

Class Imbalanced NIPT Data Processing based on Ensemble Learning: Focusing on Early Detection of Chromosomal Abnormalities in Female Fetuses

Lu Kun^{a, †}, Gaojun Jin^{b, †}, Jiaming Huang^c

International College Zhengzhou University, Zhengzhou, China

^a15236760356@163.com, ^b15290976079@163.com, ^c13782369281@163.com

[†]These authors contribute equally

Abstract

To address class imbalance in non-invasive prenatal testing (NIPT) data for pregnant women, improve the accuracy of early prenatal detection for female fetuses, and reduce the risk of shortened treatment windows, this paper comprehensively compares the classification performance of four ensemble learning models—EasyEnsemble, RUSBoost, AdaBoost, and BalancedRandomForest—for chromosomal aneuploidy labeling. This aims to determine the optimal approach for handling imbalanced data. First, the dataset underwent preprocessing: irrelevant feature columns were removed, categorical data was converted using label encoding, and the dataset was split into training (80%) and testing (20%) sets. Next, the trained models were applied to the testing set to generate predictions, which were then compared against the ground truth labels. Finally, classification results were visualized using confusion matrices and other charts, while classification performance metrics such as accuracy, precision, and F1 score were calculated to compare the classification effectiveness of the four models from different perspectives. Experimental results indicate that the four models exhibit differences in classifying chromosomal aneuploidy labels. Overall, the RUSBoost model demonstrates optimal classification performance with an accuracy rate of 87.91% and good recall, showcasing strong robustness and generalization capabilities. This provides a practical solution for early detection of female fetal abnormalities in clinical settings.

Keywords

Class imbalance, ensemble learning, RUSBoost, non-invasive prenatal testing, recall rate.

1. Introduction

Non-invasive prenatal testing (NIPT), a technique that sequences cell-free DNA fragments in maternal blood to determine fetal chromosomal abnormalities, has gained widespread clinical application [1]. Due to its strong specificity in identifying trisomy 13, 18, and 21, these chromosomal number abnormalities are commonly used as indicators for fetal abnormalities [2]. In recent years, the rapid advancement of machine learning models has presented new opportunities for NIPT technology. However, the performance of these models heavily relies on large amounts of standardized and accurate real-world sampling data. The severe class imbalance in data from pregnant women with abnormalities continues to pose a significant bottleneck for clinical deployment [3].

Class-imbalanced data refers to datasets where the distribution of categories exhibits significant disparities [4]. Class imbalance causes the majority class to dominate model training, resulting in poor prediction performance for the minority class—which typically carries higher task priority. Existing approaches to addressing class imbalance can be broadly categorized

into three types: data-level techniques such as undersampling and oversampling, and algorithm-level methods like ensemble learning. The former two achieve balance by reducing or increasing the number of majority and minority class samples, respectively [5], while the latter enhances classification performance through data fusion, modeling, and mining techniques [6].

Regarding the under-sampling technique (RUS), Reference [7] identified the optimal under-sampling ratio across four datasets from different domains through random under-sampling at varying proportions, without altering the original data structure. While this method is simple and efficient, it may introduce sampling bias and lose crucial information from the original data. Reference [8] compared distance-based undersampling methods (NearMiss) employing multiple sampling principles, pioneering the application of KNN combined with undersampling techniques in actual biological research. Among oversampling techniques, Adaptive Synthetic Oversampling (ADASYN) is the most widely applied. This method assigns corresponding weighted distributions to minority class samples based on their learning difficulty, thereby adaptively generating more data for harder-to-learn minority classes to achieve category balance [9].

Compared to data-level approaches for handling class imbalance, algorithm-level ensemble learning methods such as EasyEnsemble [10], BalanceCascade [11], and SMOTEBoost [12] construct multiple subsets from the original dataset and train multiple classifiers. This effectively mitigates the negative impact of sample imbalance on model learning. Their core principle involves transforming class imbalance into multiple localized, more manageable balanced or near-balanced learning tasks through differentiated mechanisms. This approach effectively circumvents the decision biases prone to occur in single models when handling severely skewed data.

