

Supplementary material

Manifold neighboring envelope sample generation mechanism for imbalanced ensemble classification

Table 1 shows the definitions of key concepts in this paper.

Table 1. Definition of the key concepts

Key concepts	Definition
Correlation information	Correlation among samples
Local correlation information	Correlation among neighboring samples
Global correlation information	Correlation among all samples in the sample set
Sample envelope	A set including some similar samples
Envelope sample	Sample envelope after sample transformation by mining correlation information

1. Joint Optimization

$$\begin{aligned}
 J_{MNESG}(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)}) &= \eta Loss_{local}(\mathbf{P}) + \gamma Loss_{global}(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)}) + \mu Loss_{consistency}(\mathbf{P}, \mathbf{V}^{(q)}) \\
 &= \eta \sum_{i=1}^{N'} \|\mathcal{M}_i - \mathbf{P} \mathbf{P}^T \mathcal{M}_i\|_2^2 + \gamma \sum_{i=1}^{N'} \sum_{j=1}^C (u_{ij})^m \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|^2 \\
 &\quad + \mu \left[\frac{1}{N'^2} \sum_{i=1}^{N'} \sum_{i'=1}^{N'} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{P}^T \mathcal{M}_{i'}) - \frac{2}{N' C} \sum_{i=1}^{N'} \sum_{j=1}^C \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) + \frac{1}{C^2} \sum_{j=1}^C \sum_{j'=1}^C \kappa(\mathbf{v}_j, \mathbf{v}_{j'}) \right] \\
 &\quad s.t. \mathbf{U} \mathbf{1} = \mathbf{1}, \quad \mathbf{U} \geq 0, \quad \mathbf{P}^T \mathbf{P} = \mathbf{I}
 \end{aligned} \tag{8}$$

In the MNESG model, there are three variables \mathbf{P} , \mathbf{U} and $\mathbf{V}^{(q)}$ to be optimized in (8). First, use MNSEP and FCM to complete the initialization of the three variables. Then, an effective alternating variable optimization strategy can be considered, i.e., to solve for one variable while fixing the rest of the variables as constants. Therefore, in solving (8), \mathbf{P} , \mathbf{U} and $\mathbf{V}^{(q)}$ can be solved in turn using the gradient descent method. Among them, η , γ , μ are three hyperparameters, which are optimized in this paper based on the grid search method.

1) Initialization of variables

Based on original subsets $\mathbf{X}^{(q)}$, MNSEP is used to initialize the projection vector \mathbf{P} and NES $\tilde{\mathbf{X}}^{(q)}$. The specific initialization steps are shown as follows.

$$\begin{aligned}
 Loss_{local}(\mathbf{P}) &= \sum_{i=1}^{N'} \|\mathcal{M}_i - \mathbf{P} \mathbf{P}^T \mathcal{M}_i\|_2^2 \\
 &= \sum_{i=1}^{N'} \mathcal{M}_i^T \mathcal{M}_i - 2 \sum_{i=1}^{N'} \mathcal{M}_i^T \mathbf{P} \mathbf{P}^T \mathcal{M}_i + \sum_{i=1}^{N'} \mathcal{M}_i^T \mathbf{P} \mathbf{P}^T \mathbf{P} \mathbf{P}^T \mathcal{M}_i \\
 &= \sum_{i=1}^{N'} \mathcal{M}_i^T \mathcal{M}_i - \sum_{i=1}^{N'} \mathcal{M}_i^T \mathbf{P} \mathbf{P}^T \mathcal{M}_i \\
 &= \sum_{i=1}^{N'} \mathcal{M}_i^T \mathcal{M}_i - \text{tr} \left[\mathbf{P}^T \left(\sum_{i=1}^{N'} \mathcal{M}_i \mathcal{M}_i^T \right) \mathbf{P} \right] \\
 &= \sum_{i=1}^{N'} \mathcal{M}_i^T \mathcal{M}_i - \text{tr} \left[\mathbf{P}^T \left(\mathcal{M}^T \right)^T \mathcal{M}^T \mathbf{P} \right]
 \end{aligned} \tag{9}$$

In Eq. (9), $\sum_{i=1}^{N'} \mathcal{M}_i^T \mathcal{M}$ is a constant, so minimizing Eq. (9) is equivalent to minimizing Eq. (10):

$$\begin{aligned} Loss_{local}(\mathbf{P}) &= -tr\left(\mathbf{P}^T \left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}} \mathbf{P}\right) \\ s.t. \mathbf{P}^T \mathbf{P} &= \mathbf{I} \end{aligned} \quad (10)$$

The objective function $Loss_{local}(\mathbf{P})$ can be optimized by the Lagrange multiplier method to obtain Eq. (11):

$$Loss_{local}(\mathbf{P}) = -tr\left[\mathbf{P}^T \left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}} \mathbf{P} + \boldsymbol{\zeta} (\mathbf{P}^T \mathbf{P} - \mathbf{I})\right] \quad (11)$$

Solve for the minimalist solution of Eq. (11) to obtain Eq. (12):

$$\left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}} \mathbf{P} = \boldsymbol{\zeta} \mathbf{P} \quad (12)$$

From Eq. (12), we can solve that \mathbf{P} is a matrix composed of the eigenvectors of $\left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}}$ and $\boldsymbol{\zeta}$ is a diagonal matrix composed of the eigenvalues of $\left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}}$. Therefore, when we reconstruct the sample envelope consisting of $k+1$ samples into a neighboring envelope sample by sample envelope projection reconstruction, we need to find the eigenvector \mathbf{P} corresponding to the largest eigenvalue of $\left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}}$ as the projection vector.

The whole process of the MNSEP algorithm is as follows.

Algorithm 1: MNSEP

Input: Original sample subset $\mathbf{X}^{(q)}$, Number of samples N' , Number of manifold nearest neighbors k .

Procedure:

- 1: Computing the neighborhood graph matrix \mathbf{G} ;
- 2: Computing the matrix of shortest paths to approximate the manifold distances $d_M(i, j)$;
- 3: For i -th sample, construct its sample envelope \mathcal{M}_i , and the transposed sample envelope \mathcal{M}_i^T ;
- 4: Repeat step 3(Construction of the manifold neighboring sample envelope) until all the samples are processed. After

that, the original samples are transformed into the centered sample envelope matrix $\mathcal{M}^{\hat{T}}$;

- 5: Based on the eigendecomposition of $\left(\mathcal{M}^{\hat{T}}\right)^T \mathcal{M}^{\hat{T}}$, the eigenvector corresponding to its largest eigenvalue is the projection vector \mathbf{P} .

- 6: The neighboring envelope samples $\tilde{\mathbf{X}}^{(q)}$ are obtained by step 4 based on $\mathcal{M}^{\hat{T}}$ and \mathbf{P} ;

- 7: Return \mathbf{P} and $\tilde{\mathbf{X}}^{(q)}$;
-

Output: Projection vector \mathbf{P} , Neighboring envelope sample subset $\tilde{\mathbf{X}}^{(q)}$.

Then, FCM is used to initialize partition matrix \mathbf{U} and NCES $\mathbf{V}^{(q)}$.

$$\begin{aligned} Loss_{global}(\mathbf{U}, \mathbf{V}^{(q)}) &= \sum_{i=1}^{N'} \sum_{j=1}^C (u_{ij})^m \|\tilde{\mathbf{x}}_i - \mathbf{v}_j\|^2 \\ s.t. \mathbf{U} \mathbf{1} &= \mathbf{1}, \mathbf{U} \geq 0 \end{aligned} \quad (13)$$

To minimize (13) by using Lagrange multiplier method, there is

$$\min J_1(\mathbf{U}, \mathbf{V}^{(q)}, \varepsilon) = \sum_{i=1}^{N'} \sum_{j=1}^C (u_{ij})^m \|\tilde{\mathbf{x}}_i - \mathbf{v}_j\|^2 + \varepsilon \left(1 - \sum_{j=1}^C u_{ij}\right) \quad (14)$$

Where ε is the Lagrange multiplier. To obtain the iterative fashion by setting the partial derivative of Eq. (14) with respect to $u_{ij}, \mathbf{v}_j, \varepsilon$ to be zero, there is

$$\begin{cases} \partial J_1(\mathbf{U}, \mathbf{V}^{(q)}, \varepsilon) / \partial u_{ij} = m u_{ij}^{m-1} \|\tilde{\mathbf{x}}_i - \mathbf{v}_j\|^2 - \varepsilon = 0 \\ \partial J_1(\mathbf{U}, \mathbf{V}^{(q)}, \varepsilon) / \partial \mathbf{v}_j = -2 \sum_{i=1}^{N'} u_{ij}^m (\tilde{\mathbf{x}}_i - \mathbf{v}_j) = 0 \\ \partial J_1(\mathbf{U}, \mathbf{V}^{(q)}, \varepsilon) / \partial \varepsilon = 1 - \sum_{j=1}^C u_{ij} = 0 \end{cases} \quad (15)$$

Through Eq. (15), the partition matrix and NCES are applied in an iterative fashion as follow.

$$u_{ij} = \frac{\|\tilde{\mathbf{x}}_i - \mathbf{v}_j\|^{-2/(m-1)}}{\sum_{p=1}^C \|\tilde{\mathbf{x}}_i - \mathbf{v}_p\|^{-2/(m-1)}}, \quad \mathbf{v}_j = \frac{\sum_{i=1}^{N'} (u_{ij})^m \tilde{\mathbf{x}}_i}{\sum_{i=1}^{N'} (u_{ij})^m} \quad (16)$$

Based on $\tilde{\mathbf{X}}^{(q)}$, the initialized \mathbf{U} and $\mathbf{V}^{(q)}$ are obtained by Eq. (16).

2) Fixing $\mathbf{V}^{(q)}$ and \mathbf{U} to solve \mathbf{P} .

By fixing $\mathbf{V}^{(q)}$ and \mathbf{U} , the problem is solved with respect to \mathbf{P} . After removing the terms unrelated to \mathbf{P} , the objective function (8) is transformed into Eq. (17).

$$\begin{aligned} \min_{\mathbf{P}} J_1(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)}) &= \eta \sum_{i=1}^{N'} \|\mathcal{M}_i - \mathbf{P} \mathbf{P}^T \mathcal{M}_i\|_2^2 + \gamma \sum_{i=1}^{N'} \sum_{j=1}^C (u_{ij})^m \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|^2 \\ &+ \mu \left[\frac{1}{N'^2} \sum_{i=1}^{N'} \sum_{i'=1}^{N'} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{P}^T \mathcal{M}_{i'}) - \frac{2}{N' C} \sum_{i=1}^{N'} \sum_{j=1}^C \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) \right] + \lambda (\mathbf{P}^T \mathbf{P} - \mathbf{I}) \end{aligned} \quad (17)$$

As shown in Eq. (17), it is difficult to obtain the closed-form solution of \mathbf{P} . Therefore, the gradient descent method is used to update \mathbf{P} . Then the iterative solution of \mathbf{P} can be expressed as Eq. (18).

$$\mathbf{P}_{K+1} = \mathbf{P}_K - \theta \cdot \nabla(\mathbf{P}) \quad (18)$$

$$\begin{aligned} \nabla(\mathbf{P}) &= -2\eta \sum_{i=1}^{N'} \left[(\mathcal{M}_i - \mathbf{P} \mathbf{P}^T \mathcal{M}_i) \mathcal{M}_i^T \mathbf{P} + \mathcal{M}_i (\mathcal{M}_i - \mathbf{P} \mathbf{P}^T \mathcal{M}_i)^T \mathbf{P} \right] \\ &+ 2 \sum_{i=1}^{N'} \sum_{j=1}^C \left[\gamma (u_{ij})^m + \frac{\mu}{N' C \sigma^2} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) \right] \mathcal{M}_i (\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j)^T \\ &- \frac{\mu}{N'^2 \sigma^2} \sum_{i=1}^{N'} \sum_{i'=1}^{N'} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{P}^T \mathcal{M}_{i'}) (\mathcal{M}_i - \mathcal{M}_{i'}) (\mathcal{M}_i - \mathcal{M}_{i'})^T \mathbf{P} + 2\lambda \mathbf{P} \end{aligned} \quad (19)$$

3) Fixing \mathbf{P} and $\mathbf{V}^{(q)}$ to solve \mathbf{U} .

By fixing \mathbf{P} and $\mathbf{V}^{(q)}$, the problem is solved with respect to \mathbf{U} . After removing the terms unrelated to \mathbf{U} , the objective function (8) is transformed into Eq. (20).

$$\min_{\mathbf{U}} J_3(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)}) = \gamma \sum_{i=1}^{N'} \sum_{j=1}^C (u_{ij})^m \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|_2^2 + \rho \left(\sum_{j=1}^C u_{ij} - 1 \right) \quad (20)$$

To the minimal value of Eq. (20), we set.

$$\frac{\partial J_3(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)})}{\partial u_{ij}} = m\gamma(u_{ij})^{m-1} \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|_2^2 + \rho = 0 \quad (21)$$

By calculation, the iterative formula of the partition matrix is obtained as follows.

$$u_{ij} = \frac{\left(1 / \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|_2^2\right)^{\frac{1}{m-1}}}{\sum_{w=1}^C \left(1 / \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_w\|_2^2\right)^{\frac{1}{m-1}}} \quad (22)$$

4) Fixing \mathbf{U} and \mathbf{P} to solve $\mathbf{V}^{(q)}$.

By fixing \mathbf{U} and \mathbf{P} , the problem is solved with respect to $\mathbf{V}^{(q)}$. After removing the terms unrelated to $\mathbf{V}^{(q)}$, the objective function (8) is transformed into Eq. (23).

$$\min_{\mathbf{V}^{(q)}} J_4(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)}) = \gamma \sum_{i=1}^C \sum_{j=1}^{N'} (u_{ij})^m \|\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j\|_2^2 + \mu \left[\frac{1}{C^2} \sum_{j=1}^C \sum_{j'=1}^C \kappa(\mathbf{v}_j, \mathbf{v}_{j'}) - \frac{2}{N'C} \sum_{i=1}^{N'} \sum_{j=1}^C \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) \right] \quad (23)$$

Based on the characteristic Gaussian kernel function $\kappa(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2)$, the minimal value of the objective function (23) is obtained by setting the partial derivative with respect to \mathbf{v}_j to be zero.

$$\frac{\partial J_4(\mathbf{P}, \mathbf{U}, \mathbf{V}^{(q)})}{\partial \mathbf{v}_j} = -2 \sum_{i=1}^{N'} \left[\gamma(u_{ij})^m + \frac{\mu}{N'C\sigma^2} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) \right] (\mathbf{P}^T \mathcal{M}_i - \mathbf{v}_j) + \frac{2\mu}{C^2\sigma^2} \sum_{j'=1}^C (\mathbf{v}_{j'} - \mathbf{v}_j) \kappa(\mathbf{v}_{j'}, \mathbf{v}_j) \quad (24)$$

Solve for the minimum solution of Eq. (24). $\mathbf{V}^{(q)}$ is obtained from Eq. (25).

$$\mathbf{V}^{(q)} = \mathbf{A}^{-1} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_C \end{bmatrix} \quad (25)$$

Among them:

$$\begin{aligned} \mathbf{A} &= \text{diag}(a_1, a_2, \dots, a_C) + \frac{\mu}{C^2\sigma^2} \hat{\mathbf{K}} \\ \hat{\mathbf{K}} &= \begin{bmatrix} \kappa(\mathbf{v}_1, \mathbf{v}_1) & \kappa(\mathbf{v}_2, \mathbf{v}_1) & \dots & \kappa(\mathbf{v}_C, \mathbf{v}_1) \\ \kappa(\mathbf{v}_1, \mathbf{v}_2) & \kappa(\mathbf{v}_2, \mathbf{v}_2) & \dots & \kappa(\mathbf{v}_C, \mathbf{v}_2) \\ \dots & \dots & \dots & \dots \\ \kappa(\mathbf{v}_1, \mathbf{v}_C) & \kappa(\mathbf{v}_2, \mathbf{v}_C) & \dots & \kappa(\mathbf{v}_C, \mathbf{v}_C) \end{bmatrix} \\ a_w &= \sum_{i=1}^{N'} \left[\gamma(u_{iw})^m + \frac{\mu}{N'C\sigma^2} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_w) \right] - \frac{\mu}{C^2\sigma^2} \sum_{j=1}^C \kappa(\mathbf{v}_j, \mathbf{v}_w), w=1, 2, \dots, C \\ b_j &= \sum_{i=1}^{N'} \left[\gamma(u_{ij})^m + \frac{\mu}{N'C\sigma^2} \kappa(\mathbf{P}^T \mathcal{M}_i, \mathbf{v}_j) \right] \mathbf{P}^T \mathcal{M}_i, j=1, 2, \dots, C \end{aligned}$$

The overall process of the MNESG algorithm is outlined as follows.

