0.25	2.49±0.79	6.95±1.33	**21.51±9.25**
W/T/L	0/0/12	0/0/12	0/0/12	0/0/12	0/0/12	0/0/12	1/0/11	1/0/11	**10/0/2**

Table 3
GM with standard deviation of different methods in classifying various datasets by SVM.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	49.11±4.50	48.76±3.39	48.07±3.26	48.19±3.06	48.33±2.67	48.92±4.45	44.61±1.14	2.03±1.14	**62.34±6.49**
vehicle2	84.99±1.94	84.96±1.44	85.24±1.74	85.25±2.05	84.19±2.67	84.81±1.97	75.39±0.46	71.92±0.24	**93.91±1.56**
glass0123vs456	88.94±4.82	87.56±4.59	87.85±4.09	88.83±2.90	88.51±5.19	86.85±5.08	84.90±0.10	**91.85±0.35**	89.64±1.21
ecoli1	87.21±1.88	88.14±1.24	86.37±1.33	86.00±1.72	87.39±1.40	74.32±5.11	87.76±0.35	**90.08±0.12**	83.91±2.75
ecoli2	84.39±2.86	81.21±3.29	83.78±3.81	83.87±3.04	84.52±2.89	73.84±6.56	**93.37±0.28**	93.16±0.09	88.49±4.17
segment0	98.69±0.28	94.59±0.67	98.67±0.30	98.67±0.28	98.49±0.35	98.65±0.25	50.71±3.68	43.78±2.95	**98.96±0.25**
ecoli034vs5	88.22±1.42	86.34±0.87	**91.27±2.62**	90.86±2.53	88.41±0.35	90.89±0.35	90.47±0.36	89.47±0.30	90.48±2.50
ecoli0234vs5	89.17±1.71	86.35±0.60	91.87±1.11	**91.88±1.03**	87.70±0.84	89.30±0.68	90.65±0.05	89.42±0.18	88.82±2.24
vowel0	94.48±0.57	91.77±2.38	93.99±0.85	92.02±0.91	94.89±0.53	92.77±0.70	97.64±0.04	96.43±0.11	**99.59±0.35**
glass06vs5	93.20±2.21	92.71±2.06	92.25±4.21	**94.28±0.75**	92.63±2.21	92.31±3.34	74.51±4.62	83.48±1.14	87.83±6.63
page-blocks13vs4	87.13±3.07	82.75±3.51	87.46±2.59	87.45±3.00	86.85±3.02	85.39±2.02	74.51±0.59	81.54±1.05	**97.61±1.07**
poker8vs6	48.18±3.64	48.48±8.30	49.56±4.06	52.70±7.61	49.01±3.25	50.33±2.56	19.82±5.31	24.26±1.01	**62.51±6.47**
W/T/L	0/0/12	1/0/11	1/0/11	2/0/10	0/0/12	0/0/12	1/0/11	2/0/10	**6/0/6**

5.3. Comparison methods

The effectiveness of the proposed DAMO in classifying imbalanced datasets is assessed in comparison to other methods. We specifically use a variety of neighbor-based over-sampling methods including SMOTE [15], Borderline-SMOTE [16], Safe-Level-SMOTE [33], CGMOS [34], OREM [22], and OHIT [18]. The reason is that DAMO involves a neighbor-based over-sampling method. Moreover, OHIT [18] and OREM [22] represent over-sampling methods based on clustering and semi-supervised strategy, respectively. Particularly, GAN-SMOTE [53] and E-EVRS [3], which are generative adversarial network (GAN) and ensemble methods respectively, are also employed to provide a broader comparative context. This inclusion allows for an evaluation of DAMO's performance relative to advanced methods from different methodological paradigms.

5.4. Performance evaluation

Here, Support Vector Machin (SVM) [50] and Multi-Layer Perceptron (MLP) [51] are utilized as basic classification tools. In experiments, the imbalanced datasets are first preprocessed using different over-sampling methods to balance the class distribution before being fed into the base classifiers. We first exhibit results of different methods in classifying various datasets. Then, we conduct a theoretical analysis of different methods based on the experimental results.

5.4.1. Experimental results

Tables 2-5 report the classification results of different methods in classifying various datasets by SVM. The best results for each dataset are highlighted in bold, making it easy to identify the top-performing method. The wins/ties/losses (W/T/L) statistics in the last row of tables further reinforce the superiority of the DAMO method. For example, the DAMO method consistently achieves the highest number of wins across multiple datasets, with fewer ties and losses compared to other methods. Thus, the findings indicate that the proposed DAMO method outperforms other methods in classifying imbalanced datasets.

Moreover, MLP is employed as the basic classifier to investigate the proposed DAMO method's performance compared to other methods when different classifiers are applied. In Tables 6-9 report the classification results of different methods in classifying various datasets by MLP. We can see from the results that DAMO is superior to other methods in most cases. In addition, the average ranking of different methods in terms of different indices is shown in Table 10. DAMO achieves the highest rankings, demonstrating

Table 4
AUC with standard deviation of different methods in classifying various datasets by SVM.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	60.35±1.97	58.85±1.70	60.79±1.64	60.97±0.78	59.88±2.36	60.02±1.36	44.73±1.12	46.12±0.43	**73.60±6.94**
vehicle2	96.71±0.22	96.05±0.41	96.66±0.20	96.61±0.15	96.71±0.19	96.53±0.30	87.60±0.13	85.20±0.38	**99.03±0.28**
glass0123vs456	96.20±1.11	96.04±0.95	96.20±1.01	96.24±0.81	96.22±1.18	96.16±1.12	96.80±0.04	96.58±0.05	**97.47±1.16**
ecoli1	94.40±0.63	93.87±0.27	94.63±0.51	93.90±0.66	94.46±0.37	94.53±0.56	94.32±0.28	93.84±0.12	**94.86±0.76**
ecoli2	93.38±0.32	92.17±0.73	93.48±0.35	93.46±0.48	93.44±0.40	93.11±0.92	**95.71±0.11**	95.08±0.14	94.78±0.47
segment0	99.84±0.01	99.68±0.04	99.84±0.01	99.84±0.01	99.88±0.01	99.88±0.01	79.85±1.88	73.09±1.27	**99.96±0.01**
ecoli034vs5	92.25±1.79	89.17±2.18	93.47±1.31	93.19±0.58	91.28±1.58	91.67±1.43	95.25±0.16	**97.16±0.17**	94.69±1.04
ecoli0234vs5	91.01±0.99	88.66±2.24	92.65±0.72	92.54±0.53	88.80±1.40	90.91±0.82	95.86±0.16	96.19±0.67	**96.20±2.48**
vowel0	98.66±0.08	97.28±1.20	98.52±0.11	98.11±0.19	98.68±0.06	98.27±0.10	99.85±0.01	99.70±0.04	**99.99±0.01**
glass06vs5	97.06±2.01	97.27±0.65	97.47±1.36	97.36±1.77	97.16±1.74	97.87±1.60	17.70±1.11	84.62±1.28	**99.10±0.89**
page-blocks13vs4	96.13±0.22	95.68±0.58	96.14±0.27	96.03±0.45	96.14±0.28	94.85±0.63	90.61±0.18	88.72±0.42	**99.63±0.42**
poker8vs6	37.37±1.96	42.40±8.88	38.69±2.66	38.73±4.84	38.15±3.50	41.82±5.12	63.64±0.39	64.21±0.77	**79.76±4.41**
W/T/L	0/0/12	0/0/12	0/0/12	0/0/12	0/0/12	0/0/12	1/0/11	1/0/11	**10/0/2**