Although the aforementioned ensemble learning methods have demonstrated significant advantages across multiple imbalanced datasets, existing research lacks comprehensive comparative analyses of the classification performance across different ensemble models. Furthermore, evaluation metrics remain relatively limited, failing to fully assess a model's classification capabilities. Addressing these issues, the core contributions of this paper can be summarized as follows:

- 1) Feature selection is performed on the non-invasive prenatal testing dataset for pregnant women. Categorical data is encoded using label coding to match the model's input format.
- 2) Employ diverse resampling strategies to augment data and mitigate class imbalance effects. Visualize prediction outcomes via confusion matrices and utilize multi-dimensional classification performance metrics-including Recall and G-mean-to ensure evaluation accuracy and comprehensiveness.
- 3) Summarize comparative characteristics of different models based on experimental results to identify reliable solutions for enhancing the accuracy of non-invasive prenatal testing outcomes for female fetuses.

2. Research Methods

This section provides an overview of four mainstream models addressing data class imbalance-AdaBoost, EasyEnsemble, RUSBoost, and BalancedRandomForest-from different perspectives including data processing, model principles, and applications.

2.1. Data and its Preprocessing

This study utilized non-invasive prenatal testing data from 605 pregnant women across different regions. To enhance model training efficiency and generalization capability, the original dataset underwent preprocessing: feature columns weakly correlated with

chromosomal abnormality classification (e.g., height, weight, last menstrual period) were removed. Subsequently, categorical features were converted to label encoding, while numerical features underwent Z-score normalization. Finally, the processed data were randomly split into training and testing sets at an 8:2 ratio for subsequent model training and evaluation.

2.2. Ensemble Learning Models

The core concept of ensemble learning models is based on a voting mechanism that integrates the prediction results from multiple models to reduce prediction errors, thereby enhancing the model's robustness and generalization capability. Each model has distinct focuses in terms of its suitable application scenarios, design objectives, and core principles.

(1) Balanced Random Forest

Balanced Random Forest is an ensemble learning method specifically designed for class-imbalanced data. While retaining the dual randomness of traditional random forests (random data sampling and random feature selection), it introduces a key optimization to the Bagging sampling strategy: When constructing the training set for each decision tree, instead of performing Bootstrap sampling from the entire dataset, it conducts independent random sampling with replacement for each category, ensuring an equal number of samples are drawn from each category. This approach constructs category-balanced sub-training sets.

During the node splitting process of each decision tree, the model randomly selects a feature subset m (typically \sqrt{M}) from all M features, then identifies the optimal splitting feature and threshold based on this subset. This process is typically accomplished using splitting criteria such as maximizing information gain or minimizing Gini impurity.

For a node Q , its Gini impurity is calculated as follows:

$$G(Q) = 1 - \sum_{i=1}^C (p_i)^2 \quad (1)$$

Here, C represents the total number of categories, and p_i denotes the proportion of samples belonging to category i in node Q .

For a candidate split point that divides node Q into two subnodes Q_l and Q_r , the Gini impurity gain from this split is:

$$\Delta G = G(Q) - \frac{|Q_l|}{|Q|} G(Q_l) - \frac{|Q_r|}{|Q|} G(Q_r) \quad (2)$$

The algorithm traverses all possible split points within the feature subset, ultimately selecting the feature and threshold that maximize ΔG (i.e., minimize impurity) as the node's split rule.

Ultimately, the classification label \hat{y} for sample x is determined by aggregating the voting results from T decision trees:

$$\hat{y} = \text{mode} \{ h_1(x), h_2(x), \dots, h_T(x) \} \quad (3)$$

Here, $h_T(x)$ denotes the prediction result of the t -th decision tree for sample x , and mode represents the mode of the prediction results.