Algorithm 2: MNESG

Input: Original sample subset $\mathbf{X}^{(q)}$, Number of manifold nearest neighbors k , Number of clusters C , Iteration number t , Iteration threshold ε .

Procedure:

- 1: Based on the original sample subset $\mathbf{X}^{(q)}$, obtain the initialized neighboring envelope sample subset $\tilde{\mathbf{X}}^{(q)}$ and the initialized projection vector \mathbf{P} by Algorithm 1 (MNSEP);
 - 2: Initialize $\mathbf{V}^{(q)}$ and \mathbf{U} based on $\tilde{\mathbf{X}}^{(q)}$ by FCM algorithm;
 - 3: Optimize the \mathbf{P} , \mathbf{U} , $\mathbf{V}^{(q)}$ by Eqs. (18), (22), and (25), respectively until $\left|J_{MNESG}^{(t+1)} - J_{MNESG}^{(t)}\right| < \varepsilon$;
 - 4: Return final \mathbf{P} , \mathbf{U} , $\mathbf{V}^{(q)}$;
 - 5: Based on \mathbf{P} , obtain optimized $\tilde{\mathbf{X}}^{(q)}$;
-

Output: Neighboring envelope sample subset $\tilde{\mathbf{X}}^{(q)}$, Neighboring cluster envelope sample subset $\mathbf{V}^{(q)}$.

2. Experimental studies

To demonstrate the performance of the proposed algorithm (MNESG_IE), groups of experiments were conducted and analyzed. First, the experimental environment is introduced. Second, the effects of relevant parameters on the performance of the proposed algorithm are analyzed. Third, ablation experiments are conducted for verification. Finally, the proposed algorithm is compared with seven classical IE algorithms, four state-of-the-art IE algorithms and six advanced imbalanced classification algorithms.

2.1. Experimental environment

The decision tree C 4.5, which is well used in imbalanced ensemble learning, is chosen as the base classifier here. The 5-fold cross-validation (5-CV) method is chosen. To avoid randomness, each experiment is repeated 5 times and the mean and standard deviation of the values are reported.

2.1.1. Datasets

The 38 representative public datasets are chosen from the KEEL and UCI databases, which cover different domains, dimensions, samples size, and imbalance ratios (1.82-100.14). The major reasons why the datasets are chosen are as follows: 1) these datasets cover wide ranges of domains, dimensions size, sample size, imbalance ratio, and have been widely used in imbalance learning. 2) the focus of this paper is to propose an envelope sample generation method based on original samples. The datasets can directly provide structured samples for verification of different methods. Table 2 provides the basic information of these datasets.

Table 2. Basic information of imbalanced datasets

I D	Name	Feat ures	Sam ples	Min ority	Maj ority	Imba lance ratio	I D	Name	Feat ures	Sam ples	Min ority	Maj ority	Imba lance ratio
1	Glass1	9	214	76	138	1.82	20	Yeast-1-vs-7	8	459	30	429	14.30
2	Wisconsin	9	683	239	444	1.86	21	Glass4	9	214	13	201	15.47
3	Pima	8	768	268	500	1.87	22	Ecoli4	7	336	20	316	15.80
4	Iris0	4	150	50	100	2.00	2	Abalone9-18	8	731	42	689	16.40

For SMOTEBoost, SMOTEBagging, the number of neighbors is set to 3. Other parameters are default.

2.1.3. Evaluation metrics and non-parametric statistical tests

In this paper, we evaluate the performance of each method based on Accuracy (ACC), AUC, F-measure (F-M), and G-mean (G-M) criteria. These evaluation metrics are calculated as follows.

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

$$AUC = \frac{Sensitivity + Specificity}{2}$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

$$G - mean = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}}$$

Where TP denotes true positive, FP denotes false positive, TN denotes true negative and FN denotes false negative. In addition, sensitivity, specificity, recall, and precision are calculated as follows.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

To determine whether there is a significant difference between the algorithms, two kinds of nonparametric statistical tests are involved. (1) Multiple comparisons, based on the Friedman test with its corresponding post hoc test to determine whether there are significant differences between all comparison algorithms. In this paper, the Holm post hoc test was chosen, and the significance level was set at $\alpha = 0.05$. (2) Pairwise comparisons, wherein the Wilcoxon paired signed-rank test was used to determine whether there was a significant difference in the classification ability between the two algorithms. This was complemented by the ranking of all compared algorithms with respect to different evaluation metrics based on the Friedman aligned rank test, where a lower rank number indicates better classification ability.

2.2. Parameter analysis

In this section, the influences of two important parameters: the number of manifold nearest neighbors *MN-num* and the number of clustering centers *C-num* on the performance of MNESG_IE is studied. Six datasets which represented two types of datasets are chosen (e.g., exhibiting high-IR and low-IR). In addition, the optimization of three hyperparameters η , γ , μ is analyzed based on the grid search method.

2.2.1. Effect analysis of the number of manifold nearest neighbors

MN-num is the number of manifold nearest neighbors selected based on each sample when performing the MNSEP. In the proposed algorithm, *MN-num* means the number of samples in each

sample envelope. This will affect the structural information in the neighboring envelope samples, which in turn affects the classification performance and diversity of the base classifier trained on the subsets indirectly. To investigate the effect of $MN-num$ on the performance of MNESG_IE, six datasets with different imbalance ratios (1.86-28.10) are selected for parametric analysis at $MN-num = 0, 1, 2, 3, 4, 5$. Figure 5 shows the four evaluation metrics based on different $MN-num$ for different datasets.

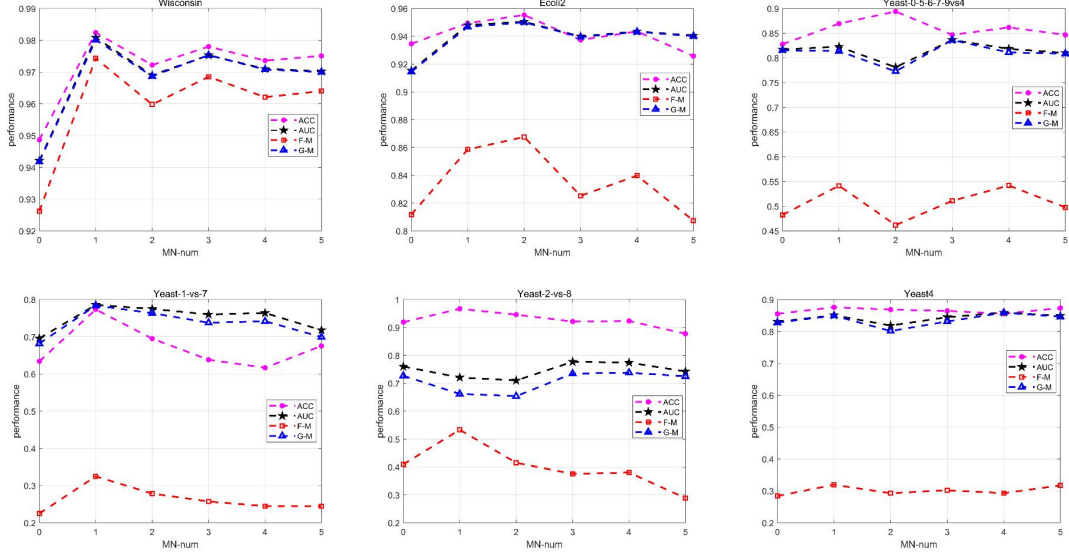


Fig. 5. MNESG_IE performance with different $MN-num$

As shown in Figure 5, when $MN-num$ changes from 0 to 1, each evaluation metric generally improves to a certain extent, which shows that the local correlation among neighboring samples is effective. The possible reason is that the envelope sample projection generation can effectively explore the local correlation among similar samples, thus generating high-quality NES. However, as $MN-num$ increases, the performance no changes or starts to decrease. Therefore, an excessive value of $MN-num$ is not suitable, probably because too many selected nearest neighbors increase the redundant information. Therefore, a reasonable value of $MN-num$ ranges from 1 to 3. To balance the accuracy and computational complexity, this paper sets $MN-num=1$.

2.2.2. Effect analysis of the number of clusters

$C-num$ is the ratio of the number of clusters to the number of samples before clustering when performing clustering. The smaller $C-num$ is, the more compact the mined correlation information will be, yet the larger the risk of missing useful correlation information. The reverse is also true.

To investigate the effect of $C-num$ on the performance of MNESG_IE, six datasets with different imbalance ratios (1.86-28.10) were selected at $C-num=30\%, 40\%, 50\%, 60\%, 70\%, 80\%$. Figure 6 shows the four evaluation metrics in terms of different $C-num$ and datasets.

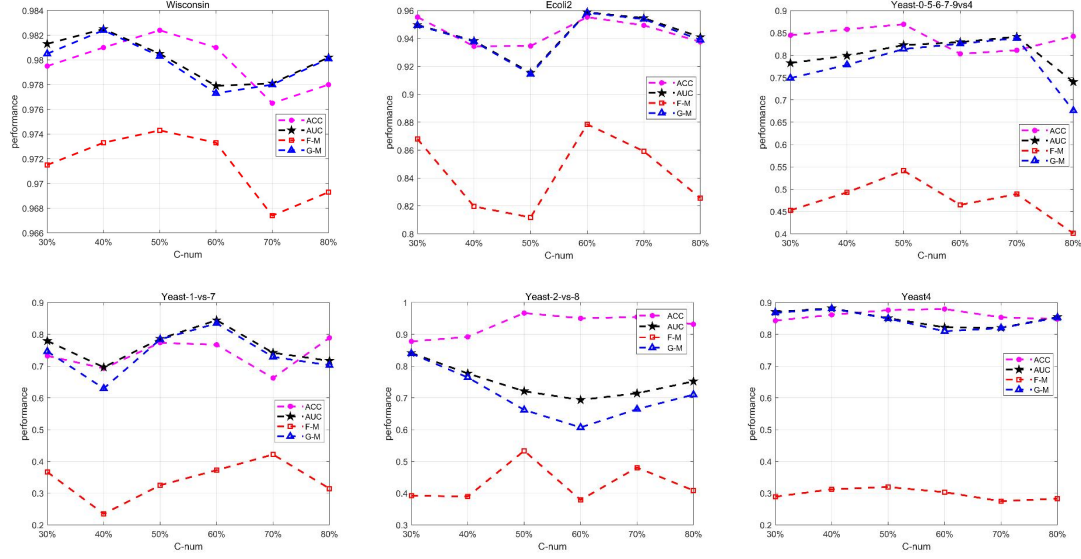


Fig. 6. MNESG_IE performance with different $C\text{-num}$

As shown in Figure 6, with the growth of $C\text{-num}$, the performance of the algorithm based on each evaluation metric tends to increase and then decrease, and the best performance is generally obtained when the $C\text{-num}$ is approximately 50%. The $C\text{-num}$ should not be too large or too small: if it is too large, some poor-quality neighboring envelope samples will be generated, and if it is too small, useful information may be lost. Therefore, a reasonable value of $C\text{-num}$ should be chosen from 40% to 60%. To balance the accuracy and computational complexity, this paper sets $C\text{-num}=50\%$.

2.2.3. Effect analysis of hyperparameters

The objective function of the proposed algorithm involves three hyperparameters η , γ and μ , and these three hyperparameters determine the contribution of the different loss items.

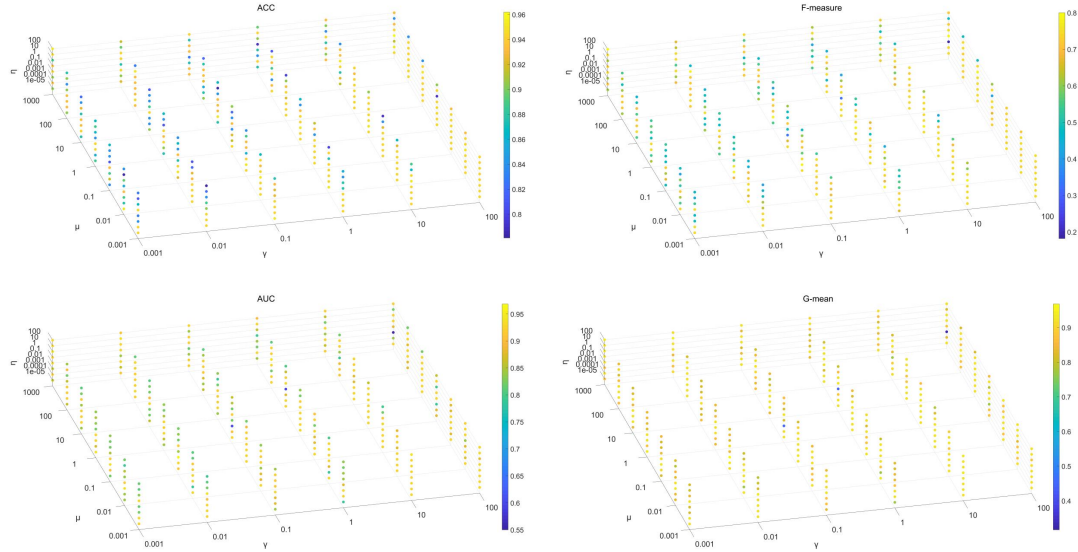


Fig. 7. MNESG_IE performance with different (η, γ, μ)

To analyze the impacts of these hyperparameters, the performance of MNESG_IE with different parameter value sets executed on Yeast-0-6-5-7-9-vs-4 is shown in Figure 7. The color of each point in

Figure 7 denotes the ACC, F-M, AUC, and G-M values for the corresponding parameter values (η, γ, μ) . It can be found in Figure 7 that better performance can be produced with relatively large values of γ and μ and a moderate value of η . The values of γ and μ of approximately 10 and the value of η of approximately 1 could be potential suitable ones.