Table 5
AP with standard deviation of different methods in classifying various datasets by SVM.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	74.60±0.96	74.23±0.98	74.69±1.15	74.22±1.11	74.53±1.33	74.20±0.95	64.11±0.74	65.25±0.35	**80.51±4.73**
vehicle2	98.15±0.08	97.87±0.16	98.14±0.08	98.12±0.07	98.15±0.07	98.09±0.11	94.25±0.09	92.72±0.21	**98.88±0.09**
glass0123vs456	95.55±0.55	95.52±0.46	95.48±0.60	95.53±0.39	95.52±0.63	95.58±0.54	**96.00±0.02**	95.95±0.02	95.96±0.57
ecoli1	96.49±0.20	96.36±0.07	**96.54±0.17**	96.35±0.21	96.51±0.12	96.51±0.18	96.38±0.11	96.09±0.15	96.51±0.18
ecoli2	96.64±0.12	96.43±0.20	96.67±0.12	96.66±0.14	96.62±0.13	96.74±0.21	**97.10±0.04**	96.44±0.08	96.92±0.17
segment0	99.72±0.00	99.70±0.01	99.72±0.00	99.72±0.00	99.73±0.00	99.73±0.00	94.22±0.50	92.04±0.28	**99.74±0.00**
ecoli034vs5	95.93±0.41	95.38±0.70	96.15±0.30	96.07±0.19	95.80±0.39	95.80±0.36	96.52±0.03	**96.86±0.06**	96.33±0.26
ecoli0234vs5	95.70±0.21	95.38±0.52	95.94±0.23	95.93±0.15	95.35±0.38	95.81±0.17	96.62±0.03	**96.90±0.03**	96.66±0.46
vowel0	99.31±0.01	99.13±0.15	99.30±0.01	99.25±0.02	99.31±0.01	99.27±0.01	99.43±0.00	99.41±0.00	**99.44±0.00**
glass06vs5	94.71±0.14	94.71±0.04	94.74±0.10	94.74±0.12	94.72±0.11	94.77±0.11	82.84±0.33	93.53±0.08	**94.86±0.09**
page-blocks13vs4	98.63±0.01	98.59±0.03	98.63±0.02	98.62±0.03	98.63±0.02	98.53±0.05	98.23±0.01	98.11±0.03	**98.85±0.03**
poker8vs6	97.98±0.08	98.46±0.26	97.92±0.13	97.85±0.24	98.07±0.16	98.15±0.15	99.06±0.02	98.82±0.02	**99.27±0.18**
W/T/L	0/0/12	0/0/12	1/0/11	0/0/12	0/0/12	0/0/12	2/0/10	2/0/10	**7/0/5**

Table 6
FM with standard deviation of different methods in classifying various datasets by MLP.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	53.61±10.29	51.73±6.84	54.80±10.16	**55.73±2.63**	54.09±6.29	55.03±7.53	21.65±3.21	12.11±3.23	53.36±7.06
vehicle2	82.58±3.78	81.91±2.90	83.96±4.69	84.48±5.01	80.73±4.61	81.49±3.34	44.65±3.04	84.56±1.70	**87.99±3.30**
glass0123vs456	80.67±5.10	76.66±8.89	81.59±5.04	83.01±7.29	75.77±11.79	80.80±6.88	54.53±5.08	20.39±3.52	**87.23±2.70**
ecoli1	49.46±5.99	55.04±7.40	59.02±8.25	49.50±6.64	56.37±7.47	43.28±6.74	75.22±0.33	**76.86±0.52**	62.35±4.16
ecoli2	49.46±5.99	55.04±7.40	59.02±8.25	49.50±6.64	56.37±7.47	43.28±6.74	**78.88±0.34**	73.84±0.52	62.35±4.16
segment0	98.04±0.30	93.14±1.01	97.91±0.76	98.00±0.33	96.97±0.63	97.51±0.66	39.67±6.30	48.17±3.05	**98.69±0.33**
ecoli034vs5	78.34±4.35	76.94±4.99	**82.01±7.06**	78.75±5.41	72.59±5.43	75.29±5.97	80.78±1.78	69.23±0.23	79.85±5.25
ecoli0234vs5	76.78±4.24	75.51±2.86	79.94±4.51	81.34±4.29	73.65±2.35	71.49±5.53	81.93±3.17	66.84±0.71	**83.22±2.32**
vowel0	94.18±2.56	94.15±0.99	93.95±1.23	92.89±2.20	90.42±2.19	90.73±2.23	83.94±2.29	70.61±0.19	**98.33±0.43**
glass06vs5	86.78±10.95	81.15±5.97	84.40±6.82	78.70±7.74	86.04±6.27	81.46±7.50	13.82±2.16	3.04±3.73	**94.40±3.93**
page-blocks13vs4	79.56±4.57	78.02±9.70	83.47±4.54	78.94±3.93	83.50±2.55	75.46±4.89	66.27±2.25	41.19±0.46	**91.84±1.43**
poker8vs6	67.13±6.63	64.05±7.33	50.12±7.13	67.61±5.87	45.33±11.55	38.37±13.05	18.65±1.03	33.18±3.76	**77.89±7.88**
W/T/L	0/0/12	0/0/12	1/0/11	1/0/11	0/0/12	0/0/12	1/0/11	1/0/11	**8/0/4**

its effectiveness and robustness irrespective of the classifier used. These findings not only highlight DAMO's capability to classify imbalanced datasets but also confirm its reliability and adaptability when applied with different classifiers.

5.4.2. Experimental analysis

Earlier studies have shown DAMO's proficiency in classifying imbalanced datasets compared to related methods. Based on these results, we conduct theoretical analyses of a range of methodologies.

- Methods such as SMOTE, Borderline-SMOTE, Safe-Level-SMOTE, and CGMOS are known for their effectiveness in addressing class imbalance by generating synthetic samples in the neighborhood of minority class samples. However, these methods often overlook the potential inaccuracies introduced by the synthetic samples themselves, which can distort the original data distribution and

Table 7
GM with standard deviation of different methods in classifying various datasets by MLP.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	59.30±6.92	58.68±5.09	60.10±8.13	60.26±3.82	60.22±4.98	60.91±5.16	5.26±2.80	6.53±4.15	**61.60±5.88**
vehicle2	90.39±2.75	89.86±1.74	91.26±2.84	91.59±3.35	89.20±4.00	90.17±2.34	36.82±10.73	91.60±1.49	**93.81±2.05**
glass0123vs456	87.06±4.30	81.89±8.86	87.74±3.14	88.15±5.84	83.22±10.38	85.27±5.42	42.34±5.97	22.74±3.80	**91.57±1.55**
ecoli1	61.21±6.28	64.19±12.75	70.40±6.71	59.97±7.46	69.59±6.79	54.77±9.08	86.67±0.32	**88.96±0.51**	71.38±3.69
ecoli2	61.21±6.28	64.19±12.75	70.40±6.71	59.97±7.46	69.59±6.79	54.77±9.08	90.24±0.17	**90.92±0.41**	71.38±3.69
segment0	99.11±0.16	98.02±0.16	99.06±0.19	99.02±0.22	99.13±0.23	98.91±0.14	43.67±6.36	50.35±2.43	**99.17±0.18**
ecoli034vs5	87.92±4.16	87.47±3.45	89.33±5.05	88.32±3.36	87.55±1.76	89.40±2.58	**89.60±0.54**	88.30±0.15	88.09±4.23
ecoli0234vs5	89.30±1.70	88.20±2.57	89.79±2.82	90.48±2.06	89.64±1.21	87.55±2.94	89.73±1.14	88.15±0.22	**91.69±1.54**
vowel0	99.24±0.38	98.44±0.72	99.22±0.21	98.80±0.48	98.88±0.28	98.71±0.64	95.46±2.05	95.45±0.14	**99.82±0.05**
glass06vs5	95.06±5.93	92.17±4.01	93.04±4.08	88.55±6.37	94.96±3.43	93.13±2.94	10.42±2.28	4.06±4.98	**96.28±3.17**
page-blocks13vs4	93.76±0.64	90.90±6.97	94.35±5.03	92.75±4.66	97.02±0.20	96.72±1.87	93.92±0.83	88.94±0.73	**97.24±2.02**
poker8vs6	90.12±3.17	84.42±8.71	71.00±4.92	91.47±5.97	79.74±7.34	**93.31±4.18**	38.72±2.65	38.15±3.42	82.77±6.01
W/T/L	0/0/12	0/0/12	0/0/12	0/0/12	0/0/12	1/0/11	1/0/11	2/0/10	**8/0/4**