(2) Random Under-Sampling Boosting Algorithm (RUSBoost)

Unlike the SMOTE Boost algorithm, RUSBoost employs random undersampling to reduce the volume of data samples in the majority class, thereby achieving data class balance [13]. The core workflow of this algorithm is illustrated below:

Algorithm 1: RUSBoost Algorithm

Require: Training set $S = \{(x_i, y_i)\}, i = 1 \text{ to } N$, where $y_i \in \{0, 1\}$; Weak learning algorithm

WeakLearn;

Number of iterations T ; Sampling ratio r

Ensure: Final strong classifier

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

1. Initialize sample weights $w_i = 1/N$, for $i = 1 \text{ to } N$
2. **for** $t=1 \text{ to } T$ **do**
3. Create a subset S_t by random undersampling majority class examples from S according to sampling ratio r , while keeping all minority class examples
4. Train weak classifier h_t on the balanced subset S_t using *WeakLearn*
5. Calculate the error of h_t on the original training set

$$S: \epsilon_t = \sum_{i: h_t(x_i) \neq y_i} w_i$$

6. **if** $\epsilon_t > 0.5$ **then**

7. Break

8. **end if**

9. Set classifier weight: $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$

10. Update sample weights:

$$w_i = w_i \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i)) \text{ for } i = 1 \text{ to } N$$

11. Normalize weights: $w_i = w_i / \sum_{j=1}^N w_j$ for $i = 1 \text{ to } N$

12. **end for**

13. **return** $H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$
-

(3) Adaptive Boosting Algorithm (AdaBoost)

Adaptive Boosting Algorithm (AdaBoost) is a Boosting ensemble algorithm that iteratively trains multiple weak classifiers and combines them with weights. This approach enhances the model's focus on hard-to-classify samples, significantly improving classification accuracy. In a binary classification scenario: , where each weak classifier outputs $h_t(x)$ with weight α_t . Let ϵ_t denote the weighted error rate for the t-th sample. The classifier weight update formula is:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{4}$$

The iterative formula for sample weights can be expressed as:

$$Q_{t+1}(i) = Q_t(i) \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i)) \tag{5}$$

The final prediction formula for the binary classification label can be derived as follows:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t \cdot h_t(x) \right) \quad (6)$$

In small dataset scenarios, this model demonstrates excellent classification accuracy. However, when encountering outliers and noisy data, the limited data volume causes the model to assign disproportionately high classification weights to these instances, leading to misclassification of label results.

(4) EasyEnsemble

EasyEnsemble is an ensemble learning algorithm designed for class-imbalanced data. By combining adaptive undersampling with model integration, it effectively overcomes the information loss issue inherent in simple undersampling methods. Its core principle involves generating multiple balanced subsamples using the Bagging approach, training a strong classifier on each subset using the AdaBoost algorithm, and finally integrating the results from all strong classifiers.

Step 1. Generate balanced subsets: Independently generate T subsets $\{S_1, S_2, \dots, S_T\}$ from the majority class sample D_{maj} using bootstrap sampling with replacement. Each subset S_t contains the same number of samples as the minority class sample set $D_{min} (N_{min})$.

Step 2. Construct Balanced Training Sets: Merge each majority-class subset S_t with the entire minority-class sample D_{min} to form T balanced training subsets $\{D_1, D_2, \dots, D_T\}$, where $D_t = S_t \cup D_{min}$.

Step 3. Train AdaBoost Ensemble Model: Independently train an AdaBoost ensemble learning model $H_t(x)$ for each balanced training subset D_t . This constitutes the algorithm's core, with the mathematical process as follows:

1) Initialize sample weights: For each sample (x_i, y_i) in D_t , initialize the weight $w_1(i) = 1/N$, where N is the total number of samples in D_t .