2.3. Ablation study on major components

The major components are the two envelope samples construction mechanism (MNESG). To verify the effectiveness of the two types of the envelope samples-NES and NCES, ablation experiment was conducted on six datasets with different imbalance ratios (1.86-28.10). The ‘NES and NCES_based’ method and ‘NES_based’ method are compared with the ‘Original_based’ method. ‘Original_based’ means the original samples-based IE methods. ‘NES_based’ means the NES based IE method. ‘NES and NCES_based’ (MNESG_IE) means NES and NCES based IE method. The comparison of the above three methods is presented in Table 3. The best results are shown in boldface.

Table 3. Ablation results for the proposed method

Dataset	Algorithms	ACC	AUC	F-M	G-M
Wisconsin	‘Original_based’	0.9400±0.0400	0.9403±0.0377	0.9178±0.0524	0.9400±0.0377
	‘NES_based’	0.9765±0.0168	0.9770±0.0212	0.9666±0.0249	0.9769±0.0214
	‘NES and NCES_based’	0.9898±0.0083	0.9921±0.0064	0.9857±0.0115	0.9921±0.0065
Ecoli2	‘Original_based’	0.9314±0.0482	0.9118±0.0620	0.8079±0.1261	0.9095±0.0655
	‘NES_based’	0.9349±0.0367	0.9311±0.0512	0.8192±0.0927	0.9307±0.0514
	‘NES and NCES_based’	0.9761±0.0257	0.9538±0.0291	0.9242±0.0787	0.9532±0.0293
Yeast-0-5-6-7-9-vs-4	‘Original_based’	0.7899±0.0483	0.8497±0.0627	0.4642±0.0765	0.8453±0.0607
	‘NES_based’	0.7916±0.0803	0.8596±0.0622	0.4811±0.1021	0.8537±0.0647
	‘NES and NCES_based’	0.9263±0.0039	0.9235±0.0779	0.6977±0.0370	0.9184±0.0873
Yeast-1-vs-7	‘Original_based’	0.7321±0.0963	0.8102±0.0401	0.3179±0.0587	0.7973±0.0409
	‘NES_based’	0.7868±0.1449	0.7930±0.0604	0.3716±0.1134	0.7784±0.0749
	‘NES and NCES_based’	0.8110±0.0531	0.8214±0.0294	0.3732±0.0614	0.8210±0.0291
Yeast-2-vs-8	‘Original_based’	0.7907±0.0759	0.7952±0.0865	0.2543±0.0871	0.7857±0.0909
	‘NES_based’	0.9143±0.0274	0.7460±0.0657	0.3612±0.0908	0.7201±0.0788
	‘NES and NCES_based’	0.9792±0.0121	0.8696±0.0978	0.7532±0.1358	0.8547±0.1152
Yeast4	‘Original_based’	0.8294±0.0494	0.8653±0.0551	0.2738±0.0529	0.8607±0.0547
	‘NES_based’	0.8330±0.0611	0.8661±0.0506	0.2837±0.0652	0.8645±0.0510
	‘NES and NCES_based’	0.9097±0.0142	0.8209±0.0693	0.3553±0.0457	0.8122±0.0787

As shown in Table 3, in terms of the four evaluation metrics, ‘NES_based’ generally outperforms ‘Original_based’. The result means the NES containing local correlation information can improve the quality of the subsets and subsequent classification performance. The ‘NES and NCES_based’ achieves significant improvement compared with ‘NES_based’ for most datasets. The result indicates that the NCES containing both local and global correlation information is effective. At the same time, the performance of the ‘NES and NCES_based’ is significantly better than that of the ‘Original_based’ method in terms of all four evaluation metrics. It means that the two types of envelope samples containing correlation information are better than original samples in the subsets, which are more helpful for imbalanced classification.

In addition, the diversity and performance of the base classifiers in the algorithms are also analyzed based on the Kappa-error diagram. Figure 8 shows the diversity performance of the base classifiers obtained using MNESG_IE, SMOTE Bagging, and Under Bagging algorithms on three datasets with different imbalance ratios. Among them, ‘Neighboring samples based’ means that the base classifiers are trained on the NES. ‘Hierarchical samples based’ means that the base classifiers are

trained on the NCES.

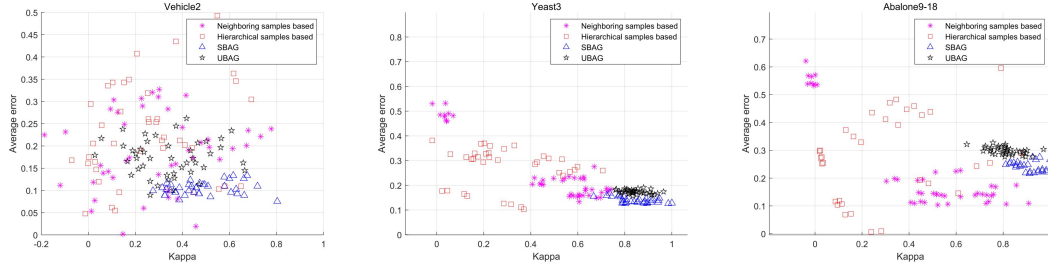


Fig. 8. Diversity and performance analysis of base classifiers

As shown in Figure 8, MNESG_IE can obtain data points with smaller Kappa values and average errors compared to the other two classical IE algorithms. This means that the base classifiers which are trained based on the envelope samples obtained by MNESG_IE have greater diversity and performance. Moreover, the Kappa values of data points obtained by ‘Hierarchical samples based’ are generally smaller than that of ‘Neighboring samples based’. The possible reason for this is that the NCES are obtained based on the NES. Therefore, NCES include both the local and global correlation information.

2.4. Algorithm comparison

To verify the effectiveness and competitiveness of the proposed algorithm, MNESG_IE is compared with seven classical IE algorithms, four state-of-the-art IE algorithms, and six advanced DL based imbalanced classification methods. For fair comparison, the results of the MNESG_IE are shown with the default and optimized parameters and compared with that of the compared methods. The best results are shown in boldface.

2.4.1. Comparison with classical IE algorithms

In this experiment, the proposed MNESG_IE was evaluated against seven classical IE methods: SMOTE Bagging, Under Bagging, SMOTE Boost, RUSBoost, EUSBoost, Balance Cascade, Easy Ensemble. The details and parameter settings of these methods are shown in subsection ‘Experimental environment’. The results are shown in Table 4, which lists the average ACC, AUC, F-M and G-M values for performing different IE algorithms based.

Table 4. Comparison with classical IE algorithms

ID	Algorithm	ACC	AUC	F-M	G-M
1	SBAG	0.7526±0.0872	0.7400±0.0747	0.6727±0.0763	0.7364±0.0728
	UBAG	0.7430±0.0882	0.7549±0.0817	0.6903±0.0807	0.7508±0.0826
	SBO	0.8354±0.0596	0.7839±0.0652	0.7232±0.0830	0.7767±0.0659
	RBO	0.7663±0.0717	0.7724±0.0864	0.7043±0.0911	0.7702±0.0838
	EBO	0.7860±0.0812	0.8013±0.0695	0.7440±0.0683	0.7979±0.0745
	BAC	0.7055±0.0693	0.7274±0.0738	0.6599±0.0703	0.7218±0.0741
	Easy	0.6903±0.0926	0.7127±0.0962	0.6468±0.0963	0.7079±0.0957
	MNESG_IE (default)	0.7614±0.0835	0.7757±0.0403	0.7130±0.0393	0.7543±0.0615
2	MNESG_IE	0.9721±0.0104	0.9662±0.0011	0.9599±0.0125	0.9656±0.0011
	SBAG	0.9648±0.0095	0.9611±0.0138	0.9501±0.0138	0.9609±0.0139
	UBAG	0.9582±0.0119	0.9609±0.0112	0.9445±0.0137	0.9607±0.0114
	SBO	0.9648±0.0141	0.9652±0.0145	0.9506±0.0196	0.9651±0.0145
	RBO	0.9736±0.0083	0.9739±0.0111	0.9627±0.0120	0.9738±0.0113

	EBO	0.9546±0.0141	0.9526±0.0154	0.9358±0.0194	0.9522±0.0157
	BAC	0.9657±0.0095	0.9473±0.0132	0.9658±0.0095	0.9619±0.0095
	Easy	0.9619±0.0119	0.9610±0.0117	0.9463±0.0166	0.9610±0.0117
	MNESG_IE (default)	0.9780±0.0052	0.9812±0.0048	0.9694±0.0070	0.9811±0.0048
	MNESG_IE	0.9898±0.0083	0.9921±0.0064	0.9857±0.0115	0.9921±0.0065
3	SBAG	0.7590±0.0158	0.7274±0.0155	0.6434±0.0212	0.7196±0.0165
	UBAG	0.7214±0.0408	0.7383±0.0352	0.6666±0.0356	0.7355±0.0363
	SBO	0.7408±0.0356	0.7291±0.0529	0.6469±0.0697	0.7257±0.0575
	RBO	0.7356±0.0309	0.7293±0.0337	0.6511±0.0414	0.7278±0.0342
	EBO	0.7792±0.0480	0.7566±0.0392	0.6852±0.0437	0.7541±0.0379
	BAC	0.6901±0.0307	0.7020±0.0256	0.6252±0.0277	0.6982±0.0258
	Easy	0.7143±0.0306	0.7124±0.0423	0.6344±0.0512	0.7114±0.0423
	MNESG_IE (default)	0.7435±0.0349	0.7220±0.0475	0.6223±0.1029	0.6945±0.0933
	MNESG_IE	0.8243±0.0302	0.7954±0.0343	0.7327±0.0494	0.7823±0.0460
4	SBAG	0.9866±0.0182	0.9800±0.0273	0.9789±0.0288	0.9794±0.0281
	UBAG	0.9866±0.0182	0.9800±0.0273	0.9789±0.0288	0.9794±0.0281
	SBO	0.9933±0.0149	0.9900±0.0223	0.9894±0.0235	0.9897±0.0229
	RBO	0.9933±0.0149	0.9900±0.0223	0.9894±0.0235	0.9897±0.0229
	EBO	0.9933±0.0149	0.9900±0.0223	0.9894±0.0235	0.9897±0.0229
	BAC	1±0	1±0	1±0	1±0
	Easy	0.9933±0.0149	0.9900±0.0223	0.9894±0.0235	0.9897±0.0229
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0
5	SBAG	0.7030±0.0246	0.6924±0.0239	0.5621±0.0347	0.6855±0.0331
	UBAG	0.7263±0.0239	0.7210±0.0347	0.5989±0.0418	0.7204±0.0351
	SBO	0.7183±0.0336	0.6946±0.0280	0.5648±0.0401	0.6865±0.0366
	RBO	0.7398±0.0253	0.7104±0.0400	0.5864±0.0522	0.7066±0.0434
	EBO	0.6913±0.0551	0.7081±0.0307	0.5857±0.0394	0.7032±0.0304
	BAC	0.6440±0.0433	0.6794±0.0080	0.5885±0.0456	0.6696±0.0192
	Easy	0.6543±0.0207	0.6752±0.0144	0.5481±0.0146	0.6728±0.0156
	MNESG_IE (default)	0.6914±0.0558	0.7028±0.0194	0.5788±0.0222	0.6957±0.0227
	MNESG_IE	0.7399±0.0516	0.7397±0.0239	0.6246±0.0357	0.7366±0.0211
6	SBAG	0.6532±0.0654	0.6304±0.0521	0.4698±0.0637	0.6208±0.0463
	UBAG	0.6598±0.0565	0.6422±0.0476	0.4861±0.0640	0.6188±0.0802
	SBO	0.6370±0.0473	0.6145±0.0540	0.4517±0.0698	0.6096±0.0572
	RBO	0.6575±0.0641	0.6639±0.0376	0.5076±0.0488	0.6575±0.0382
	EBO	0.7018±0.0427	0.6548±0.0647	0.4901±0.0849	0.6375±0.0719
	BAC	0.6172±0.0830	0.6210±0.0920	0.6581±0.1180	0.6189±0.0919
	Easy	0.6843±0.0758	0.6008±0.1185	0.3728±0.2543	0.5165±0.2490
	MNESG_IE (default)	0.7284±0.1128	0.6819±0.0402	0.5259±0.0744	0.6333±0.0653
	MNESG_IE	0.9905±0.0130	0.9947±0.0072	0.9556±0.0609	0.9947±0.0073
7	SBAG	0.9621±0.0130	0.9525±0.0245	0.9270±0.0261	0.9532±0.0244
	UBAG	0.9566±0.0179	0.9536±0.0082	0.9235±0.0320	0.9599±0.0175
	SBO	0.9657±0.0188	0.9661±0.0140	0.9328±0.0375	0.9660±0.0140