Table 8
AUC with standard deviation of different methods in classifying various datasets by MLP.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	67.98±2.96	65.49±6.20	70.03±5.38	69.03±3.61	68.20±2.95	67.71±3.02	56.63±3.39	53.73±4.44	**72.18±3.80**
vehicle2	94.92±3.41	94.59±2.22	95.36±3.47	95.94±2.52	95.60±2.59	96.23±2.19	62.43±6.62	97.43±0.21	**98.66±0.81**
glass0123vs456	93.17±2.01	90.43±7.30	94.63±1.28	93.96±3.35	94.34±4.07	90.99±6.02	84.08±2.25	90.41±2.76	**97.62±1.16**
ecoli1	73.98±8.01	78.87±6.35	79.49±5.54	74.07±7.70	79.19±7.60	70.99±7.46	95.71±0.12	**95.77±0.06**	86.54±9.14
ecoli2	73.98±8.01	78.87±6.35	79.49±5.54	74.07±7.70	79.19±7.60	70.99±7.46	**94.22±0.09**	94.01±0.20	86.54±9.14
segment0	99.93±0.08	99.78±0.04	99.89±0.05	99.91±0.07	99.95±0.02	99.85±0.11	89.76±0.60	76.54±0.90	**99.99±0.01**
ecoli034vs5	93.58±2.97	95.14±1.04	95.64±1.44	95.69±2.72	94.92±2.98	93.72±1.67	90.74±0.42	92.73±0.31	**95.78±2.91**
ecoli0234vs5	94.63±1.85	95.20±2.16	94.33±2.56	95.37±1.88	94.14±0.98	91.71±2.16	90.65±0.53	90.73±1.18	**96.19±1.93**
vowel0	99.89±0.08	99.68±0.34	99.84±0.19	99.86±0.20	99.92±0.07	99.90±0.07	98.72±0.89	98.84±0.04	**100.00±0.00**
glass06vs5	99.19±1.31	96.87±3.77	99.10±1.47	97.40±1.78	99.10±0.82	97.08±3.30	50.35±5.71	73.50±0.89	**100.00±0.00**
page-blocks13vs4	98.27±1.71	98.00±1.73	99.04±0.35	98.18±1.15	98.63±0.88	98.85±1.48	94.93±1.60	93.73±0.07	**99.68±0.23**
poker8vs6	97.48±2.62	98.04±1.78	93.18±4.39	**99.67±0.20**	89.59±4.58	98.25±2.28	78.90±1.08	83.93±0.70	96.22±2.41
W/T/L	0/0/12	0/0/12	0/0/12	1/0/11	0/0/12	0/0/12	1/0/11	1/0/11	**9/0/3**

Table 9
AP with standard deviation of different methods in classifying various datasets by MLP.

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
glass1	76.97±2.75	75.27±4.98	78.18±3.80	78.60±1.57	77.73±1.66	76.20±2.05	69.46±1.45	69.75±2.25	**80.82±2.28**
vehicle2	97.40±1.08	97.43±0.79	97.76±1.09	97.81±0.86	97.83±0.83	98.03±0.66	81.95±3.36	98.33±0.11	**98.79±0.24**
glass0123vs456	94.09±1.12	93.51±2.65	94.66±0.81	94.16±1.38	95.07±1.51	93.30±2.91	90.32±1.01	93.26±1.14	**96.02±0.44**
ecoli1	87.81±3.21	89.84±3.29	90.10±1.86	87.49±3.07	89.92±3.29	86.64±3.13	96.83±0.03	**96.85±0.02**	92.88±4.50
ecoli2	87.81±3.21	89.84±3.29	90.10±1.86	87.49±3.07	89.92±3.29	86.64±3.13	**96.77±0.04**	96.27±0.15	92.88±4.50
segment0	99.73±0.02	99.71±0.01	99.73±0.01	99.73±0.01	99.74±0.00	99.72±0.02	97.26±0.17	93.28±0.21	**99.75±0.00**
ecoli034vs5	96.28±0.48	**96.58±0.17**	96.56±0.32	96.45±0.74	96.39±0.79	96.11±0.46	95.42±0.08	96.20±0.14	96.50±0.71
ecoli0234vs5	96.45±0.43	96.55±0.47	96.33±0.56	96.57±0.32	96.37±0.39	95.92±0.51	95.88±0.17	96.35±0.18	**96.66±0.49**
vowel0	99.43±0.01	99.39±0.08	99.43±0.02	99.43±0.02	99.44±0.01	99.43±0.01	99.29±0.14	99.33±0.00	**99.44±0.00**
glass06vs5	94.87±0.14	94.56±0.54	94.85±0.16	94.68±0.23	94.88±0.06	94.71±0.22	88.02±0.78	91.54±0.38	**94.95±0.00**
page-blocks13vs4	98.71±0.21	98.73±0.13	98.81±0.02	98.75±0.08	98.76±0.10	98.74±0.22	98.15±0.35	98.46±0.01	**98.85±0.02**
poker8vs6	99.60±0.08	99.63±0.02	99.56±0.07	**99.65±0.00**	99.43±0.17	99.64±0.03	99.32±0.05	99.24±0.02	**99.60±0.04**
W/T/L	0/0/12	1/0/11	0/0/12	1/0/11	0/0/12	0/0/12	1/0/11	1/0/11	**8/0/4**

introduce uncertainty into the classification process. They primarily focus on quantity rather than quality, potentially leading to the creation of samples in regions of low density or unrealistic data points that do not reflect the underlying data manifold, thus negatively impacting classification accuracy.

- OHIT and OREM adopt clustering and semi-supervised learning paradigms, respectively, to address the class imbalance problem. While these methods attempt to preserve the intrinsic structure of the data by focusing on the distributional aspects of the minority class, they may fall short in accurately capturing the nuances of the data manifold. OHIT and OREM might not sufficiently account for the inherent uncertainties and variabilities in the minority class, potentially leading to over-generalization or under-representation of the minority class samples. Additionally, the reliance on clustering or semi-supervised mechanisms might introduce biases if the underlying assumptions about the data distribution are not met, affecting the robustness and reliability of the classification results.

Table 10
Average ranking of different methods in terms of different indices.