2) For the k th iteration (out of K rounds):

a. Train a weak classifier (e.g., a shallow decision tree) using the current sample weight distribution.

b. Calculate the weighted error rate for this weak classifier:

$$\epsilon_{t,k} = \sum_{i=1}^N W_k(i) \cdot \mathbb{I}(h_{t,k}(c_i) \neq y_i) \quad (7)$$

c. Calculate the voting weight $\alpha_{t,k}$ for this weak classifier:

$$\alpha_{t,k} = \frac{1}{2} \left(\frac{1 - \epsilon_{t,k}}{\epsilon_{t,k}} \right) \quad (8)$$

d. Update sample weights by assigning different weights to correctly classified and incorrectly classified samples based on their classification status:

$$w_{k+1}(i) = \frac{w_k(i) \cdot \exp(-\alpha_{t,k} \cdot y_i \cdot h_{t,k}(x_i))}{Z_k} \quad (9)$$

where Z_k is the normalization factor, ensuring $\sum_i w_{k+1}(i) = 1$.

$$H_t(x) = \text{sign} \left(\sum_{k=1}^K \alpha_{t,k} \cdot h_{t,k}(x) \right) \quad (10)$$

Step4. Ensemble prediction for all base estimators: Combine the predictions from $H_t(x)$ AdaBoost models with T samples to output the final category label \hat{y} for sample x:

$$\hat{y} = \text{sign} \left(\sum_{t=1}^T H_t(x) \right) \quad (11)$$

To accommodate the scarcity of minority class samples and the data distribution characteristics in this study, and to optimize the model's adaptability to imbalanced data, experiments employed grid search and 10-fold cross-validation to obtain the optimal hyperparameter settings for each model, as shown in Table 1:

3. Experimental Results and Analysis

This section will conduct an in-depth analysis of the predictive performance of various ensemble learning models on NIPT testing data for female fetuses from a data-driven perspective, supplemented by visualizations. This analysis aims to identify the optimal classification model for chromosomal anomaly detection scenarios.

Table 1. Hyperparameter Configuration Table

Model	BalancedRandomForest	RUSBoost	AdaBoost	Easy-Ensemble
Parameters Settings	(1)sampling strategy: all (2)replacement: True (3)n-estimators: 400 (4)min weight fraction leaf: 0.0 (5)min samples split: 20 (6)min samples leaf: 2 (7)max leaf nodes: 100 (8)max features: 0.3 (9)max depth: 7 (10)criterion: entropy (11)class weight: None (12)bootstrap: False	(1)sampling strategy: majority (2)replacement: False (3)n-estimators: 300 (4)learning_rate: 0.001 (5)algorithm: SAMME.R	(1)n-estimators: 100 (2)learning_rate: 1.5 (3)estimator: DecisionTreeClassifier(max_depth=2, random_state=42) (4)algorithm: SAMME	(1)sampling strategy: all (2)replacement: True (3)n-jobs: -1 (4)n-estimators: 25
Ten-fold cross-validation on custom score	0.5418	0.4719	0.3707	0.5002

3.1. Model Evaluation Metrics

To quantify the classification performance of the model, this paper employs a comprehensive evaluation using multiple metrics, including accuracy, precision, recall, F1-score, area under the ROC curve (ROC-AUC), geometric mean (G-Mean), and area under the precision-recall curve (PR-AUC).

3.2. Experimental Results and Model Comparison

Based on the aforementioned evaluation metrics, this section applies the trained models to the test set for prediction. By systematically comparing the performance of each model in the chromosomal aneuploidy labeling classification task, we aim to identify classification strategies that demonstrate greater applicability and robustness in clinical practice. This approach provides reliable technical support for the early screening of fetal abnormalities in female fetuses.

Additionally, to conduct a more comprehensive evaluation of model performance, this paper further introduces two visualization tools: confusion matrices and radar charts. Confusion matrices provide an intuitive display of the model's classification performance across various categories, while radar charts clearly reveal the strengths and weaknesses of each model across different evaluation dimensions from a geometric perspective. These tools offer an intuitive and comprehensive basis for model selection, as detailed in Figures 1 and 2.