	RBO	0.9621±0.0184	0.9671±0.0206	0.9304±0.0340	0.9669±0.0206
	EBO	0.9704±0.0200	0.9681±0.0210	0.9442±0.0373	0.9680±0.0212
	BAC	0.9515±0.0153	0.9539±0.0050	0.9114±0.0241	0.9536±0.0051
	Easy	0.9455±0.0211	0.9559±0.0176	0.9031±0.0352	0.9555±0.0178
	MNESG_IE (default)	0.9468±0.0120	0.9447±0.0146	0.9011±0.0218	0.9446±0.0146
	MNESG_IE	0.9503±0.0456	0.9606±0.0313	0.9148±0.0694	0.9599±0.0325
8	SBAG	0.7677±0.0092	0.7618±0.0170	0.6190±0.0195	0.7613±0.0177
	UBAG	0.7411±0.0234	0.7768±0.0235	0.6213±0.0265	0.7720±0.0219
	SBO	0.7718±0.0287	0.7441±0.0348	0.6015±0.0423	0.7402±0.0386
	RBO	0.7564±0.0295	0.7493±0.0545	0.5993±0.0643	0.7464±0.0584
	EBO	0.7494±0.0343	0.7902±0.0315	0.6364±0.0377	0.7854±0.0312
	BAC	0.7055±0.0446	0.7360±0.0400	0.5765±0.0459	0.7314±0.0391
	Easy	0.7257±0.0351	0.7386±0.0360	0.5830±0.0420	0.7377±0.0362
	MNESG_IE (default)	0.7448±0.0397	0.7555±0.0376	0.6034±0.0442	0.7503±0.0401
	MNESG_IE	0.8120±0.0383	0.7870±0.0509	0.6601±0.0586	0.7785±0.0612
9	SBAG	0.9203±0.0487	0.9198±0.0466	0.8509±0.0872	0.9183±0.0465
	UBAG	0.8831±0.0520	0.8885±0.0503	0.7903±0.0863	0.8862±0.0491
	SBO	0.9392±0.0390	0.9185±0.0514	0.8730±0.0807	0.9171±0.0524
	RBO	0.9203±0.0428	0.9269±0.0324	0.8529±0.0740	0.9258±0.0322
	EBO	0.9161±0.0533	0.9101±0.0514	0.8414±0.0880	0.9078±0.0522
	BAC	0.9157±0.0359	0.8892±0.0517	0.8255±0.0715	0.9054±0.0435
	Easy	0.9219±0.0268	0.9171±0.0313	0.8536±0.0521	0.9159±0.0319
	MNESG_IE (default)	0.9628±0.0202	0.9548±0.0185	0.9254±0.0332	0.9539±0.0188
	MNESG_IE	0.9953±0.0104	0.9970±0.0068	0.9905±0.0213	0.9969±0.0068
10	SBAG	0.9337±0.0215	0.9359±0.0308	0.8703±0.0416	0.9355±0.0311
	UBAG	0.9349±0.0131	0.9523±0.0189	0.8769±0.0251	0.9516±0.0187
	SBO	0.9396±0.0252	0.9311±0.0385	0.8768±0.0519	0.9305±0.0389
	RBO	0.9550±0.0106	0.9618±0.0138	0.9108±0.0206	0.9617±0.0137
	EBO	0.9397±0.0234	0.9466±0.0279	0.8827±0.0440	0.9465±0.0278
	BAC	0.9255±0.0239	0.9356±0.0223	0.8588±0.0393	0.9353±0.0225
	Easy	0.9302±0.0224	0.9370±0.0312	0.8652±0.0418	0.9365±0.0315
	MNESG_IE (default)	0.8208±0.0199	0.9007±0.0545	0.8336±0.0542	0.8966±0.0618
	MNESG_IE	0.9657±0.0190	0.9655±0.0118	0.9311±0.0349	0.9652±0.0119
11	SBAG	0.8814±0.0376	0.8998±0.0196	0.7865±0.0489	0.8975±0.0187
	UBAG	0.8721±0.0540	0.8990±0.0501	0.7777±0.0761	0.8961±0.0510
	SBO	0.8840±0.0436	0.8474±0.0587	0.7562±0.0916	0.8436±0.0621
	RBO	0.8839±0.0266	0.9068±0.0290	0.7908±0.0348	0.9042±0.0289
	EBO	0.8750±0.0734	0.8918±0.0578	0.7820±0.0953	0.8878±0.0616
	BAC	0.8600±0.0584	0.8637±0.0439	0.7475±0.0773	0.8616±0.0434
	Easy	0.8481±0.0611	0.8698±0.0564	0.7383±0.0855	0.8666±0.0570
	MNESG_IE (default)	0.9139±0.0504	0.9070±0.0425	0.8331±0.0843	0.9046±0.0453
	MNESG_IE	0.9910±0.0082	0.9800±0.0183	0.9793±0.0189	0.9797±0.0186
12	SBAG	0.9168±0.0379	0.8926±0.0721	0.7606±0.1042	0.8870±0.0744
	UBAG	0.8929±0.0408	0.8899±0.0353	0.7247±0.0834	0.8897±0.0353

	SBO	0.9346±0.0301	0.9090±0.0786	0.8023±0.0962	0.9054±0.0842
	RBO	0.9046±0.0252	0.8905±0.0332	0.7408±0.0364	0.8878±0.0356
	EBO	0.8990±0.0570	0.8862±0.0639	0.7357±0.1204	0.8857±0.0645
	BAC	0.8539±0.0376	0.8670±0.0341	0.6552±0.0653	0.8653±0.0332
	Easy	0.8212±0.0427	0.8485±0.0298	0.6095±0.0525	0.8459±0.0284
	MNESG_IE (default)	0.9316±0.0266	0.9202±0.0526	0.8035±0.0748	0.9191±0.0534
	MNESG_IE	0.9761±0.0257	0.9538±0.0291	0.9242±0.0787	0.9532±0.0293
13	SBAG	0.9346±0.0191	0.8923±0.0965	0.7657±0.1005	0.8836±0.1123
	UBAG	0.8972±0.0421	0.9159±0.0576	0.7159±0.0953	0.9115±0.0553
	SBO	0.9345±0.0106	0.8504±0.0815	0.7447±0.0615	0.8353±0.0970
	RBO	0.9108±0.0518	0.9227±0.0342	0.7507±0.1082	0.9182±0.0340
	EBO	0.8877±0.0894	0.8932±0.0700	0.7121±0.1598	0.8912±0.0701
	BAC	0.8875±0.0395	0.8932±0.0359	0.6893±0.0791	0.8913±0.0352
	Easy	0.8550±0.0387	0.8463±0.0935	0.6050±0.1065	0.8390±0.1036
	MNESG_IE (default)	0.9673±0.0126	0.9359±0.0408	0.8823±0.0369	0.9333±0.0429
	MNESG_IE	0.9766±0.0233	0.9586±0.0709	0.9132±0.0881	0.9551±0.0783
14	SBAG	0.9413±0.0218	0.9401±0.0197	0.7822±0.0661	0.9401±0.0197
	UBAG	0.9279±0.0258	0.9353±0.0155	0.7472±0.0673	0.9350±0.0156
	SBO	0.9386±0.0086	0.8795±0.0148	0.7427±0.0290	0.8761±0.0157
	RBO	0.9225±0.0263	0.9188±0.0240	0.7260±0.0698	0.9188±0.0240
	EBO	0.9198±0.0252	0.9281±0.0166	0.7244±0.0597	0.9277±0.0168
	BAC	0.9076±0.0257	0.9079±0.0316	0.6871±0.0697	0.9079±0.0316
	Easy	0.9130±0.0222	0.9109±0.0274	0.6989±0.0580	0.9108±0.0274
	MNESG_IE (default)	0.9373±0.0182	0.8819±0.0776	0.7349±0.0974	0.8753±0.0897
	MNESG_IE	0.9905±0.0130	0.9947±0.0072	0.9556±0.0609	0.9947±0.0073
15	SBAG	0.8898±0.0227	0.8754±0.0467	0.6197±0.0525	0.8736±0.0481
	UBAG	0.8424±0.0520	0.8994±0.0386	0.5710±0.0843	0.8955±0.0396
	SBO	0.9107±0.0179	0.8744±0.0735	0.6563±0.0814	0.8693±0.0823
	RBO	0.8720±0.0522	0.8402±0.0332	0.5778±0.0686	0.8346±0.0337
	EBO	0.8621±0.0174	0.8605±0.0746	0.5621±0.0549	0.8552±0.0810
	BAC	0.8090±0.1216	0.8112±0.0749	0.4820±0.1089	0.7996±0.0831
	Easy	0.8451±0.0822	0.8504±0.0463	0.5544±0.0845	0.8439±0.0493
	MNESG_IE (default)	0.8958±0.0153	0.9166±0.0357	0.6539±0.0376	0.9154±0.0350
	MNESG_IE	0.9628±0.0584	0.9442±0.0516	0.8806±0.1533	0.9413±0.0544
16	SBAG	0.9163±0.0108	0.8910±0.0753	0.6664±0.0600	0.8861±0.0830
	UBAG	0.9124±0.0305	0.9335±0.0229	0.6914±0.0662	0.9324±0.0229
	SBO	0.9435±0.0187	0.8893±0.0455	0.7428±0.0829	0.9058±0.0581
	RBO	0.9299±0.0371	0.9263±0.0296	0.7342±0.0819	0.9247±0.0309
	EBO	0.9144±0.0510	0.9267±0.0196	0.7024±0.1029	0.9253±0.0209
	BAC	0.9066±0.0520	0.8866±0.0358	0.6632±0.0875	0.8827±0.0396
	Easy	0.9046±0.0221	0.9212±0.0433	0.6642±0.0618	0.9199±0.0435
	MNESG_IE (default)	0.9590±0.0273	0.9068±0.0702	0.8080±0.1153	0.9006±0.0796
	MNESG_IE	0.9845±0.0128	0.9833±0.0251	0.9286±0.0532	0.9832±0.0253
17	SBAG	0.8541±0.0340	0.8217±0.0822	0.5089±0.1030	0.8165±0.0885

	UBAG	0.7916±0.0179	0.7969±0.0385	0.4271±0.0403	0.7963±0.0386
	SBO	0.8900±0.0321	0.7727±0.0907	0.5235±0.1359	0.7523±0.1114
	RBO	0.8351±0.0295	0.7862±0.0899	0.4581±0.0987	0.7788±0.0982
	EBO	0.8180±0.0324	0.8206±0.0471	0.4692±0.0679	0.8197±0.0473
	BAC	0.7557±0.0313	0.7511±0.0658	0.3686±0.0432	0.7456±0.0636
	Easy	0.8047±0.0516	0.7801±0.0640	0.4298±0.0675	0.7757±0.0693
	MNESG_IE (default)	0.8146±0.0478	0.8176±0.0934	0.4618±0.0965	0.8112±0.1015
	MNESG_IE	0.9263±0.0039	0.9235±0.0779	0.6977±0.0370	0.9184±0.0873
18	SBAG	0.8331±0.0475	0.6676±0.1698	0.4020±0.1666	0.5605±0.3434
	UBAG	0.6667±0.0261	0.7035±0.1160	0.2771±0.0862	0.6863±0.1213
	SBO	0.8491±0.0423	0.6014±0.1304	0.3234±0.1658	0.4569±0.2871
	RBO	0.7809±0.0607	0.6999±0.1416	0.3196±0.1500	0.6692±0.1733
	EBO	0.7182±0.0896	0.7930±0.1055	0.3669±0.1032	0.7851±0.1015
	BAC	0.6300±0.1074	0.6254±0.1762	0.2199±0.0984	0.5965±0.1821
	Easy	0.6345±0.1013	0.7390±0.1278	0.3066±0.1166	0.7259±0.1226
	MNESG_IE (default)	0.8961±0.0019	0.6244±0.0194	0.3333±0.0006	0.5267±0.0479
	MNESG_IE	0.9795±0.0115	0.9000±0.0559	0.8857±0.0639	0.8928±0.0599
19	SBAG	0.8174±0.0518	0.7019±0.1399	0.3847±0.1180	0.6077±0.3421
	UBAG	0.5888±0.0579	0.7766±0.0316	0.2796±0.0375	0.7428±0.0422
	SBO	0.8594±0.0585	0.7330±0.0945	0.4141±0.1420	0.7114±0.1150
	RBO	0.7662±0.1562	0.6375±0.1506	0.3190±0.2350	0.6110±0.1610
	EBO	0.7053±0.0626	0.7646±0.0711	0.3129±0.0714	0.7569±0.0638
	BAC	0.5750±0.1166	0.6174±0.1515	0.2064±0.0854	0.6134±0.1478
	Easy	0.5933±0.1077	0.7795±0.0565	0.2903±0.0845	0.7445±0.0766
	MNESG_IE (default)	0.6957±0.1946	0.7595±0.0043	0.3018±0.0874	0.7378±0.0187
	MNESG_IE	0.9814±0.0104	0.8607±0.0745	0.8400±0.0894	0.8532±0.0821
20	SBAG	0.8237±0.0556	0.7208±0.0367	0.3174±0.0480	0.7074±0.0455
	UBAG	0.7363±0.0262	0.6883±0.0798	0.2361±0.0525	0.6778±0.0981
	SBO	0.8065±0.0742	0.7259±0.0319	0.3142±0.0662	0.7170±0.0356
	RBO	0.8257±0.0558	0.7827±0.0943	0.3601±0.0813	0.7749±0.1005
	EBO	0.7996±0.0336	0.7222±0.1005	0.2946±0.0899	0.7105±0.1192
	BAC	0.6404±0.0785	0.6526±0.1016	0.1974±0.0634	0.6425±0.1109
	Easy	0.6817±0.0821	0.7212±0.0346	0.2425±0.0201	0.7113±0.0416
	MNESG_IE (default)	0.6455±0.1230	0.8023±0.0658	0.2690±0.0662	0.7753±0.0848
	MNESG_IE	0.8110±0.0531	0.8214±0.0294	0.3732±0.0614	0.8210±0.0291
21	SBAG	0.9060±0.0610	0.9025±0.0885	0.5733±0.1382	0.8921±0.1066
	UBAG	0.8773±0.0577	0.8593±0.1031	0.4666±0.1027	0.8487±0.1153
	SBO	0.9343±0.0355	0.8866±0.1011	0.6189±0.1166	0.8749±0.1178
	RBO	0.9018±0.0665	0.9004±0.1392	0.5563±0.2347	0.8962±0.1469
	EBO	0.8975±0.1390	0.8666±0.1520	0.6187±0.2352	0.8578±0.1610
	BAC	0.8036±0.0881	0.8480±0.1249	0.3915±0.2176	0.8393±0.1301
	Easy	0.8693±0.0706	0.8208±0.1130	0.4389±0.1094	0.8108±0.1214
	MNESG_IE (default)	0.9765±0.0001	0.9104±0.1090	0.8000±0.0001	0.9020±0.1209
	MNESG_IE	0.9902±0.0134	0.9949±0.0070	0.9200±0.1095	0.9948±0.0071

22	SBAG	0.9524±0.0193	0.8810±0.0497	0.7463±0.1452	0.8753±0.0523
	UBAG	0.8363±0.0802	0.8427±0.0773	0.4118±0.1298	0.8394±0.0756
	SBO	0.9582±0.0164	0.8373±0.0964	0.6666±0.1020	0.8177±0.1151
	RBO	0.8927±0.0867	0.8961±0.0714	0.5680±0.2540	0.8920±0.0737
	EBO	0.8600±0.1016	0.9021±0.1029	0.5121±0.2370	0.9005±0.1026
	BAC	0.8539±0.0492	0.8989±0.0593	0.4491±0.1020	0.8956±0.0596
	Easy	0.8869±0.0248	0.8930±0.0531	0.4888±0.0248	0.8897±0.0541
	MNESG_IE (default)	0.9045±0.0977	0.9258±0.0542	0.6436±0.2550	0.9214±0.0582
	MNESG_IE	0.9911±0.0081	0.9953±0.0043	0.9333±0.0609	0.9952±0.0043
23	SBAG	0.8617±0.0601	0.7257±0.1537	0.3392±0.1904	0.6932±0.1850
	UBAG	0.7369±0.0566	0.7819±0.0526	0.2755±0.0453	0.7770±0.0536
	SBO	0.9137±0.0458	0.7533±0.1430	0.4560±0.2876	0.7149±0.1897
	RBO	0.8002±0.0353	0.7906±0.0992	0.3146±0.0807	0.7834±0.1052
	EBO	0.7523±0.0597	0.7326±0.0941	0.2504±0.0572	0.7196±0.1073
	BAC	0.7358±0.0713	0.7502±0.1094	0.2628±0.0932	0.7463±0.1094
	Easy	0.6880±0.0665	0.7220±0.0595	0.2257±0.0444	0.7165±0.0586
	MNESG_IE (default)	0.9110±0.0581	0.8057±0.0724	0.4882±0.2028	0.7968±0.0756
	MNESG_IE	0.9623±0.0048	0.9212±0.0807	0.7166±0.0151	0.9173±0.0859
24	SBAG	0.9221±0.0109	0.9558±0.0077	0.6334±0.0383	0.9551±0.0076
	UBAG	0.9384±0.0842	0.9473±0.0447	0.7666±0.3248	0.9658±0.0467
	SBO	0.9340±0.0241	0.9114±0.0904	0.6269±0.0362	0.9027±0.1095
	RBO	0.9692±0.0688	0.9840±0.0357	0.8666±0.2981	0.9833±0.0373
	EBO	1±0	1±0	1±0	1±0
	BAC	1±0	1±0	1±0	1±0
	Easy	1±0	1±0	1±0	1±0
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0
25	SBAG	0.9488±0.0382	0.9256±0.0998	0.6619±0.1980	0.9164±0.1185
	UBAG	0.9495±0.0310	0.5266±0.2712	0.9512±0.0298	0.9065±0.0569
	SBO	0.9673±0.0126	0.9353±0.1037	0.6933±0.0596	0.9266±0.1228
	RBO	0.8976±0.0879	0.9463±0.0461	0.5815±0.3015	0.9438±0.0487
	EBO	0.9297±0.0572	0.9634±0.0298	0.6066±0.2832	0.9623±0.0310
	BAC	0.9018±0.0476	0.9487±0.0249	0.4866±0.1849	0.9471±0.0261
	Easy	0.9390±0.0538	0.9682±0.0280	0.6266±0.2385	0.9674±0.0292
	MNESG_IE (default)	0.9884±0.0164	0.9939±0.0086	0.9000±0.1414	0.9939±0.0087
	MNESG_IE	0.9884±0.0164	0.9939±0.0086	0.9000±0.1414	0.9939±0.0087
26	SBAG	0.9584±0.0181	0.8109±0.1566	0.5473±0.1410	0.7697±0.2034
	UBAG	0.7590±0.1437	0.7307±0.1349	0.2327±0.1257	0.7215±0.1378
	SBO	0.9646±0.0174	0.7902±0.1070	0.5833±0.1666	0.7528±0.1564
	RBO	0.9335±0.0059	0.8457±0.0030	0.4846±0.0210	0.8403±0.0027
	EBO	0.7820±0.0375	0.7666±0.0844	0.2238±0.0514	0.7613±0.0878
	BAC	0.6823±0.0972	0.7625±0.1149	0.1910±0.0758	0.7485±0.1142
	Easy	0.7637±0.0961	0.7571±0.1331	0.2214±0.0736	0.7471±0.1340
	MNESG_IE (default)	0.8961±0.0927	0.7066±0.1603	0.3026±0.0901	0.6246±0.2253