Classifier	Indices	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN-SMOTE	E-EVRS	DAMO
SVM	FM	4.92	6.17	4.75	4.42	6.33	6.58	5.17	5.42	**1.25**
	GM	4.42	6.33	4.42	3.92	5.33	5.67	6.17	5.67	**3.08**
	AUC	5.58	7.42	4.33	5.08	5.08	5.67	5.08	5.42	**1.33**
	AP	5.08	7.08	4.42	5.92	5.08	5.08	4.92	5.92	**1.50**
MLP	FM	4.50	5.75	3.58	4.08	5.67	6.33	6.17	7.00	**1.92**
	GM	4.83	6.75	4.08	4.67	4.83	5.17	6.08	6.50	**2.08**
	AUC	5.17	6.00	4.00	4.17	4.25	5.50	7.50	6.67	**1.67**
	AP	5.25	5.50	4.25	4.50	3.83	5.92	7.58	6.33	**1.83**

- GAN-SMOTE employs a shallow GAN architecture with small-batch discriminant and discretization constraints to generate diverse synthetic samples. E-EVRS filters unreliable synthetic samples using belief function theory and integrates them into a dynamically weighted ensemble learning framework to enhance classification performance, particularly in high-noise and complex multi-class scenarios. While GAN-SMOTE effectively learns the original data distribution through its generative network, it fails to account for the uncertainty and potential bias of the generated samples, which can negatively impact classification. Although E-EVRS demonstrates robustness in filtering unreliable samples and improving ensemble performance, it does not address the specific issue of distribution inconsistency during the sampling process. Thus, these methods fall short of addressing the challenges tackled by the proposed DAMO method.
- DAMO stands out by explicitly addressing the limitations of comparison methods. It quantifies the inconsistency of data distribution post-sampling, ensuring that the synthetic samples are not only numerous but also accurate and representative of the true data landscape. This method minimizes data shift caused by inaccurately synthesized samples, thereby preserving the integrity of the data distribution. Moreover, DAMO identifies and calibrates imprecise samples that are indistinguishable between classes, utilizing neighboring information to refine classification decisions. This method not only reduces the likelihood of misclassification due to inaccurate synthetic samples but also enhances the reliability of classification results by leveraging the collective wisdom of the neighbors. Consequently, DAMO achieves a superior balance between data augmentation and data fidelity, outperforming other methods in terms of classification accuracy and robustness when dealing with imbalanced datasets.

5.5. Discussions of DAMO

Beyond the previously described experimental findings, we conducted extensive studies to examine and analyze the proposed DAMO method. Section 5.5.1 delves into the efficacy of DAMO using various statistical indices to assess data distribution, while Section 5.5.2 details the impact of the DAMO-related parameter K. Furthermore, an ablation analysis is conducted to evaluate the impact of each phase of DAMO, as detailed in Section 5.5.3. Ultimately, Section 5.5.4 details the examination of DAMO's computational complexity and the time taken to execute various methods. Moreover, Section 5.5.5 clarifies how the proposed DAMO method can be adapted for multi-class scenarios.

5.5.1. Assessment of data distribution

The DAMO method measures the distribution's inconsistency using the statistical metric MMD. However, alternative measures such as the Kullback-Leibler divergence (KLD) [54] are applicable in this context. This part delves into the efficacy of DAMO using various statistical measures, with findings presented in Fig. 2, where each point represents the result of a dataset. x-axis and y-axis represent index values of the proposed DAMO method with the MMD and KLD measures, respectively. It is evident that DAMO using MMD typically surpasses the performance of KLD. Consequently, our suggestion is to use MMD as the standard statistical measure in DAMO, with other statistical indices being viable as substitute options in various applications.

5.5.2. Influence of parameters

Within the proposed DAMO method, it is essential to adjust the parameters K and T in applications.

The symbol K signifies not only the count of neighboring elements used to create new samples but also the quantity used to calibrate the classification outcomes of imprecise samples. Fig. 3 illustrates the performance of the DAMO method across various values of $K \in [3, 12]$. In Fig. 3, the x-axis represents different values of K, while the y-axis shows the corresponding results for various performance indices. Different colors and symbols are used to distinguish between these indices in a point-line plot, which helps visualize how the performance of DAMO changes with varying K values. As depicted in Fig. 3, with the increase of K, the variations in the DAMO result are small. The parameter K exhibits robustness in the proposed DAMO method due to its adaptive weight optimization mechanism. Unlike traditional over-sampling methods that rigidly depend on a fixed KNN structure, DAMO dynamically adjusts the weights of these neighbors to align the generated synthetic samples with the original data distribution. This adaptive weighting ensures that even when K is set suboptimally (*e.g.*, too small or too large), the method prioritizes neighbors that preserve the intrinsic data structure while suppressing those that introduce noise. Moreover, in the process of calibrating the classification results of imprecise samples, different weights are given according to the distance from the test sample to its neighbors. The closer the neighbor is to the test sample, the greater the weight. This ensures that distant neighbors have less influence, while

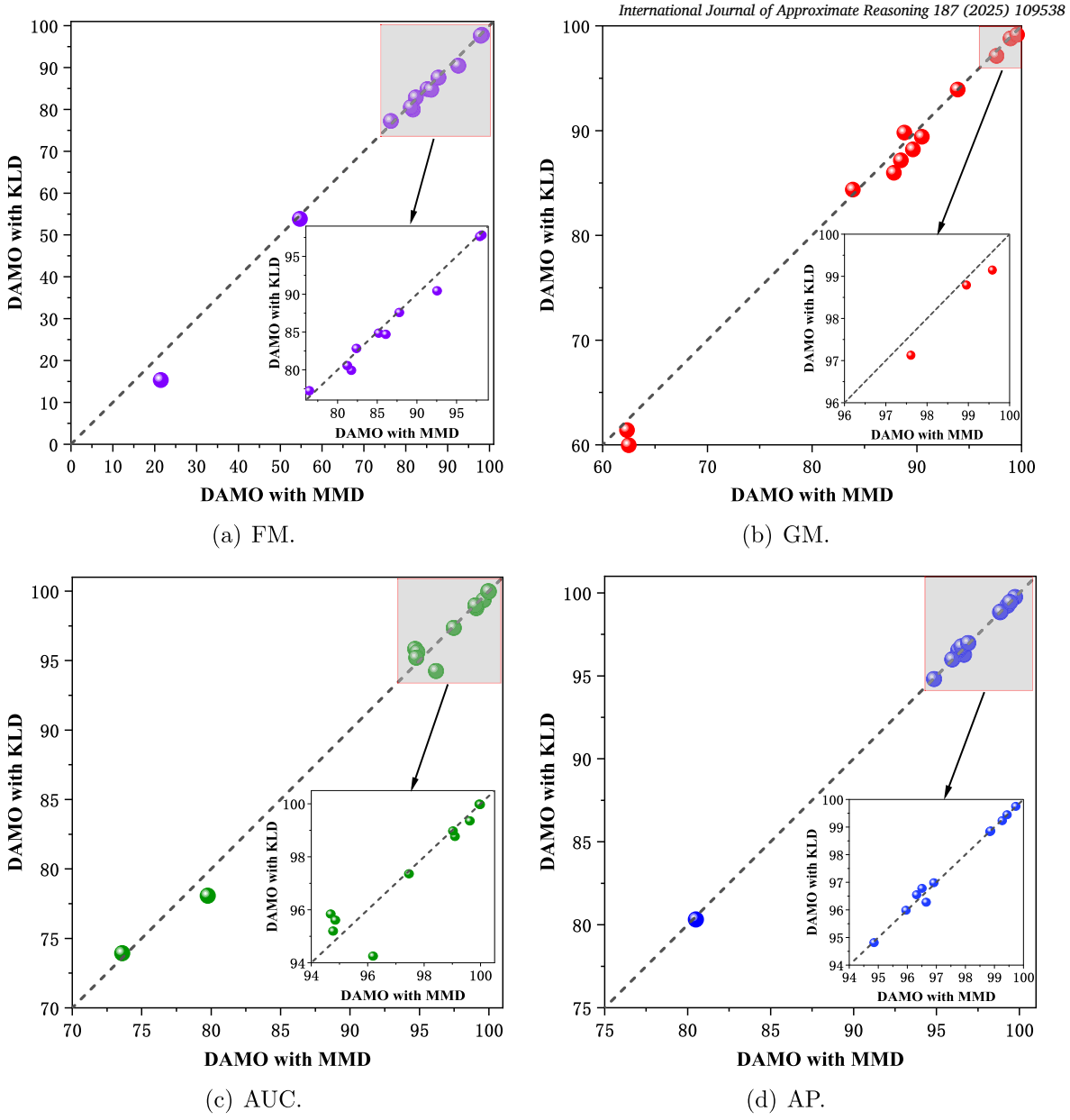

Fig. 2. Classification results of DAMO with various statistical indices.

closer neighbors play a key role. Therefore, DAMO can achieve stable performance across a wide range of K values. Experimental results further validate this robustness, demonstrating minimal performance degradation even when K varies between 3 and 12. In applications, we set $K = 7$ as the default value for applications because it typically produces high performance.