The quantitative evaluation of the model and the comparison of results are shown in Table 2

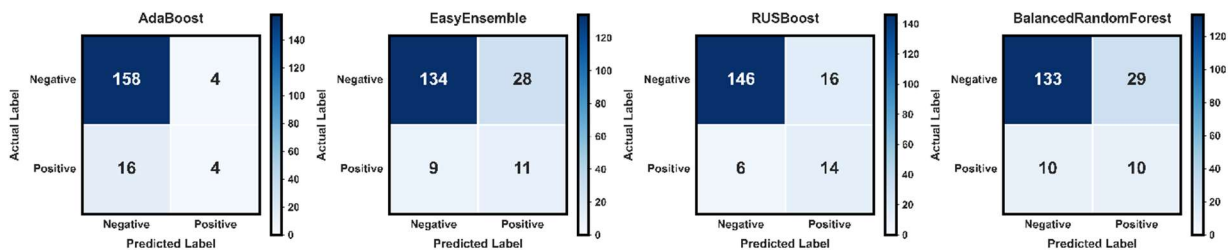


Fig.1 Confusion Matrix

Based on the data in the table above and the actual meanings of different indicators, a comprehensive performance analysis of the model can be derived:

(1) AdaBoost: High Specificity but Insufficient Recall

Although the AdaBoost model demonstrates excellent performance in accuracy (89.01%) and specificity (97.53%), its extremely low recall rate (20.00%) reveals a severe category bias issue. This occurs because the traditional AdaBoost algorithm, when handling imbalanced data, is easily dominated by majority class samples during training. Consequently, the model overemphasizes correctly classifying the majority class while neglecting the learning of minority class samples. In practical clinical applications, such a low recall rate implies that 80% of cases with chromosomal abnormalities would be missed, which is unacceptable in prenatal screening scenarios. Therefore, despite AdaBoost's overall good accuracy, it is unsuitable for direct application to the female fetal anomaly detection task in this study.

(2) RUSBoost: Optimal Overall Performance

RUSBoost effectively mitigates the negative impact of class imbalance by randomly undersampling majority-class samples in each iteration. Experimental results demonstrate that this model achieves the best performance across multiple key metrics: Recall: 70.00%, F1 Score: 0.5600, ROC-AUC: 0.8776, G-Mean: 0.7943. This outstanding performance demonstrates the effective integration of random undersampling with the Boosting framework, enabling the model to maintain high sensitivity toward abnormal samples while preserving strong

recognition capabilities for normal samples. Notably, RUSBoost exhibits a significant advantage over other models in recall rate, which is crucial for reducing the risk of clinical misdiagnosis. However, it should be noted that random undersampling may result in the loss of useful information within the majority class samples, which to some extent limits the potential for further improving the model's performance.

(3) EasyEnsemble: Balanced Performance but Lower Accuracy

EasyEnsemble mitigates information loss to some extent by constructing multiple under-sampled subsets and integrating multiple AdaBoost classifiers. This model achieves a recall rate of 55.00%, positioned between AdaBoost and RUSBoost, demonstrating relatively balanced classification characteristics. However, its precision rate is only 0.2821, the lowest among the four models. This indicates a high proportion of false positives among samples predicted as abnormal. In clinical applications, an excessively high false positive rate may lead to unnecessary follow-up examinations and psychological burden. Therefore, while EasyEnsemble improves recall, its overall performance still requires optimization.

(4) Balanced Random Forest: Robust but Not Outstanding in Performance

Balanced Random Forest ensures equitable attention to samples across different categories by applying balanced sampling during the training process of each decision tree.