	MNESG_IE	0.9792±0.0121	0.8696±0.0978	0.7532±0.1358	0.8547±0.1152
	SBAG	0.7978±0.0482	0.8339±0.0374	0.2345±0.0278	0.8307±0.0327
	UBAG	0.7742±0.0275	0.8452±0.0970	0.2097±0.0219	0.8179±0.0717
	SBO	0.8215±0.0885	0.7979±0.0779	0.2466±0.0488	0.7835±0.0916
	RBO	0.8200±0.0347	0.8199±0.0472	0.2398±0.0170	0.8165±0.0462
27	EBO	0.8126±0.0403	0.8372±0.0611	0.2437±0.0397	0.8340±0.0594
	BAC	0.7850±0.0532	0.7747±0.0966	0.2031±0.0592	0.7710±0.1002
	Easy	0.7998±0.0588	0.8191±0.0865	0.2337±0.0750	0.8159±0.0880
	MNESG_IE (default)	0.8653±0.0190	0.8579±0.0243	0.2990±0.0123	0.8573±0.0247
	MNESG_IE	0.9097±0.0142	0.8209±0.0693	0.3553±0.0457	0.8122±0.0781
	SBAG	0.8930±0.0156	0.6523±0.0732	0.1970±0.0716	0.5921±0.1039
	UBAG	0.6897±0.0454	0.6666±0.0750	0.1206±0.0285	0.6600±0.0797
	SBO	0.9392±0.0047	0.5867±0.0582	0.2180±0.0403	0.3999±0.2266
	RBO	0.8116±0.0324	0.6228±0.0453	0.1330±0.0335	0.5879±0.0552
28	EBO	0.7535±0.0876	0.6110±0.0788	0.1051±0.0124	0.5630±0.1616
	BAC	0.6113±0.0831	0.6500±0.0791	0.1021±0.0186	0.6335±0.0727
	Easy	0.6611±0.1143	0.6403±0.0669	0.1083±0.0217	0.6206±0.0824
	MNESG_IE (default)	0.6953±0.1522	0.7326±0.0554	0.1584±0.0331	0.7074±0.0423
	MNESG_IE	0.9677±0.0018	0.6034±0.1003	0.2699±0.1787	0.4239±0.2120
	SBAG	0.8253±0.0177	0.6711±0.1485	0.1498±0.0788	0.6200±0.2011
	UBAG	0.7285±0.0661	0.7147±0.0517	0.1423±0.0212	0.7089±0.0561
	SBO	0.8171±0.0096	0.7299±0.1313	0.1851±0.0714	0.7065±0.1625
	RBO	0.8456±0.0239	0.7389±0.0896	0.1992±0.0365	0.7191±0.1208
29	EBO	0.7696±0.0990	0.6714±0.0611	0.1440±0.0471	0.6537±0.0801
	BAC	0.6472±0.0484	0.6727±0.0260	0.1122±0.0093	0.6702±0.0229
	Easy	0.5957±0.1497	0.6783±0.0562	0.1130±0.0231	0.6591±0.0712
	MNESG_IE (default)	0.6943±0.1227	0.7919±0.0580	0.1916±0.1012	0.7826±0.0852
	MNESG_IE	0.8468±0.0976	0.8000±0.0066	0.2628±0.0997	0.7954±0.0137
	SBAG	0.9703±0.0122	0.9618±0.0320	0.6669±0.0925	0.9612±0.0325
	UBAG	0.9393±0.0189	0.9458±0.0293	0.4917±0.0811	0.9451±0.0294
	SBO	0.9757±0.0087	0.9444±0.0555	0.7018±0.0541	0.9418±0.0589
	RBO	0.9440±0.1475	0.9604±0.0282	0.5158±0.0678	0.9600±0.0282
30	EBO	0.9602±0.0096	0.9458±0.0308	0.5849±0.0624	0.9452±0.0312
	BAC	0.9333±0.0197	0.9534±0.0220	0.4736±0.0642	0.9526±0.0221
	Easy	0.9198±0.0191	0.9479±0.0223	0.4253±0.0464	0.9470±0.0222
	MNESG_IE (default)	0.9158±0.0381	0.9566±0.0196	0.4342±0.1065	0.9554±0.0207
	MNESG_IE	0.9636±0.0140	0.9597±0.0291	0.6231±0.0788	0.9592±0.0295
	SBAG	0.9521±0.0114	0.8360±0.1076	0.4103±0.0799	0.8179±0.1288
	UBAG	0.8679±0.0335	0.8766±0.0594	0.2466±0.0502	0.8746±0.0593
	SBO	0.9676±0.0084	0.7743±0.1489	0.4407±0.1667	0.7252±0.1973
31	RBO	0.8746±0.0377	0.8382±0.0795	0.2353±0.0390	0.8307±0.0848
	EBO	0.8712±0.0517	0.8225±0.1086	0.2206±0.0575	0.8135±0.1236
	BAC	0.7776±0.0334	0.8162±0.0426	0.1548±0.0158	0.8135±0.0405
	Easy	0.7755±0.0493	0.8293±0.0513	0.1603±0.0266	0.8246±0.0488

	MNESG_IE (default)	0.8814±0.0237	0.8974±0.0921	0.2685±0.0570	0.8921±0.0994
	MNESG_IE	0.9137±0.0173	0.8868±0.0170	0.3234±0.0472	0.8861±0.0175
32	SBAG	0.9333±0.0143	0.6605±0.0675	0.1980±0.0557	0.5865±0.1133
	UBAG	0.7411±0.0529	0.7698±0.0618	0.1237±0.0255	0.7674±0.0581
	SBO	0.9518±0.0032	0.7310±0.1207	0.3025±0.1000	0.6772±0.1771
	RBO	0.7866±0.0873	0.7198±0.1365	0.1347±0.0713	0.7101±0.1409
	EBO	0.7688±0.0550	0.7107±0.1565	0.1176±0.0589	0.6940±0.1782
	BAC	0.7069±0.0517	0.7279±0.1075	0.1043±0.0323	0.7223±0.1054
	Easy	0.5966±0.1318	0.6715±0.0420	0.0779±0.0115	0.6521±0.0527
	MNESG_IE (default)	0.8694±0.0904	0.8111±0.0462	0.2398±0.1322	0.8082±0.0429
	MNESG_IE	0.9889±0.0045	0.8111±0.0722	0.7113±0.1216	0.7855±0.0916
33	SBAG	0.9005±0.0354	0.7349±0.0847	0.1937±0.0712	0.7054±0.1165
	UBAG	0.6596±0.0886	0.7446±0.0555	0.0956±0.230	0.7323±0.0490
	SBO	0.9502±0.0139	0.7298±0.1264	0.2955±0.1582	0.6764±0.1776
	RBO	0.8070±0.0423	0.7139±0.1216	0.1151±0.0452	0.6886±0.1536
	EBO	0.6830±0.1012	0.6750±0.0523	0.0813±0.0108	0.6610±0.0447
	BAC	0.6783±0.1094	0.7625±0.1077	0.1043±0.0379	0.7467±0.1106
	Easy	0.6549±0.0726	0.6933±0.1589	0.0819±0.0382	0.6748±0.1727
	MNESG_IE (default)	0.6932±0.1517	0.6599±0.0788	0.1222±0.0629	0.5730±0.1182
	MNESG_IE	0.8302±0.1735	0.9136±0.0082	0.2679±0.2273	0.9069±0.0974
34	SBAG	0.9294±0.0140	0.9564±0.0120	0.6436±0.0659	0.9562±0.0118
	UBAG	0.8376±0.0610	0.8779±0.0683	0.1802±0.0277	0.8704±0.0744
	SBO	0.9374±0.0122	0.9177±0.0611	0.6434±0.0661	0.9151±0.0642
	RBO	0.8945±0.0467	0.9462±0.0238	0.2836±0.0897	0.9444±0.0252
	EBO	0.9095±0.0337	0.9539±0.0172	0.3081±0.0860	0.9526±0.0180
	BAC	0.8362±0.0679	0.9165±0.0347	0.1970±0.0545	0.9128±0.0391
	Easy	0.8602±0.0333	0.8732±0.0909	0.1979±0.0683	0.8693±0.0940
	MNESG_IE (default)	0.9534±0.0093	0.9436±0.0745	0.4269±0.0647	0.9408±0.0801
	MNESG_IE	0.9932±0.0080	0.9965±0.0041	0.8684±0.0239	0.9965±0.0041
35	SBAG	1±0	1±0	1±0	1±0
	UBAG	1±0	1±0	1±0	1±0
	SBO	1±0	1±0	1±0	1±0
	RBO	0.9834±0.0371	0.9915±0.0188	0.8533±0.3279	0.9913±0.0192
	EBO	1±0	1±0	1±0	1±0
	BAC	1±0	1±0	1±0	1±0
	Easy	1±0	1±0	1±0	1±0
	MNESG_IE (default)	0.9970±0.0024	0.9262±0.0631	0.8895±0.0890	0.9212±0.0697
	MNESG_IE	1±0	1±0	1±0	1±0
36	SBAG	0.9996±0.0010	0.9833±0.0373	0.9818±0.0407	0.9826±0.0390
	UBAG	0.9861±0.0168	0.9929±0.0085	0.7471±0.2704	0.9929±0.0086
	SBO	0.9986±0.0020	0.9828±0.0370	0.9532±0.0666	0.9821±0.0387
	RBO	0.9605±0.0232	0.9800±0.0117	0.4848±0.2898	0.9797±0.0119
	EBO	0.9852±0.0201	0.9760±0.0345	0.7387±0.2556	0.9752±0.0361
	BAC	0.9771±0.0223	0.9884±0.0113	0.6533±0.3207	0.9882±0.0114

	Easy	0.9520±0.0078	0.9757±0.0039	0.3622±0.0357	0.9754±0.0041
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0
37	SBAG	0.9694±0.0142	0.9570±0.0492	0.7258±0.0399	0.9553±0.0523
	UBAG	0.9334±0.0196	0.9662±0.0099	0.2830±0.0757	0.9656±0.0102
	SBO	0.9660±0.0153	0.9462±0.0542	0.7178±0.0448	0.9439±0.0577
	RBO	0.9402±0.0347	0.9697±0.0176	0.3391±0.1441	0.9691±0.0181
	EBO	0.9544±0.0340	0.9769±0.0172	0.4431±0.2470	0.9765±0.0176
	BAC	0.9366±0.0292	0.9679±0.0148	0.3257±0.1712	0.9672±0.0151
	Easy	0.9357±0.0235	0.9674±0.0119	0.2950±0.0914	0.9668±0.0122
	MNESG_IE (default)	0.9973±0.0019	0.9788±0.0441	0.9005±0.0651	0.9777±0.0466
	MNESG_IE	0.9995±0.0010	0.9833±0.0622	0.9818±0.0407	0.9826±0.0390
38	SBAG	0.9854±0.0644	0.9543±0.0621	0.8769±0.1315	0.9525±0.0644
	UBAG	0.9825±0.0279	0.9115±0.0506	0.8420±0.1243	0.9073±0.0528
	SBO	0.9858±0.0290	0.9545±0.0623	0.8987±0.1244	0.9528±0.0646
	RBO	0.9818±0.0283	0.9327±0.0620	0.8057±0.2371	0.9300±0.0643
	EBO	0.9829±0.0282	0.9315±0.0631	0.8642±0.1432	0.9285±0.0658
	BAC	0.9863±0.0292	0.9547±0.0625	0.9169±0.1327	0.9530±0.0648
	Easy	0.9784±0.0266	0.9094±0.0514	0.7531±0.2110	0.9055±0.0535
	MNESG_IE (default)	0.9991±0.0012	0.9550±0.0622	0.9492±0.0705	0.9521±0.0664
	MNESG_IE	0.9996±0.0010	0.9998±0.0001	0.9778±0.0497	0.9998±0.0001

As shown in the above table, the proposed MNESG_IE achieves the best performance on more than 30 datasets. It generally shows a significant improvement in each evaluation metric compared to the classical IE algorithms. For example, MNESG_IE obtains the best performance on Kr-vs-k-zero_vs_eight with the mean ACC, AUC, F-M, and G-M results of 0.9932, 0.9965, 0.8684, and 0.9965, which are 5.58%, 4.01%, 22.48%, and 4.03% better than the second-best results, respectively. It means that the correlation information among samples and envelope samples produced by MNESG_IE is effective in improving the classification performance.

The nonparametric tests are used to evaluate the difference between multiple methods. Figure 9 and Table 5 present comparative results of the MNESG_IE and seven classical IE algorithms using the Friedman test and Holm test. The conclusion is that the MNESG_IE achieves the highest average ranking among all the methods in terms of four evaluation metrics from Figure 9. The results in Table 5 show the significance levels between MNESG_IE and other IE algorithms. All the p-value in Table 5 is less than 0.05, which means MNESG_IE and the compared algorithms have significant differences in terms of performance. From results in Figure 9, we know that MNESG_IE achieves the highest average ranking among all the methods. In general, the nonparametric tests confirm the significant performance improvement of MNESG_IE over other classical IE algorithms.