T represents the count of different forms of synthetic samples generated. The study employs various unbalanced data sets to evaluate the effectiveness of DAMO at different T values, facilitating the implementation of T. The classification results of DAMO with varying T values are presented in Fig. 4, where the x-axis represents different values of T, ranging from 2 to 8, while the y-axis shows the corresponding performance indices. Different colors and symbols are used to distinguish between these indices in a point-line plot, which facilitates the visualization of how the performance of DAMO changes with different T values. We can observe that DAMO maintains stability regardless of the value of T, exhibiting slight variations as T rises. The proposed DAMO method exhibits robustness to T due to the local reliability assessment mechanism embedded in the fusion framework. While increasing T introduces diversity in the generated datasets, the subsequent evaluation of each classifier's reliability on the test sample's local neighbors ensures that suboptimal datasets are dynamically suppressed. Specifically, classifiers trained on datasets with poor over-sampling quality (e.g., those deviating from the original distribution) are assigned lower weights in the final decision fusion, as their performance on the test sample's neighbors is weaker. Conversely, classifiers that generalize well to the local data structure are weighted more heavily.

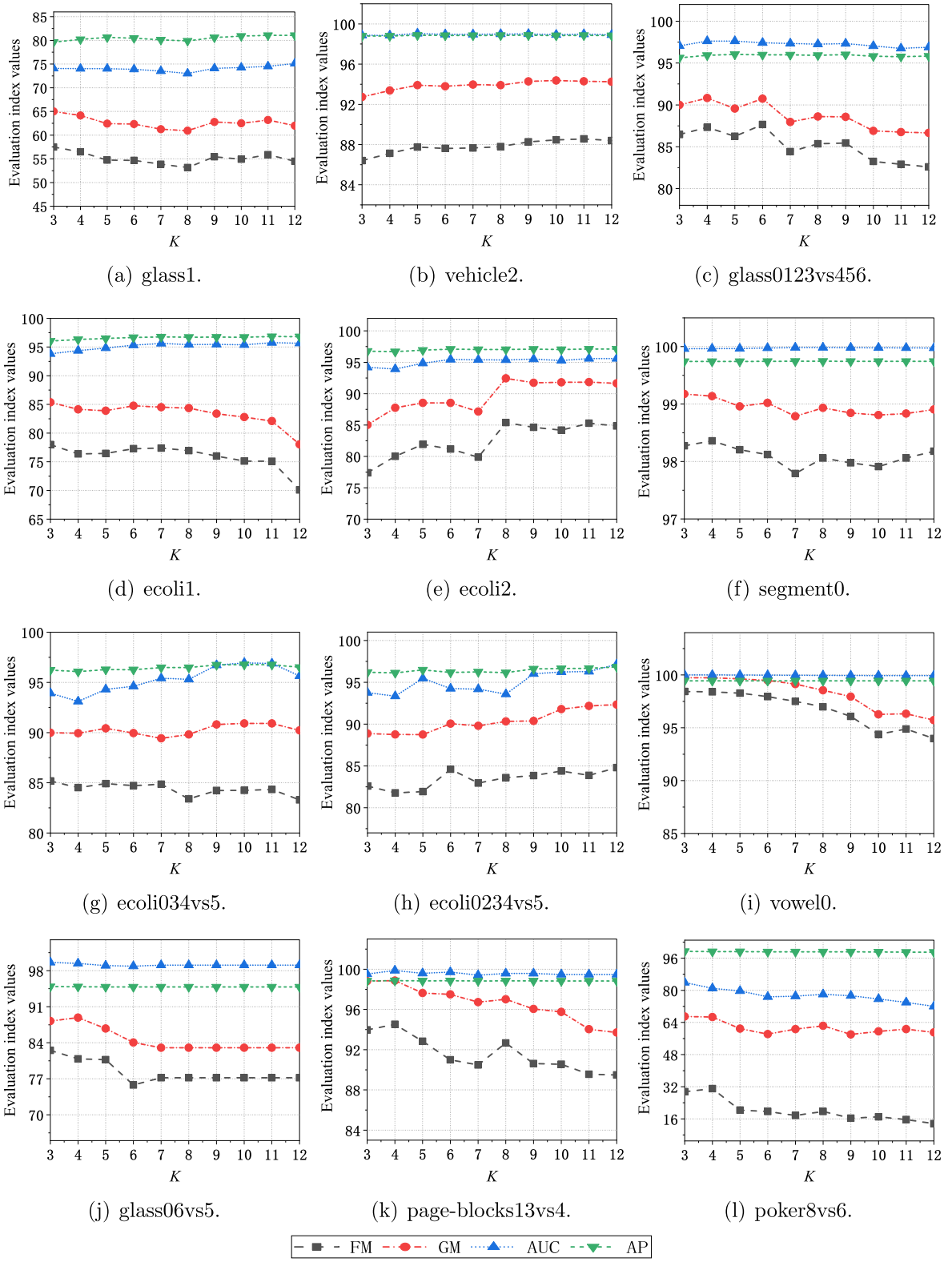

Fig. 3. Classification results for datasets with various K.