However, experimental results indicate that this model performs at an average level across all metrics, achieving an accuracy of 78.57%, a recall rate of 50.00%, and an F1 score of 0.3390. This phenomenon may be attributed to the following factors: First, while the parallel ensemble strategy of random forests reduces variance, its advantages are less pronounced than sequential ensemble methods (e.g., Boosting) when handling highly imbalanced data. Second, the relatively limited sample size of this study's dataset (147 cases) may have prevented random forests from fully leveraging their strengths in processing large-scale data.

Statistical Significance Analysis

To verify whether the performance advantage of RUSBoost over other models is statistically significant, this study calculated the relative improvement in key metrics. Results show that compared to the second-best model EasyEnsemble, RUSBoost achieved a 27.27% increase in recall; a 50.23% improvement in F1 score; G-Mean by 17.76%; and ROC-AUC by 13.42%. These findings conclusively demonstrate RUSBoost's superiority in handling imbalanced NIPT data

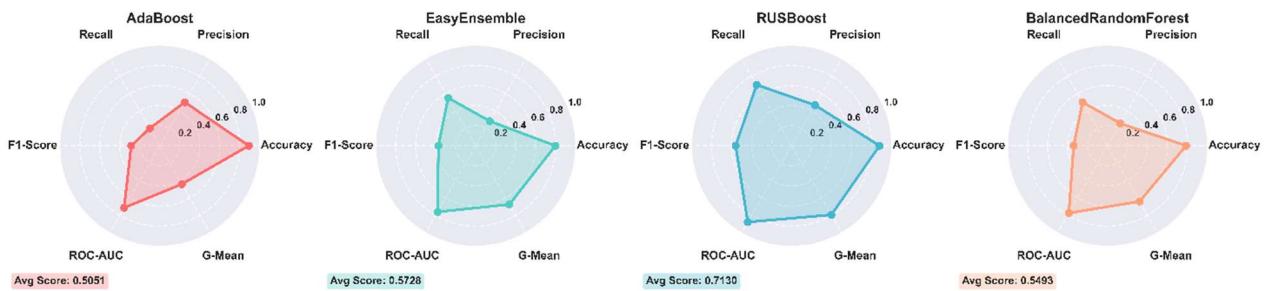


Fig.2 Performance Metrics Radar Chart

Table 2. Experimental Results Table

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	G-Mean	PR-AUC	Specificity
AdaBoost	0.89	0.50	0.20	0.29	0.71	0.44	0.38	0.98
EasyEnsemble	0.80	0.28	0.55	0.37	0.76	0.67	0.42	0.83
RUSBoost	0.88	0.47	0.70	0.56	0.88	0.79	0.55	0.90
BalanceRandomForest	0.79	0.26	0.50	0.34	0.77	0.64	0.40	0.82

3.3. Assessment of Clinical Application Value

From a clinical application perspective, this study further analyzed the practical value of the four models:

(1) Risk of missed diagnosis: RUSBoost achieved a recall rate of 70.00%, meaning it correctly identified 70 out of 100 cases with chromosomal abnormalities, resulting in a missed diagnosis rate of 30.00%. In contrast, AdaBoost exhibited a missed diagnosis rate as high as 80.00%, posing significant safety concerns in practical applications.

(2) False positive burden: Although RUSBoost's precision (46.67%) is lower than AdaBoost's (50.00%), this difference is clinically acceptable given its significantly higher recall rate. In prenatal screening scenarios, moderately increasing false positives to reduce missed diagnosis risk is a more reasonable strategy.

(3) Comprehensive Trade-off: As the geometric mean of true positive rate and true negative rate, G-Mean comprehensively reflects a model's overall performance across both sample categories. RUSBoost's G-Mean value (0.7943) significantly outperforms other models, indicating that it maintains strong recognition capability for normal samples while reducing false negative risks. This provides crucial reference value for clinical decision-making.