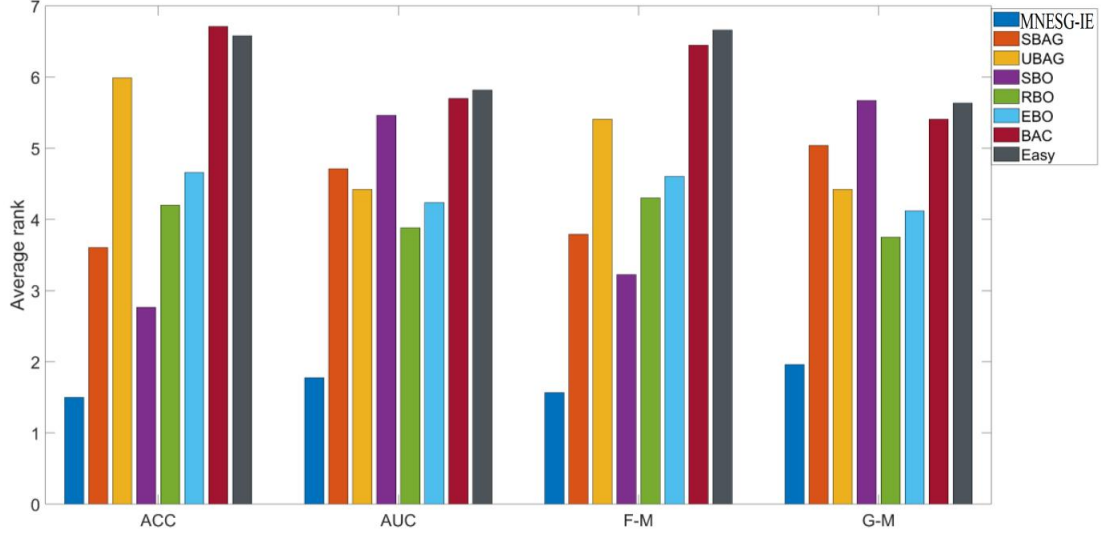


Fig. 9. Average ranks of all compared imbalanced ensemble methods

Table 5. P values from Holm's test

Algorithm	ACC	AUC	F-M	G-M	Hypothesis (0.05)
SBAG	1.4394e-07	9.8556e-10	1.8974e-08	8.1672e-12	Rejected
UBAG	2.2882e-27	5.0702e-09	4.4469e-20	1.3204e-08	Rejected
SBO	1.1707e-03	2.4488e-15	1.2966e-05	5.8939e-16	Rejected
RBO	1.7973e-11	3.3785e-06	2.7651e-11	1.9413e-05	Rejected
EBO	1.7601e-16	3.6705e-09	1.8417e-14	1.4852e-07	Rejected
BAC	1.5289e-35	7.6447e-18	2.7724e-30	9.4960e-15	Rejected
Easy	1.8088e-34	1.3937e-18	5.4957e-32	3.9370e-16	Rejected

2.4.2. Comparison with state-of-the-art IE algorithms

To further verify the performance of the proposed MNESG_IE, four state-of-the-art IE algorithms: EASE, SPE, HUE, and ECUBoost were chosen for comparison. The specific results are shown in Table 6, and the complete results are shown in the supplementary material.

Table 6. The comparison results between EASE, SPE, HUE, ECUBoost and MNESG_IE

ID	Algorithm	ACC	AUC	F-M	G-M
1	EASE	0.7756±0.0801	0.7687±0.0901	0.7008±0.1109	0.7668±0.0911
	SPE	0.7754±0.0646	0.7546±0.0679	0.6845±0.0901	0.7509±0.0694
	HUE	0.7942±0.0351	0.7872±0.0430	0.7238±0.0462	0.7848±0.0426
	ECUBoost	0.7620±0.0776	0.7543±0.0821	0.6849±0.0934	0.7511±0.0816
	MNESG_IE (default)	0.7614±0.0835	0.7757±0.0403	0.7130±0.0393	0.7543±0.0615
	MNESG_IE	0.9721±0.0104	0.9662±0.0011	0.9599±0.0125	0.9656±0.0011
2	EASE	0.9326±0.0224	0.9337±0.0234	0.9072±0.0303	0.9335±0.0234
	SPE	0.9238±0.0149	0.9154±0.0055	0.8915±0.0150	0.9140±0.0057
	HUE	0.9253±0.0202	0.9262±0.0182	0.8975±0.0255	0.9259±0.0183
	ECUBoost	0.9385±0.0228	0.9401±0.0218	0.9154±0.0308	0.9399±0.0218
	MNESG_IE (default)	0.9780±0.0052	0.9812±0.0048	0.9694±0.0070	0.9811±0.0048
	MNESG_IE	0.9898±0.0083	0.9921±0.0064	0.9857±0.0115	0.9921±0.0065

3	EASE	0.6523±0.0306	0.6474±0.0297	0.5590±0.0319	0.6466±0.0297
	SPE	0.6470±0.0277	0.6335±0.0344	0.5372±0.0446	0.6311±0.0359
	HUE	0.6588±0.0338	0.6419±0.0339	0.5455±0.0387	0.6392±0.0339
	ECUBoost	0.6745±0.0314	0.6437±0.0312	0.5357±0.0431	0.6326±0.0367
	MNESG_IE (default)	0.7435±0.0349	0.7220±0.0475	0.6223±0.1029	0.6945±0.0933
	MNESG_IE	0.8243±0.0302	0.7954±0.0343	0.7327±0.0494	0.7823±0.0460
4	EASE	1±0	1±0	1±0	1±0
	SPE	1±0	1±0	1±0	1±0
	HUE	1±0	1±0	1±0	1±0
	ECUBoost	1±0	1±0	1±0	1±0
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0
5	EASE	0.7358±0.0238	0.7256±0.0244	0.6054±0.0311	0.7241±0.0248
	SPE	0.7237±0.0408	0.7040±0.0357	0.5804±0.0441	0.7015±0.0355
	HUE	0.7243±0.0278	0.7058±0.0278	0.5818±0.0353	0.7042±0.0280
	ECUBoost	0.7385±0.0191	0.7193±0.0161	0.5986±0.0205	0.7175±0.0162
	MNESG_IE (default)	0.6914±0.0558	0.7028±0.0194	0.5788±0.0222	0.6957±0.0227
	MNESG_IE	0.7399±0.0516	0.7397±0.0239	0.6246±0.0357	0.7366±0.0211
6	EASE	0.6312±0.0802	0.5877±0.0779	0.4199±0.0827	0.5792±0.0792
	SPE	0.5846±0.0690	0.5823±0.0730	0.4259±0.0793	0.5807±0.0720
	HUE	0.5819±0.0796	0.5339±0.0715	0.3559±0.0764	0.5188±0.0707
	ECUBoost	0.4932±0.0634	0.5519±0.0743	0.4132±0.0719	0.5327±0.0719
	MNESG_IE (default)	0.7284±0.1128	0.6819±0.0402	0.5259±0.0744	0.6333±0.0653
	MNESG_IE	0.9905±0.0130	0.9947±0.0072	0.9556±0.0609	0.9947±0.0073
7	EASE	0.9645±0.0134	0.9559±0.0189	0.9257±0.0280	0.9556±0.0191
	SPE	0.9645±0.0158	0.9595±0.0166	0.9328±0.0294	0.9593±0.0167
	HUE	0.9538±0.0219	0.9525±0.0178	0.9149±0.0375	0.9523±0.0178
	ECUBoost	0.9704±0.0106	0.9636±0.0126	0.9434±0.0183	0.9632±0.0128
	MNESG_IE (default)	0.9468±0.0120	0.9447±0.0146	0.9011±0.0218	0.9446±0.0146
	MNESG_IE	0.9503±0.0456	0.9606±0.0313	0.9148±0.0694	0.9599±0.0325
8	EASE	0.7623±0.0278	0.7281±0.0367	0.5815±0.0494	0.7242±0.0387
	SPE	0.7553±0.0079	0.7206±0.0073	0.5713±0.0100	0.7168±0.0088
	HUE	0.7270±0.0374	0.7333±0.0376	0.5783±0.0438	0.7311±0.0363
	ECUBoost	0.7647±0.0285	0.6860±0.0193	0.5308±0.0293	0.6672±0.0172
	MNESG_IE (default)	0.7623±0.0278	0.7281±0.0367	0.5815±0.0494	0.7242±0.0387
	MNESG_IE	0.8120±0.0383	0.7870±0.0509	0.6601±0.0586	0.7785±0.0612
9	EASE	0.9346±0.0307	0.9173±0.0380	0.8668±0.0569	0.9157±0.0391
	SPE	0.8967±0.0448	0.8558±0.0552	0.7840±0.0720	0.8472±0.0607
	HUE	0.7270±0.0374	0.7333±0.0376	0.5783±0.0438	0.7311±0.0363
	ECUBoost	0.9345±0.0175	0.9172±0.0419	0.8640±0.0376	0.9133±0.0461
	MNESG_IE (default)	0.9346±0.0307	0.9173±0.0380	0.8668±0.0569	0.9157±0.0391
	MNESG_IE	0.9953±0.0104	0.9970±0.0068	0.9905±0.0213	0.9969±0.0068
10	EASE	0.9597±0.0196	0.9545±0.0184	0.9181±0.0376	0.9544±0.0184
	SPE	0.9598±0.0137	0.9459±0.0160	0.9154±0.0267	0.9451±0.0168

	HUE	0.9491±0.0087	0.9581±0.0087	0.9004±0.0162	0.9579±0.0086
	ECUBoost	0.9539±0.0195	0.9542±0.0203	0.9084±0.0357	0.9535±0.0206
	MNESG_IE (default)	0.8208±0.0199	0.9007±0.0545	0.8336±0.0542	0.8966±0.0618
	MNESG_IE	0.9657±0.0190	0.9655±0.0118	0.9311±0.0349	0.9652±0.0119
11	EASE	0.8989±0.0314	0.8607±0.0500	0.7802±0.0693	0.8560±0.0536
	SPE	0.8841±0.0349	0.8484±0.0653	0.7524±0.0802	0.8406±0.0732
	HUE	0.8571±0.0545	0.8704±0.0412	0.7476±0.0791	0.8693±0.0418
	ECUBoost	0.9136±0.0404	0.8569±0.0625	0.7981±0.0987	0.8488±0.0684
	MNESG_IE (default)	0.9139±0.0504	0.9070±0.0425	0.8331±0.0843	0.9046±0.0453
	MNESG_IE	0.9910±0.0082	0.9800±0.0183	0.9793±0.0189	0.9797±0.0186
12	EASE	0.9256±0.0523	0.8871±0.0689	0.7886±0.1235	0.8840±0.0704
	SPE	0.9435±0.0303	0.8942±0.0525	0.8214±0.0922	0.8895±0.0569
	HUE	0.8838±0.0176	0.8846±0.0208	0.7035±0.0223	0.8831±0.0203
	ECUBoost	0.8746±0.1476	0.8575±0.0961	0.7371±0.1897	0.8454±0.1095
	MNESG_IE (default)	0.9316±0.0266	0.9202±0.0526	0.8035±0.0748	0.9191±0.0534
	MNESG_IE	0.9761±0.0257	0.9538±0.0291	0.9242±0.0787	0.9532±0.0293
13	EASE	0.9486±0.0092	0.9250±0.0623	0.8197±0.0512	0.9210±0.0676
	SPE	0.9439±0.0347	0.8944±0.0963	0.7921±0.1383	0.8861±0.1090
	HUE	0.9302±0.0465	0.9175±0.0435	0.7967±0.1158	0.9157±0.0441
	ECUBoost	0.9485±0.0519	0.9250±0.0669	0.8387±0.1382	0.9203±0.0721
	MNESG_IE (default)	0.9673±0.0126	0.9359±0.0408	0.8823±0.0369	0.9333±0.0429
	MNESG_IE	0.9766±0.0233	0.9586±0.0709	0.9132±0.0881	0.9551±0.0783
14	EASE	0.9353±0.0062	0.8826±0.0331	0.7335±0.0327	0.8791±0.0354
	SPE	0.9366±0.0157	0.8810±0.0263	0.7392±0.0576	0.8779±0.0281
	HUE	0.9110±0.0208	0.8964±0.0203	0.6876±0.0478	0.8955±0.0213
	ECUBoost	0.9366±0.0152	0.8969±0.0334	0.7477±0.0422	0.8942±0.0365
	MNESG_IE (default)	0.9373±0.0182	0.8819±0.0776	0.7349±0.0974	0.8753±0.0897
	MNESG_IE	0.9905±0.0130	0.9947±0.0072	0.9556±0.0609	0.9947±0.0073
15	EASE	0.9078±0.0472	0.8097±0.1166	0.6100±0.1820	0.7882±0.1378
	SPE	0.8986±0.0259	0.7919±0.0953	0.5695±0.1207	0.7721±0.1094
	HUE	0.8572±0.0444	0.8698±0.0447	0.5736±0.0911	0.8693±0.0443
	ECUBoost	0.9195±0.0336	0.8414±0.0921	0.6571±0.1281	0.8274±0.1096
	MNESG_IE (default)	0.8958±0.0153	0.9166±0.0357	0.6539±0.0376	0.9154±0.0350
	MNESG_IE	0.9628±0.0584	0.9442±0.0516	0.8806±0.1533	0.9413±0.0544
16	EASE	0.9416±0.0261	0.8623±0.0856	0.7188±0.1360	0.8517±0.0979
	SPE	0.9572±0.0131	0.9316±0.0273	0.8096±0.0486	0.9302±0.0285
	HUE	0.8910±0.0187	0.9047±0.0378	0.6277±0.0546	0.9037±0.0374
	ECUBoost	0.9163±0.0265	0.8661±0.0434	0.6630±0.0430	0.8594±0.0508
	MNESG_IE (default)	0.9590±0.0273	0.9068±0.0702	0.8080±0.1153	0.9006±0.0796
	MNESG_IE	0.9845±0.0128	0.9833±0.0251	0.9286±0.0532	0.9832±0.0253
17	EASE	0.8806±0.0233	0.8095±0.0620	0.5387±0.0817	0.8021±0.0676
	SPE	0.8599±0.0494	0.8016±0.0910	0.5109±0.1306	0.7956±0.0947
	HUE	0.8086±0.0549	0.7813±0.0486	0.4440±0.1000	0.7794±0.0493
	ECUBoost	0.8977±0.0299	0.7840±0.0367	0.5538±0.0394	0.7672±0.0488