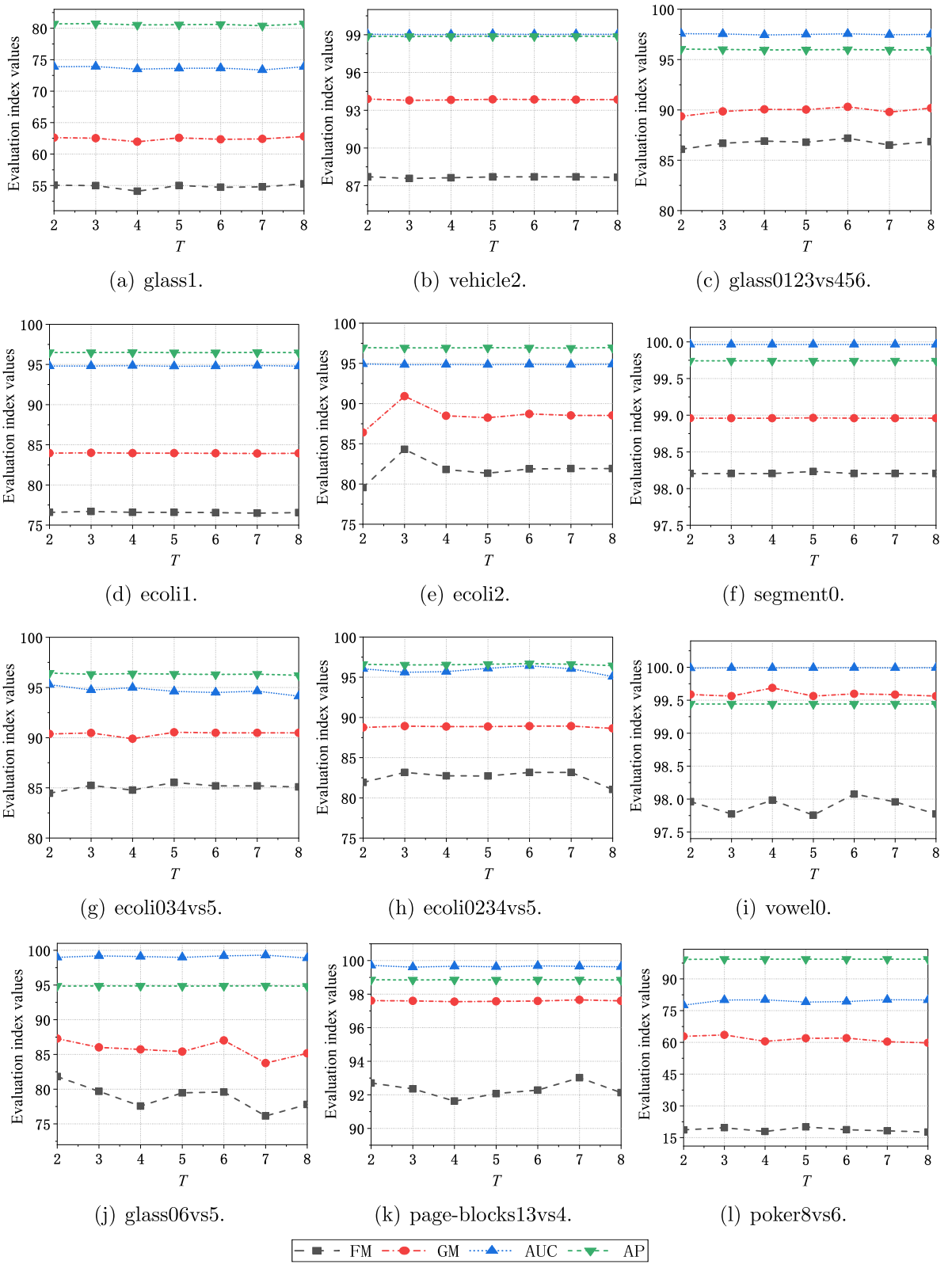

Fig. 4. Classification results for datasets with various T.

Table 11
Ablation experimental results obtained by different steps in DAMO.

	FM			GM		
Step 1	✗	✓	✓	✗	✓	✓
Step 2	✗	✗	✓	✗	✗	✓
glass1	51.00$_{\pm 5.63}$	51.17$_{\pm 3.35}$	54.72$_{\pm 7.86}$	50.35$_{\pm 3.64}$	51.58$_{\pm 2.70}$	62.34$_{\pm 6.49}$
vehicle2	76.47$_{\pm 1.92}$	76.79$_{\pm 2.06}$	87.81$_{\pm 2.50}$	86.91$_{\pm 1.50}$	87.21$_{\pm 1.42}$	93.91$_{\pm 1.56}$
glass0123vs456	81.21$_{\pm 3.00}$	82.33$_{\pm 3.82}$	86.10$_{\pm 2.56}$	86.87$_{\pm 1.99}$	87.55$_{\pm 2.72}$	89.64$_{\pm 1.21}$
ecoli1	72.57$_{\pm 2.85}$	73.47$_{\pm 1.60}$	76.45$_{\pm 4.20}$	84.01$_{\pm 2.62}$	85.09$_{\pm 0.87}$	83.91$_{\pm 2.75}$
ecoli2	69.04$_{\pm 3.44}$	69.13$_{\pm 3.39}$	81.75$_{\pm 4.86}$	84.14$_{\pm 2.86}$	84.17$_{\pm 2.82}$	88.49$_{\pm 4.17}$
segment0	96.42$_{\pm 1.22}$	96.51$_{\pm 1.22}$	98.20$_{\pm 0.29}$	98.59$_{\pm 0.34}$	98.60$_{\pm 0.34}$	98.96$_{\pm 0.25}$
ecoli034vs5	70.02$_{\pm 2.90}$	70.73$_{\pm 1.48}$	85.20$_{\pm 3.24}$	87.39$_{\pm 2.47}$	88.02$_{\pm 1.07}$	90.48$_{\pm 2.50}$
ecoli0234vs5	55.91$_{\pm 4.02}$	65.76$_{\pm 3.68}$	82.40$_{\pm 4.08}$	84.31$_{\pm 1.72}$	87.82$_{\pm 0.86}$	88.82$_{\pm 2.24}$
vowel0	67.20$_{\pm 1.83}$	68.34$_{\pm 1.98}$	97.96$_{\pm 0.89}$	94.52$_{\pm 0.54}$	94.69$_{\pm 0.79}$	99.59$_{\pm 0.35}$
glass06vs5	60.63$_{\pm 2.93}$	61.03$_{\pm 5.75}$	81.24$_{\pm 5.17}$	91.69$_{\pm 1.44}$	91.96$_{\pm 1.82}$	87.83$_{\pm 6.63}$
page-blocks13vs4	51.83$_{\pm 2.36}$	51.91$_{\pm 2.54}$	92.57$_{\pm 2.90}$	87.16$_{\pm 3.02}$	87.46$_{\pm 2.55}$	97.61$_{\pm 1.07}$
poker8vs6	2.24$_{\pm 0.21}$	2.22$_{\pm 0.25}$	21.51$_{\pm 9.25}$	46.18$_{\pm 2.80}$	46.11$_{\pm 3.30}$	62.51$_{\pm 6.47}$

Table 12
Execution time of different methods (in seconds).

Datasets	SMOTE	Borderline -SMOTE	Safe-Level -SMOTE	CGMOS	OREM	OHIT	GAN -SMOTE	E-EVRS	DAMO
glass1	**0.1531**	0.7185	0.9142	0.1985	1.0232	0.7636	3.6731	0.3133	0.2803
vehicle2	1.3127	5.1624	6.5841	**0.3214**	6.2213	6.5040	9.9763	3.8533	0.7725
glass0123vs456	1.0382	1.4387	1.4425	1.0032	1.2121	0.9820	3.5884	**0.4124**	0.8653
ecoli1	**0.1652**	2.1173	0.7429	0.2492	1.7996	0.8140	3.9108	1.0231	0.3909
ecoli2	0.3725	2.3475	1.2928	**0.1978**	1.1623	0.4784	4.0140	1.0208	0.4163
segment0	7.0708	25.8692	25.1647	**1.4264**	20.9975	11.9459	26.6708	4.8235	2.8498
ecoli034vs5	0.3003	1.2422	1.4128	**0.2487**	0.3510	0.4598	3.9354	0.7011	0.4463
ecoli0234vs5	**0.2999**	0.6712	1.0772	0.4242	0.4215	0.3011	3.9127	0.692	0.5416
vowel0	2.3195	10.3309	5.3242	**0.4692**	0.8395	1.4284	3.8665	2.7858	1.1128
glass06vs5	0.1479	0.3460	0.5403	0.1566	**0.1354**	0.1504	3.7872	2.4819	0.3018
page-blocks13vs4	1.0229	4.1984	4.0922	**0.1966**	0.2850	0.3340	3.7064	1.7243	0.5168
poker8vs6	4.2884	19.2012	20.8678	1.2716	**0.9741**	1.0718	4.6304	4.4847	2.2491

This adaptive weighting mechanism effectively mitigates the impact of T, as even when T is varied, the fusion process inherently prioritizes reliable models and suppresses noise. Furthermore, the magnitude of T should not be too small, which cannot accurately characterize the uncertainty in the samples produced. However, setting the T value too high will result in significant computational strain. Hence, we propose to establish $T \in [3,5]$ as the standard value of the application.