3.4. Experimental Conclusions

In the detection of chromosomal abnormalities in female fetuses, the RUSBoost model demonstrated the best overall performance. By combining random undersampling with a boosting framework, it effectively alleviated class imbalance, significantly outperforming other models in key metrics like recall and F1-score, while achieving a 70.00% recall rate and 87.91% accuracy to reduce clinical misdiagnosis risk. Although the traditional AdaBoost model showed excellent accuracy (89.01%) and specificity (97.53%), its extremely low recall (20.00%) made it unsuitable for this imbalanced data task. EasyEnsemble and BalancedRandomForest performed moderately-improving minority class recognition slightly but remaining far inferior to RUSBoost. Clinically, RUSBoost is recommended for early screening as it balances lower missed diagnosis risk and a reasonable false positive rate (46.67% precision), also offering methodological guidance for future work on highly imbalanced medical datasets.

4. Summary and Outlook

This study systematically analyzed the performance of four ensemble learning models in classifying NIPT data for female fetuses through a comprehensive comparison of multi-dimensional evaluation metrics. In medical diagnostic scenarios, model recall is a critical performance metric as it directly relates to the ability to identify abnormal cases. Higher recall effectively reduces the risk of missed diagnoses, preventing severe consequences from undetected fetal chromosomal abnormalities. This holds significant clinical importance for safeguarding maternal and fetal health. This study provides a scientific basis for model selection in early screening for chromosomal abnormalities.

References

- [1] Kotsopoulou I, Tsoplou P, Mavrommatis K, et al. Non-invasive prenatal testing (NIPT): limitations on the way to become diagnosis[J]. *Diagnosis*, 2015, 2(3): 141-158.
- [2] ayashankar S S, Nasaruddin M L, Hassan M F, et al. Non-invasive prenatal testing (NIPT): reliability, challenges, and future directions[J]. *Diagnostics*, 2023, 13(15): 2570.
- [3] Salmi M, Atif D, Oliva D, et al. Handling imbalanced medical datasets: review of a decade of research[J]. *Artificial intelligence review*, 2024, 57(10): 273.
- [4] He H, Garcia E A. Learning from imbalanced data[J]. *IEEE Transactions on knowledge and data engineering*, 2009, 21(9): 1263-1284.

- [5] Mohammed R, Rawashdeh J, Abdullah M. Machine learning with oversampling and undersampling techniques: overview study and experimental results[C]//2020 11th international conference on information and communication systems (ICICS). IEEE, 2020: 243-248.
- [6] Dong X, Yu Z, Cao W, et al. A survey on ensemble learning[J]. *Frontiers of Computer Science*, 2020, 14(2): 241-258.
- [7] Hasanin T, Khoshgoftaar T. The effects of random undersampling with simulated class imbalance for big data[C]//2018 IEEE international conference on information reuse and integration (IRI). IEEE, 2018: 70-79.
- [8] Mani I, Zhang I. kNN approach to unbalanced data distributions: a case study involving information extraction[C]//Proceedings of workshop on learning from imbalanced datasets. United States: ICML, 2003, 126(1): 1-7.
- [9] He H, Bai Y, Garcia E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]//2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). Ieee, 2008: 1322-1328.
- [10] Liu T Y. Easyensemble and feature selection for imbalance data sets[C]//2009 international joint conference on bioinformatics, systems biology and intelligent computing. IEEE, 2009: 517-520.
- [11] Raghuvanshi B S, Shukla S. Classifying imbalanced data using BalanceCascade-based kernelized extreme learning machine[J]. *Pattern Analysis and Applications*, 2020, 23(3): 1157-1182.
- [12] Chawla N V, Lazarevic A, Hall L O, et al. SMOTEBoost: Improving prediction of the minority class in boosting[C]//European conference on principles of data mining and knowledge discovery. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003: 107-119.
- [13] Seiffert C, Khoshgoftaar T M, Van Hulse J, et al. RUSBoost: A hybrid approach to alleviating class imbalance[J]. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, 2009, 40(1): 185-197.