	MNESG_IE (default)	0.8146±0.0478	0.8176±0.0934	0.4618±0.0965	0.8112±0.1015
	MNESG_IE	0.9263±0.0039	0.9235±0.0779	0.6977±0.0370	0.9184±0.0873
18	EASE	0.7916±0.0600	0.5700±0.0783	0.2004±0.1249	0.4436±0.2274
	SPE	0.7759±0.0543	0.7054±0.1384	0.2966±0.1573	0.6121±0.3086
	HUE	0.6399±0.1113	0.7059±0.0675	0.2799±0.0489	0.6843±0.0737
	ECUBoost	0.6197±0.1118	0.5061±0.1231	0.1207±0.0815	0.3878±0.2236
	MNESG_IE (default)	0.8961±0.0019	0.6244±0.0194	0.3333±0.0006	0.5267±0.0479
	MNESG_IE	0.9795±0.0115	0.9000±0.0559	0.8857±0.0639	0.8928±0.0599
19	EASE	0.7944±0.0165	0.7202±0.1439	0.3147±0.1203	0.6918±0.1640
	SPE	0.7849±0.0237	0.6842±0.0887	0.2911±0.0997	0.6617±0.1144
	HUE	0.6966±0.0859	0.7596±0.0851	0.3145±0.0845	0.7518±0.0819
	ECUBoost	0.7988±0.0549	0.7003±0.0758	0.3233±0.0975	0.6831±0.0928
	MNESG_IE (default)	0.6957±0.1946	0.7595±0.0043	0.3018±0.0874	0.7378±0.0187
	MNESG_IE	0.9814±0.0104	0.8607±0.0745	0.8400±0.0894	0.8532±0.0821
20	EASE	0.8323±0.0389	0.7088±0.0945	0.2996±0.0735	0.6767±0.1230
	SPE	0.7581±0.0205	0.7001±0.0920	0.2519±0.0564	0.6888±0.0945
	HUE	0.6950±0.0887	0.6663±0.1180	0.2188±0.0905	0.6430±0.1286
	ECUBoost	0.8583±0.0308	0.7537±0.0489	0.3722±0.0445	0.7390±0.0603
	MNESG_IE (default)	0.6455±0.1230	0.8023±0.0658	0.2690±0.0662	0.7753±0.0848
	MNESG_IE	0.8110±0.0531	0.8214±0.0294	0.3732±0.0614	0.8210±0.0291
21	EASE	0.9486±0.0270	0.8160±0.1426	0.5895±0.2248	0.7847±0.1711
	SPE	0.9486±0.0227	0.8943±0.1058	0.6599±0.1688	0.8842±0.1187
	HUE	0.8321±0.0919	0.8626±0.0873	0.4204±0.1110	0.8502±0.0989
	ECUBoost	0.9627±0.0237	0.9492±0.0710	0.7690±0.1489	0.9466±0.0759
	MNESG_IE (default)	0.9765±0.0001	0.9104±0.1090	0.8000±0.0001	0.9020±0.1209
	MNESG_IE	0.9902±0.0134	0.9949±0.0070	0.9200±0.1095	0.9948±0.0071
22	EASE	0.9582±0.0274	0.8373±0.1251	0.6716±0.2234	0.8096±0.1669
	SPE	0.9642±0.0223	0.8873±0.0871	0.7365±0.1263	0.8756±0.1012
	HUE	0.8747±0.0515	0.8865±0.1137	0.4833±0.1598	0.8808±0.1211
	ECUBoost	0.9612±0.0152	0.8623±0.1014	0.6978±0.0939	0.8436±0.1202
	MNESG_IE (default)	0.9045±0.0977	0.9258±0.0542	0.6436±0.2550	0.9214±0.0582
	MNESG_IE	0.9911±0.0081	0.9953±0.0043	0.9333±0.0609	0.9952±0.0043
23	EASE	0.9083±0.0205	0.7935±0.0862	0.4554±0.1128	0.7758±0.1065
	SPE	0.8768±0.0096	0.7574±0.0591	0.3636±0.0334	0.7407±0.0688
	HUE	0.8371±0.0160	0.7571±0.0417	0.3202±0.0416	0.7501±0.0469
	ECUBoost	0.8617±0.0470	0.7951±0.0890	0.3780±0.0694	0.7766±0.1180
	MNESG_IE (default)	0.9110±0.0581	0.8057±0.0724	0.4882±0.2028	0.7968±0.0756
	MNESG_IE	0.9623±0.0048	0.9212±0.0807	0.7166±0.0151	0.9173±0.0859
24	EASE	1±0	1±0	1±0	1±0
	SPE	0.9923±0.0153	0.9000±0.2000	0.8000±0.4000	0.8000±0.4000
	HUE	1±0	1±0	1±0	1±0
	ECUBoost	0.9923±0.0153	0.9500±0.0999	0.9333±0.1333	0.9414±0.1171
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0

25	EASE	0.9812±0.0271	0.8926±0.1968	0.7142±0.3938	0.7925±0.3965
	SPE	0.9812±0.0271	0.9902±0.0142	0.8476±0.1890	0.9900±0.0144
	HUE	0.8788±0.0677	0.9365±0.0356	0.4504±0.1291	0.9336±0.0382
	ECUBoost	0.9112±0.0629	0.9536±0.0330	0.5454±0.2082	0.9518±0.0348
	MNESG_IE (default)	0.9884±0.0164	0.9939±0.0086	0.9000±0.1414	0.9939±0.0087
	MNESG_IE	0.9884±0.0164	0.9939±0.0086	0.9000±0.1414	0.9939±0.0087
26	EASE	0.9169±0.0287	0.7653±0.0626	0.3915±0.1006	0.7433±0.0756
	SPE	0.8609±0.0267	0.7600±0.1574	0.2637±0.1019	0.7180±0.2037
	HUE	0.7923±0.0579	0.7959±0.0973	0.2517±0.0684	0.7895±0.0990
	ECUBoost	0.8318±0.0444	0.7448±0.0453	0.2510±0.0470	0.7334±0.0569
	MNESG_IE (default)	0.8961±0.0927	0.7066±0.1603	0.3026±0.0901	0.6246±0.2253
	MNESG_IE	0.9792±0.0121	0.8696±0.0978	0.7532±0.1358	0.8547±0.1152
27	EASE	0.9083±0.0164	0.7893±0.0782	0.3316±0.0671	0.7725±0.0917
	SPE	0.8712±0.0095	0.8000±0.0392	0.2791±0.0350	0.7951±0.0437
	HUE	0.8167±0.0300	0.8112±0.0534	0.2351±0.0373	0.8094±0.0520
	ECUBoost	0.9204±0.0229	0.8176±0.0429	0.3891±0.0734	0.8087±0.0485
	MNESG_IE (default)	0.8653±0.0190	0.8579±0.0243	0.2990±0.0123	0.8573±0.0247
	MNESG_IE	0.9097±0.0142	0.8209±0.0693	0.3553±0.0457	0.8122±0.0781
28	EASE	0.7785±0.0302	0.6124±0.0568	0.1142±0.0266	0.5779±0.0831
	SPE	0.7173±0.0542	0.6168±0.0693	0.1091±0.0345	0.6010±0.0795
	HUE	0.6960±0.0204	0.6892±0.0451	0.1294±0.0156	0.6877±0.0459
	ECUBoost	0.6292±0.1777	0.5906±0.0374	0.0932±0.0192	0.5529±0.0582
	MNESG_IE (default)	0.6953±0.1522	0.7326±0.0554	0.1584±0.0331	0.7074±0.0423
	MNESG_IE	0.9677±0.0018	0.6034±0.1003	0.2699±0.1787	0.4239±0.2120
29	EASE	0.7877±0.0520	0.6646±0.0267	0.1424±0.0266	0.6491±0.0288
	SPE	0.6980±0.0373	0.6989±0.0447	0.1274±0.0117	0.6940±0.0469
	HUE	0.6737±0.0439	0.6703±0.0719	0.1129±0.0223	0.6592±0.0845
	ECUBoost	0.6737±0.1902	0.7025±0.1043	0.1559±0.0735	0.6843±0.1096
	MNESG_IE (default)	0.6943±0.1227	0.7919±0.0580	0.1916±0.1012	0.7826±0.0852
	MNESG_IE	0.8468±0.0976	0.8000±0.0066	0.2628±0.0997	0.7954±0.0137
30	EASE	0.9770±0.0068	0.9006±0.0427	0.6854±0.0728	0.8956±0.0481
	SPE	0.9757±0.0107	0.8892±0.0673	0.6666±0.1285	0.8815±0.0777
	HUE	0.9501±0.0100	0.9513±0.0299	0.5349±0.0514	0.9509±0.0300
	ECUBoost	0.9770±0.0077	0.8885±0.0738	0.6720±0.0938	0.8794±0.0842
	MNESG_IE (default)	0.9158±0.0381	0.9566±0.0196	0.4342±0.1065	0.9554±0.0207
	MNESG_IE	0.9636±0.0140	0.9597±0.0291	0.6231±0.0788	0.9592±0.0295
31	EASE	0.9420±0.0124	0.8030±0.0698	0.3509±0.0635	0.7842±0.0874
	SPE	0.9103±0.0311	0.8425±0.1048	0.3038±0.0909	0.8302±0.1192
	HUE	0.8786±0.0248	0.8542±0.0245	0.2482±0.0312	0.8529±0.0265
	ECUBoost	0.9615±0.0083	0.8130±0.0597	0.4507±0.0919	0.7946±0.0768
	MNESG_IE (default)	0.8814±0.0237	0.8974±0.0921	0.2685±0.0570	0.8921±0.0994
	MNESG_IE	0.9137±0.0173	0.8868±0.0170	0.3234±0.0472	0.8861±0.0175
32	EASE	0.9433±0.0214	0.7511±0.1385	0.2923±0.1600	0.6427±0.3257
	SPE	0.8888±0.0312	0.7232±0.0841	0.1834±0.0531	0.6890±0.1201

	HUE	0.8022±0.0191	0.7034±0.0887	0.1153±0.0293	0.6816±0.1219
	ECUBoost	0.6188±0.1883	0.6829±0.0831	0.1007±0.0560	0.6595±0.0964
	MNESG_IE (default)	0.8694±0.0904	0.8111±0.0462	0.2398±0.1322	0.8082±0.0429
	MNESG_IE	0.9889±0.0045	0.8111±0.0722	0.7113±0.1216	0.7855±0.0916
33	EASE	0.7964±0.0550	0.5859±0.1207	0.0732±0.0488	0.4735±0.2590
	SPE	0.6257±0.0731	0.6213±0.0944	0.0623±0.0188	0.5999±0.1044
	HUE	0.5918±0.0551	0.6367±0.0701	0.0638±0.0046	0.6250±0.0588
	ECUBoost	0.2350±0.1469	0.5604±0.0913	0.0480±0.0092	0.4130±0.1613
	MNESG_IE (default)	0.6932±0.1517	0.6599±0.0788	0.1222±0.0629	0.5730±0.1182
	MNESG_IE	0.8302±0.1735	0.9136±0.0082	0.2679±0.2273	0.9069±0.0974
34	EASE	0.9863±0.0075	0.9244±0.0983	0.7045±0.1528	0.9151±0.1137
	SPE	0.9678±0.0073	0.8627±0.1268	0.4549±0.1158	0.8407±0.1573
	HUE	0.9171±0.0135	0.9577±0.0069	0.3116±0.0331	0.9568±0.0072
	ECUBoost	0.8712±0.0521	0.8951±0.0940	0.2418±0.1103	0.8914±0.0970
	MNESG_IE (default)	0.9534±0.0093	0.9436±0.0745	0.4269±0.0647	0.9408±0.0801
	MNESG_IE	0.9932±0.0080	0.9965±0.0041	0.8684±0.0239	0.9965±0.0041
35	EASE	1±0	1±0	1±0	1±0
	SPE	1±0	1±0	1±0	1±0
	HUE	1±0	1±0	1±0	1±0
	ECUBoost	1±0	1±0	1±0	1±0
	MNESG_IE (default)	0.9970±0.0024	0.9262±0.0631	0.8895±0.0890	0.9212±0.0697
	MNESG_IE	1±0	1±0	1±0	1±0
36	EASE	1±0	1±0	1±0	1±0
	SPE	1±0	1±0	1±0	1±0
	HUE	0.9995±0.0008	0.9833±0.0333	0.9818±0.0363	0.9825±0.0348
	ECUBoost	0.9991±0.0010	0.9666±0.0408	0.9636±0.0445	0.9651±0.0426
	MNESG_IE (default)	1±0	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0	1±0
37	EASE	0.9990±0.0011	0.9797±0.0398	0.9595±0.0498	0.9786±0.0421
	SPE	0.9995±0.0009	0.9997±0.0004	0.9846±0.0307	0.9997±0.0004
	HUE	0.9872±0.0099	0.9771±0.0317	0.6964±0.1634	0.9763±0.0331
	ECUBoost	1±0	1±0	1±0	1±0
	MNESG_IE (default)	0.9973±0.0019	0.9788±0.0441	0.9005±0.0651	0.9777±0.0466
	MNESG_IE	0.9995±0.0010	0.9833±0.0622	0.9818±0.0407	0.9826±0.0390
38	EASE	0.9995±0.0008	0.9800±0.0399	0.9777±0.0444	0.9788±0.0422
	SPE	0.9991±0.0011	0.9550±0.0556	0.9492±0.0630	0.9520±0.0593
	HUE	0.9950±0.0078	0.9479±0.0637	0.8300±0.2357	0.9446±0.0678
	ECUBoost	0.9991±0.0017	0.9600±0.0799	0.9500±0.1000	0.9549±0.0901
	MNESG_IE (default)	0.9991±0.0012	0.9550±0.0622	0.9492±0.0705	0.9521±0.0664
	MNESG_IE	0.9996±0.0010	0.9998±0.0001	0.9778±0.0497	0.9998±0.0001

As shown in Table 6, the four evaluation metrics obtained by MNESG_IE with default and optimized parameters are all better than the compared methods. For example, MNESG_IE obtains the best performance on Ecoli4 with the mean ACC, AUC, F-M, and G-M results of 0.9911, 0.9953,

0.9333, and 0.9952, which are 2.69%, 10.8%, 19.68%, and 11.44% better than the second-best (except for the proposed algorithm with default parameters) results, respectively. In addition, the proposed algorithm obtained the best performance on 32 datasets for ACC, AUC and G-M metrics and on 33 datasets for F-M metric. This means that MNESG_IE can effectively mine the correlation information among samples to generate high quality envelope samples with better separability and diversity, which is helpful for imbalanced classification.

The Friedman test and Holm test are also used to evaluate the difference between MNESG_IE and four SOTA imbalanced ensemble methods, the specific results are shown in Tables 7-8.

Table 7. Average rank numbers of compared IE methods

Algorithm	ACC	AUC	F-M	G-M
MNESG_IE	1.5658	1.3816	1.5395	1.3947
EASE	2.6053	3.2500	2.6842	3.3684
SPE	3.3158	3.6184	3.4605	3.6316
HUE	4.3816	3.2895	4.1842	3.1053
ECUBoost	3.1316	3.4605	3.1316	3.5000

Table 8. P values from Holm's test

Algorithm	ACC	AUC	F-M	G-M	Hypothesis (0.05)
EASE	2.0387e-04	4.5857e-09	9.3157e-06	1.1536e-10	Rejected
SPE	1.3621e-09	6.2702e-12	9.5064e-12	4.9002e-14	Rejected
HUE	3.6905e-20	1.1890e-08	1.0609e-17	3.2926e-10	Rejected
ECUBoost	8.2446e-08	1.4279e-10	1.5560e-08	7.8126e-12	Rejected

As shown in Table 7, the rank numbers of MNESG_IE in the four evaluation metrics are 1.5658, 1.3816, 1.5395, and 1.3947 in order, which are the lowest. Table 8 shows that the equivalence hypotheses between MNESG_IE and the compared algorithms are all rejected. Therefore, the proposed algorithm MNESG_IE achieve significant improvements compared to the state-of-the-art IE algorithms.

To further evaluate the statistically significant difference between methods, the Wilcoxon paired signed-rank test was adopted to conduct the four comparisons: MNESG_IE vs EASE, MNESG_IE vs SPE, MNESG_IE vs HUE, and MNESG_IE vs ECUBoost. The specific results are shown in Table 9.