5.5.3. Ablation study

This section delves into an ablation research conducted across various datasets to evaluate the impact of each step in DAMO. Each step was systematically amalgamated to assess the degree of improvement in the classification outcome. Table 11 displays the outcomes of ablation experiments conducted at various DAMO steps on various datasets, with ✓ and ✗ representing the integration and non-integration of distinct DAMO steps, respectively. The findings show that the incorporation of various steps enhances DAMO's overall efficiency, thus verifying the effectiveness of each step in DAMO.

5.5.4. Execution time

The computational complexity of DAMO consists mainly of two parts. In the first part, it is necessary to calculate the distances between minority class samples and neighbors during the process of over-sampling. Specifically, within the minority class Y_{min}, the sample $\mathbf{y}_i$ calculates the distances from $\mathbf{y}_i$ to every minority sample, with the computational complexity indicated as $\mathcal{O}(|Y_{min}|)$, where $|Y_{min}|$ indicates the count of samples in the minority class Y_{min}. It is presumed that, upon concluding the over-sampling procedure, M' samples require neighboring searches. Consequently, DAMO's overall computational complexity amounts to $\mathcal{O}(M'|Y_{min}|)$. In the second part, the maximum mean discrepancy (MMD) is utilized to measure the variability of the data distribution after over-sampling, with computational complexity denoted as $\mathcal{O}(|Y_{min}|^2)$. Consequently, DAMO's overall computational complexity amounts to $\mathcal{O}(M'|Y_{min}| + |Y_{min}|^2)$.

Table 12 enumerates the execution time in seconds for various methods across diverse datasets. Typically, SMOTE generally exhibits shorter execution times in most cases owing to its less complex operations. Particularly, during over-sampling, Borderline-SMOTE and Safe-level-SMOTE require time to calculate the threshold and evaluate the levels of safety. Additionally, CGMOS requires a certain duration to calculate how adding minority data affects the dataset's overall reliability, while OHIT also needs time to clustering datasets. Furthermore, the operational duration of DAMO mainly comes from the procedure to explore KNNs and evaluate the distribution of the data. Despite the DAMO's execution duration not being short, it remained within permissible ranges. Consequently, DAMO is ideal for scenarios where achieving precise classification is crucial and the efficiency of computation is not essential.

5.5.5. Extension to multi-class imbalanced classification

Although the proposed DAMO method is designed for classifying imbalanced data with binary classes, it can be naturally extended to handle multi-class (non-binary) imbalanced scenarios.

In the first step, DAMO introduces a multiple over-sampling strategy based on distribution assessment to balance the two classes. For the multi-class problem, this strategy is naturally extended by treating each minority class independently. Specifically, for each minority class, we apply our multiple over-sampling strategy to generate synthetic samples, ensuring that the intrinsic distribution of each class is preserved. This process results in T different balanced training sets. These datasets are then used to train classifiers thereby classifying test samples.

In the second step, the processes of evidence fusion and neighbors calibration can also be extended to a multi-class scenario. For each test sample, we quantify the local reliability of the trained classifiers, where the classification results for neighbors and test samples are multi-class rather than binary classification. For precise samples with high-reliability predictions, different classification results are directly fused using the DS discounting fusion rule, where more reliable models are assigned higher weights. For imprecise samples with low-reliability predictions, we use the classification results from the sample's neighbors to optimize different calibration matrices. Particularly, the size of these calibration matrices is $C \times C$ rather than 2×2, where C represents the number of classes, such that $C > 2$. The calibrated classification results are then fused using the DS discounting fusion rule to make a final decision.

6. Conclusion

In this paper, we investigate the distribution assessment-based optimal over-sampling with neighbor calibration (DAMO) method for classifying imbalanced data. DAMO has the advantages of minimizing the shift in data distribution and overcoming, to some extent, the negative impact of inaccurate sampling on the classification process. Experiments on various imbalanced datasets have demonstrated the efficacy of DAMO compared to other over-sampling methods. Using the Friedman test for statistical analysis reveals notable disparities between DAMO and its counterparts across different indices. In addition, we explore how the parameter impacts DAMO and offer guidelines for its implementation. Moreover, an ablation study can demonstrate the effectiveness of each step in DAMO. In the future, the DAMO application will expand to include more practical, real-world activities. Although the present work primarily addresses binary imbalanced classification, we recognize the importance and complexity of multi-class imbalanced learning in applications. In our future work, we will not only extend the proposed method to multi-class scenarios, but also investigate effective strategies to enhance its performance in such settings.

CRediT authorship contribution statement

Hongpeng Tian: Writing – original draft, Software, Methodology, Conceptualization. **Zuowei Zhang:** Writing – review & editing, Methodology, Conceptualization. **Zhunga Liu:** Writing – review & editing, Supervision, Conceptualization. **Jingwei Zuo:** Supervision, Methodology. **Caixing Yang:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

[1] D. Bai, G. Li, D. Jiang, J. Yun, B. Tao, G. Jiang, Y. Sun, Z. Ju, Surface defect detection methods for industrial products with imbalanced samples: a review of progress in the 2020s, Eng. Appl. Artif. Intell. 130 (2024) 107697.
[2] J. Yang, R. El-Bouri, O. O'Donoghue, A.S. Lachapelle, A.A. Soltan, D.W. Eyre, L. Lu, D.A. Clifton, Deep reinforcement learning for multi-class imbalanced training: applications in healthcare, Mach. Learn. 113 (5) (2024) 2655–2674.
[3] F. Grina, Z. Elouedi, E. Lefevre, Re-sampling of multi-class imbalanced data using belief function theory and ensemble learning, Int. J. Approx. Reason. 156 (2023) 1–15.
[4] S. Wang, J. Tian, P. Liang, X. Xu, Z. Yu, S. Liu, D. Zhang, Single and simultaneous fault diagnosis of gearbox via wavelet transform and improved deep residual network under imbalanced data, Eng. Appl. Artif. Intell. 133 (2024) 108146.
[5] A. Telikani, N.E. Rudbardeh, S. Soleymanpour, A. Shahbahrami, J. Shen, G. Gaydadjiev, R. Hassanpour, A cost-sensitive machine learning model with multitask learning for intrusion detection in iot, IEEE Trans. Ind. Inform. 20 (3) (2023) 3880–3890.
[6] H. Tian, Z. Zhang, A. Martin, Z. Liu, Reliability-based imbalanced data classification with Dempster-Shafer theory, in: International Conference on Belief Functions, Springer, 2022, pp. 77–86.
[7] Y. Li, S. Wang, J. Jin, F. Zhu, L. Zhao, J. Liang, C.P. Chen, Imbalanced complemented subspace representation with adaptive weight learning, Expert Syst. Appl. 249 (2024) 123555.
[8] A. Rosales-Pérez, S. García, F. Herrera, Handling imbalanced classification problems with support vector machines via evolutionary bilevel optimization, IEEE Trans. Cybern. 53 (8) (2022) 4735–4747.
[9] S. Xia, Y. Zheng, G. Wang, P. He, H. Li, Z. Chen, Random space division sampling for label-noisy classification or imbalanced classification, IEEE Trans. Cybern. 52 (10) (2021) 10444–10457.