Table 9. Results of the Wilcoxon pairwise test

Algorithm	Measure	R+	R-	P value	Hypothesis (0.05)
MNESG_IE vs EASE	ACC	676	65	1.0708e-05	Rejected
	AUC	681	22	8.1248e-07	Rejected
	F-M	698	43	2.4761e-06	Rejected
	G-M	697	44	5.2082e-06	Rejected
MNESG_IE vs SPE	ACC	679	24	1.4692e-06	Rejected
	AUC	683	20	6.8159e-07	Rejected
	F-M	715	26	8.1248e-07	Rejected
	G-M	664	39	4.0111e-06	Rejected
MNESG_IE vs HUE	ACC	697.5	5.5	2.9491e-07	Rejected
	AUC	713	28	1.2457e-06	Rejected
	F-M	698.5	4.5	2.7019e-07	Rejected

	G-M	667.5	35.5	3.5642e-06	Rejected
	ACC	638.5	64.5	1.5890e-05	Rejected
	AUC	685.5	17.5	3.8848e-07	Rejected
MNESG_IE vs ECUBoost	F-M	686.5	54.5	4.4861e-06	Rejected
	G-M	697.5	43.5	2.6341e-06	Rejected

The statistical test results in Table 9 rejects the all hypotheses of equivalence in four comparisons. Meanwhile, R+ in the table indicates that the rank sum of MNESG_IE outperforms the compared algorithm based on 38 datasets, and R- indicates that the rank sum of the compared algorithm outperforms MNESG_IE. Table 9 shows that R+ is much larger than R-, which means that MNESG_IE is significantly better than the four state-of-the-art IE algorithms.

2.4.3. Comparison with deep learning based IE algorithms

To verify the competitiveness of the proposed algorithm, MNESG_IE is further compared with six advanced DL based imbalanced classification methods: CNN+SMOTE, CNN+AE+GAN, BED, RVGAN-TL, EAL-GAN, DLE-ISMOTE. Since the DL based methods are time consuming, 10 representative datasets are chosen for comparison. The specific results are shown in Table 10, and the complete results are shown in the supplementary material.

Table 10. The comparison results with deep learning based imbalanced classification methods

Dataset	Algorithm	AUC	F-M	G-M
Ecoli1	CNN+SMOTE	0.6522±0.0381	0.4668±0.0509	0.6405±0.0399
	CNN+AE+GAN	0.8376±0.0541	0.4303±0.0690	0.8295±0.0556
	BED	0.8495±0.0251	0.7584±0.0268	0.8467±0.0269
	RVGAN-TL	0.8123±0.0510	0.7252±0.0654	0.7991±0.0583
	EAL-GAN	0.9226±0.0441	0.6968±0.0960	0.8624±0.0518
	DLE-ISMOTE	0.7663±0.0599	0.5826±0.0710	0.7463±0.0910
	MNESG_IE (default)	0.9070±0.0425	0.8331±0.0843	0.9046±0.0453
	MNESG_IE	0.9800±0.0183	0.9793±0.0189	0.9797±0.0186
Ecoli3	CNN+SMOTE	0.8288±0.0831	0.4810±0.0636	0.8209±0.0851
	CNN+AE+GAN	0.8325±0.0129	0.4944±0.0305	0.8321±0.0134
	BED	0.9197±0.0212	0.5941±0.0679	0.9159±0.0231
	RVGAN-TL	0.7443±0.0883	0.5621±0.1560	0.7031±0.1180
	EAL-GAN	0.9200±0.0574	0.6086±0.1530	0.8650±0.0664
	DLE-ISMOTE	0.8452±0.0183	0.5227±0.0227	0.8451±0.0185
	MNESG_IE (default)	0.9166±0.0357	0.6539±0.0376	0.9154±0.0350
	MNESG_IE	0.9442±0.0516	0.8806±0.1533	0.9413±0.0544
Glass-0-1-6_vs_2	CNN+SMOTE	0.6025±0.0765	0.2151±0.0601	0.5629±0.0842
	CNN+AE+GAN	0.6065±0.0932	0.2314±0.0872	0.5952±0.1070
	BED	0.7133±0.0775	0.2488±0.0416	0.6887±0.0682
	RVGAN-TL	0.5935±0.1600	0.1967±0.2410	0.3269±0.1560
	EAL-GAN	0.6389±0.1290	0.3704±0.1250	0.5992±0.0997
	DLE-ISMOTE	0.6698±0.1210	0.2269±0.0769	0.6143±0.1400
	MNESG_IE (default)	0.6244±0.0194	0.3333±0.0006	0.5267±0.0479
	MNESG_IE	0.9000±0.0559	0.8857±0.0639	0.8928±0.0599
Shuttle-c2-vs-c4	CNN+SMOTE	0.9957±0.0088	0.9918±0.0112	0.9957±0.0089

	CNN+AE+GAN	0.9997±0.0007	0.9959±0.0091	0.9997±0.0007
	BED	1±0	1±0	1±0
	RVGAN-TL	1±0	1±0	1±0
	EAL-GAN	0.9960±0.0120	0.9940±0.0143	0.9905±0.0234
	DLE-ISMOTE	0.9791±0.0022	0.9787±0.0054	0.9789±0.0028
	MNESG_IE (default)	1±0	1±0	1±0
	MNESG_IE	1±0	1±0	1±0
Yeast4	CNN+SMOTE	0.8279±0.0325	0.2963±0.0204	0.8250±0.0349
	CNN+AE+GAN	0.7970±0.0335	0.2983±0.0538	0.7907±0.0380
	BED	0.8573±0.0080	0.2508±0.0155	0.8562±0.0084
	RVGAN-TL	0.6352±0.0755	0.3328±0.0716	0.4948±0.0937
	EAL-GAN	0.8764±0.0604	0.3045±0.1130	0.4533±0.1130
	DLE-ISMOTE	0.8552±0.0179	0.2910±0.0252	0.8545±0.0181
	MNESG_IE (default)	0.8579±0.0243	0.2990±0.0123	0.8573±0.0247
	MNESG_IE	0.8209±0.0693	0.3553±0.0457	0.8122±0.0781
Yeast5	CNN+SMOTE	0.9242±0.0513	0.4922±0.1120	0.9228±0.0560
	CNN+AE+GAN	0.9183±0.0398	0.4335±0.0560	0.9168±0.0408
	BED	0.9785±0.0016	0.5925±0.0168	0.9782±0.0016
	RVGAN-TL	0.7991±0.0721	0.6638±0.0878	0.7685±0.0977
	EAL-GAN	0.9715±0.0391	0.5556±0.0242	0.8094±0.0264
	DLE-ISMOTE	0.9687±0.0086	0.4810±0.0707	0.9682±0.0089
	MNESG_IE (default)	0.9566±0.0196	0.4342±0.1065	0.9554±0.0207
	MNESG_IE	0.9597±0.0291	0.6231±0.0788	0.9592±0.0295
Yeast6	CNN+SMOTE	0.8739±0.0517	0.2753±0.0567	0.8724±0.0532
	CNN+AE+GAN	0.8663±0.0744	0.2485±0.0550	0.8639±0.0754
	BED	0.8829±0.0291	0.2586±0.0147	0.8823±0.0282
	RVGAN-TL	0.7098±0.0678	0.4821±0.1540	0.6415±0.1160
	EAL-GAN	0.9336±0.0355	0.5476±0.0982	0.7020±0.0496
	DLE-ISMOTE	0.9345±0.0186	0.2717±0.0261	0.9321±0.0192
	MNESG_IE (default)	0.8974±0.0921	0.2685±0.0570	0.8921±0.0994
	MNESG_IE	0.8868±0.0170	0.3234±0.0472	0.8861±0.0175
Winequality-red-8_vs_6-7	CNN+SMOTE	0.6434±0.0762	0.0896±0.0288	0.6250±0.0934
	CNN+AE+GAN	0.6026±0.0315	0.0550±0.0041	0.6002±0.0327
	BED	0.6613±0.0621	0.0801±0.0142	0.6587±0.0619
	RVGAN-TL	0.6470±0.0193	0.3800±0.0653	0.5447±0.0372
	EAL-GAN	0.5294±0.1690	0.0400±0.1000	0.0000±0.0000
	DLE-ISMOTE	0.5722±0.0810	0.0957±0.0791	0.5714±0.0865
	MNESG_IE (default)	0.6599±0.0788	0.1222±0.0629	0.5730±0.1182
	MNESG_IE	0.9136±0.0082	0.2679±0.2273	0.9069±0.0974
Shuttle-2_vs_5	CNN+SMOTE	0.9994±0.0008	0.9636±0.0498	0.9994±0.008
	CNN+AE+GAN	0.9989±0.0004	0.9351±0.0237	0.9989±0.004
	BED	0.9991±0.0008	0.9453±0.0471	0.9991±0.008
	RVGAN-TL	1±0	1±0	1±0
	EAL-GAN	1±0	1±0	1±0

	DLE-ISMOTE	0.9992±0.0008	0.9545±0.0455	0.9992±0.008
	MNESG_IE (default)	0.9970±0.0024	0.9262±0.0631	0.8895±0.0890
	MNESG_IE	1±0	1±0	1±0
Rootkit_imapvsback	CNN+SMOTE	0.9993±0.0006	0.9414±0.0541	0.9993±0.0006
	CNN+AE+GAN	0.9050±0.0102	0.8532±0.1380	0.8935±0.1210
	BED	1±0	1±0	1±0
	RVGAN-TL	1±0	1±0	1±0
	EAL-GAN	0.9538±0.0092	0.9190±0.0174	0.9298±0.0161
	DLE-ISMOTE	0.9375±0.0625	0.9286±0.0714	0.9330±0.0670
	MNESG_IE (default)	0.9550±0.0622	0.9492±0.0705	0.9521±0.0664
	MNESG_IE	0.9998±0.0001	0.9778±0.0497	0.9998±0.0001

It can be seen the performance of MNESG_IE is superior to the other DL based imbalanced classification methods on all criteria. In addition, MNESG_IE with default parameters can outperform most of the compared DL based imbalanced classification methods. For example, when considering F-M and G-M, MNESG_IE (default) improves 7.47% and 4.22% on the Ecoli1 dataset compared to the best performance obtained by the comparison algorithms, respectively. MNESG_IE improves the performance on AUC, F-M, and G-M by 5.47%, 22.09%, and 11.73%, respectively, which indicates that the MNESG_IE algorithm is competitive.

3. Time complexity analysis

The computational complexity of MNESG_IE consists of the following four components. (1) The Q sample subsets are divided based on the random undersampling method. (2) The initialization of \mathbf{P} , \mathbf{U} , and \mathbf{V} in the MNESG algorithm is performed by the MNSEP and FCM algorithms. (3) Iterative updating of \mathbf{P} , \mathbf{U} , and \mathbf{V} is conducted by the MNESG algorithm. (4) 2D sparse fusion is performed based on the prediction results.

The computational complexity of the first part is related to the number of sample subsets Q . The second part of initializing \mathbf{P} by MNSEP involves manifold distance calculation and eigendecomposition, so the complexity is $O(N_{min}^3)$. The complexity of FCM algorithm to initialize

\mathbf{U} and \mathbf{V} , the complexity is $O(N_{min} C^2 dt)$, where N_{min} is the number of minority class samples, C is the number of cluster centers, d is the sample dimension, and t is the number of iterations in the FCM algorithm. In the third part, the complexities of updating \mathbf{P} , \mathbf{U} , and \mathbf{V} are $O(N_{min}^2)$,

$O(N_{min} Cd)$, and $O(C^2 d)$, respectively. The computational complexity of the fourth part of the 2D-SFM is related to the number of sample subsets Q and the number of test samples N_t , which can be expressed as $O(QN_t)$. Assuming that the number of iterations is T , the total computational

complexity of the proposed MNESG_IE algorithm can be expressed as $Q + Q \cdot (O(N_{min}^3) + O(N_{min} C^2 dt) + T \cdot (O(N_{min}^2) + O(N_{min} Cd) + O(C^2 d))) + O(QN_t)$. It is worth mentioning that in the case of relatively high imbalance ratio, N_{min} is much smaller than the number

of total samples, so the increased time cost is not significant and relatively close compared to the classical imbalanced ensemble methods. The specific compared results of the running time between the proposed MNESG_IE and the state-of-the-art algorithms are shown in Table 11.

Table 11. The time cost comparison between EASE, SPE, HUE, ECUBoost and MNESG_IE

ID	EASE (s)	SPE (s)	HUE (s)	ECUBoost (s)	MNESG_IE (s)
1	0.3383	0.1446	0.2756	1.1255	6.3436
2	0.1966	0.1465	0.1316	1.4987	14.6192
3	0.2413	0.1994	0.1915	1.7194	37.8634
4	0.0929	0.0578	0.0548	1.0686	1.0621
5	0.3690	0.2565	0.6038	2.6864	58.8042
6	0.1409	0.1226	0.1431	1.1678	3.6503
7	0.2542	0.2293	0.4158	2.0981	28.6764
8	0.3655	0.2473	0.5224	1.9837	58.4714
9	0.0977	0.0987	0.0937	1.2348	2.3798
10	0.2536	0.1705	0.2929	1.7264	25.4692
11	0.1176	0.0847	0.1505	1.2563	3.9906
12	0.1189	0.0817	0.1612	1.2865	2.1714
13	0.1017	0.0767	0.1409	1.1774	1.6414
14	0.2227	0.1457	0.3785	2.0747	9.6333
15	0.1027	0.0907	0.1715	1.1882	1.5926
16	0.1317	0.0917	0.1635	1.4541	2.0436
17	0.1236	0.1216	0.1795	1.2366	2.4931
18	0.1207	0.0768	0.1935	1.1289	1.2314
19	0.1605	0.0738	0.2493	1.2199	1.1017
20	0.1238	0.0867	0.2652	1.1741	1.6201
21	0.0897	0.0817	0.2344	1.0877	0.9547
22	0.1498	0.0787	0.2433	1.1359	0.9520
23	0.2054	0.0957	0.3241	1.3310	2.4676
24	0.1011	0.0588	0.2632	1.0033	0.9593
25	0.0887	0.0877	0.3630	1.1617	0.7127
26	0.1253	0.0897	0.5096	1.3894	1.0945
27	0.1667	0.1370	0.6711	2.0650	2.4580
28	0.2253	0.1220	0.8973	1.8235	5.9541
29	0.1635	0.1028	0.5877	1.5818	1.6469
30	0.2962	0.1119	0.5597	1.7273	1.8731
31	0.1428	0.1167	0.5789	2.2304	1.5674
32	0.1326	0.0907	1.1943	1.3666	1.4893
33	0.1256	0.0887	1.1906	1.5584	1.4865
34	0.1593	0.1008	0.2924	1.6109	0.9908
35	0.2664	0.1207	1.2852	2.8180	2.5935
36	0.1534	0.1052	1.1983	2.4106	2.1295
37	0.1466	0.1087	0.3134	2.2223	1.0250
38	0.2048	0.1256	2.1970	2.4046	1.8002

Based on the analysis above, although there is no significant improvement in the computational complexity of the model for the proposed algorithm, the results in Table 6 show that the proposed algorithm has the best classification performance compared to the state-of-the-art algorithms. Therefore, in summary, the proposed algorithm is still highly competitive.