[10] J. Guo, H. Wu, X. Chen, W. Lin, Adaptive sv-borderline smote-svm algorithm for imbalanced data classification, Appl. Soft Comput. 150 (2024) 110986.
[11] X.Y. Liu, J.X. Wu, Z.H. Zhou, Exploratory undersampling for class-imbalance learning, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 39 (2) (2008) 539–550.
[12] W.W. Ng, S. Xu, J. Zhang, X. Tian, T. Rong, S. Kwong, Hashing-based undersampling ensemble for imbalanced pattern classification problems, IEEE Trans. Cybern. 52 (2) (2020) 1269–1279.
[13] H. Tian, Z. Zhang, Z. Liu, J. Zuo, Multi-oversampling with evidence fusion for imbalanced data classification, in: International Conference on Belief Functions, Springer, 2024, pp. 68–77.
[14] Y.T. Yan, Z.B. Wu, X.Q. Du, J. Chen, S. Zhao, Y.P. Zhang, A three-way decision ensemble method for imbalanced data oversampling, Int. J. Approx. Reason. 107 (2019) 1–16.
[15] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
[16] H. Han, W.Y. Wang, B.H. Mao, Borderline-smote: a new over-sampling method in imbalanced data sets learning, in: International Conference on Intelligent Computing, Springer, 2005, pp. 878–887.
[17] X. Lu, X. Ye, Y. Cheng, An overlapping minimization-based over-sampling algorithm for binary imbalanced classification, Eng. Appl. Artif. Intell. 133 (2024) 108107.
[18] T. Zhu, C. Luo, Z. Zhang, J. Li, S. Ren, Y. Zeng, Minority oversampling for imbalanced time series classification, Knowl.-Based Syst. 247 (2022) 108764.
[19] M.K. Paul, B. Pal, A.S. Sattar, A.M.R. Siddique, M.A.M. Hasan, Carbo: clustering and rotation based oversampling for class imbalance learning, Knowl.-Based Syst. (2024) 112196.
[20] X. Kong, X. Wei, X. Liu, J. Wang, W. Xing, W. Lu, Fgbc: flexible graph-based balanced classifier for class-imbalanced semi-supervised learning, Pattern Recognit. 143 (2023) 109793.
[21] Z. Jiang, L. Zhao, Y. Lu, Y. Zhan, Q. Mao, A semi-supervised resampling method for class-imbalanced learning, Expert Syst. Appl. 221 (2023) 119733.
[22] T. Zhu, X. Liu, E. Zhu, Oversampling with reliably expanding minority class regions for imbalanced data learning, IEEE Trans. Knowl. Data Eng. 35 (6) (2022) 6167–6181.
[23] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Stat. 38 (1967) 325–339.
[24] G. Shafer, A Mathematical Theory of Evidence, vol. 42, Princeton Univ. Press, 1976.
[25] T. Denœux, Uncertainty quantification in logistic regression using random fuzzy sets and belief functions, Int. J. Approx. Reason. 168 (2024) 109159.
[26] T. Denœux, Decision-making with belief functions: a review, Int. J. Approx. Reason. 109 (2019) 87–110.
[27] H. Tian, Z. Zhang, W. Ding, Incomplete data transfer calibration classification, IEEE Trans. Emerg. Top. Comput. Intell. (2024), https://doi.org/10.1109/TETCI.2024.3449902.
[28] T. Denœux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, IEEE Trans. Syst. Man Cybern. 25 (5) (1995) 804–813.
[29] T. Denœux, A neural network classifier based on Dempster-Shafer theory, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 30 (2) (2000) 131–150.
[30] L. Jiao, H. Zhang, X. Geng, Q. Pan, Belief rule learning and reasoning for classification based on fuzzy belief decision tree, Int. J. Approx. Reason. 175 (2024) 109300.
[31] Z. Zhang, Z. Liu, H. Tian, A. Martin, Mixed-type imputation for missing data credal classification via quality matrices, IEEE Trans. Syst. Man Cybern. Syst. 54 (8) (2024) 4772–4785.
[32] T. Denœux, P.P. Shenoy, An interval-valued utility theory for decision making with Dempster-Shafer belief functions, Int. J. Approx. Reason. 124 (2020) 194–216.
[33] C. Bunkhumpornpat, K. Sinapiromsaran, Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2009, pp. 475–482.
[34] X. Zhang, D. Ma, L. Gan, S. Jiang, G. Agam, Cgmos: certainty guided minority oversampling, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2016, pp. 1623–1631.
[35] X. Chao, L. Zhang, Few-shot imbalanced classification based on data augmentation, Multimed. Syst. 29 (5) (2023) 2843–2851.
[36] P. Sun, Z. Wang, L. Jia, Z. Xu, Smote-ktlnn: a hybrid re-sampling method based on smote and a two-layer nearest neighbor classifier, Expert Syst. Appl. 238 (2024) 121848.
[37] D. Dablain, B. Krawczyk, N.V. Chawla, Deepsmote: fusing deep learning and smote for imbalanced data, IEEE Trans. Neural Netw. Learn. Syst. 34 (9) (2022) 6390–6404.
[38] A. Özdemir, K. Polat, A. Alhudhaif, Classification of imbalanced hyperspectral images using SMOTE-based deep learning methods, Expert Syst. Appl. 178 (2021) 114986.
[39] F. Shen, X. Zhao, G. Kou, F.E. Alsaadi, A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique, Appl. Soft Comput. 98 (2021) 106852.
[40] S.A. Alex, Imbalanced data learning using smote and deep learning architecture with optimized features, Neural Comput. Appl. 37 (2) (2025) 967–984.
[41] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science 344 (6191) (2014) 1492–1496.
[42] P. Smets, Decision making in the tbm: the necessity of the pignistic transformation, Int. J. Approx. Reason. 38 (2) (2005) 133–147.
[43] T. Denœux, M.-H. Masson, Evclus: evidential clustering of proximity data, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 34 (1) (2004) 95–109.
[44] M.-H. Masson, T. Denœux, Ecm: an evidential version of the fuzzy c-means algorithm, Pattern Recognit. 41 (4) (2008) 1384–1397.
[45] Z. Zhang, Z. Liu, A. Martin, K. Zhou, Bsc: belief shift clustering, IEEE Trans. Syst. Man Cybern. Syst. 53 (3) (2022) 1748–1760.
[46] Z. Zhang, S. Ye, Y. Zhang, W. Ding, H. Wang, Belief combination of classifiers for incomplete data, IEEE/CAA J. Autom. Sin. 9 (4) (2022) 652–667.
[47] X. Jia, M. Zhao, Y. Di, Q. Yang, J. Lee, Assessment of data suitability for machine prognosis using maximum mean discrepancy, IEEE Trans. Ind. Electron. 65 (7) (2017) 5872–5881.
[48] A.S. Nemirovski, M.J. Todd, Interior-point methods for optimization, Acta Numer. 17 (2008) 191–234.
[49] J.A. Hartigan, M.A. Wong, A k-means clustering algorithm, Appl. Stat. 28 (1) (1979) 100–108.
[50] D.A. Pisner, D.M. Schnyer, Support vector machine, in: Machine Learning, Elsevier, 2020, pp. 101–121.
[51] M. Riedmiller, A. Lernen, Multi layer perceptron, in: Machine Learning Lab Special Lecture, Univ. Freiburg 24 (2014).
[52] T.S. Wong, Linear approximation of f-measure for the performance evaluation of classification algorithms on imbalanced data sets, IEEE Trans. Knowl. Data Eng. 34 (2) (2020) 753–763.
[53] M. Scott, J. Plested, Gan-smote: a generative adversarial network approach to synthetic minority oversampling, Aust. J. Intell. Inf. Process. Syst. 15 (2) (2019) 29–35.
[54] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao, Y. Gao, Kullback–Leibler divergence metric learning, IEEE Trans. Cybern. 52 (4) (2020) 2047–2058.