Hyperspectral Image Classification with Imbalanced Data Based on a Semi-Supervised Learning Algorithm

Nian Zhang
Department of Electrical and Computer Engineering
University of the District of Columbia
Washington, D.C., USA
nzhang@udc.edu

Paul Cotae
Department of Electrical and Computer Engineering
University of the District of Columbia
Washington, D.C., USA
pcotae@udc.edu

***Abstract* - Hyperspectral image classification is often challenged by severe class imbalance and limited labeled samples, which significantly degrade the performance of conventional supervised learning methods. To address these issues, this paper proposes a semi-supervised learning framework that iteratively incorporates informative unlabeled samples based on neighborhood consistency to improve minority-class representation. Unlike synthetic oversampling techniques such as SMOTE, the proposed method leverages the intrinsic data distribution without generating artificial samples.**

Extensive experiments under varying class imbalance ratios demonstrate that the proposed approach consistently outperforms baseline preprocessing methods in terms of Recall, F1-score, G-mean, and AUC. Paired *t*-tests further confirm that the improvements over SMOTE are statistically significant, particularly under severe imbalance conditions. Moreover, the proposed framework achieves stable performance across different base classifiers and exhibits low sensitivity to hyperparameter settings, highlighting its robustness and practical applicability.

Index Terms - Hyperspectral Image Classification, Hyperspectral Imaging, Semi-Supervised Learning, Minority-Class Detection, Class Imbalance, Imbalanced Data, Pseudo-Labeling.

I. INTRODUCTION

Hyperspectral image (HSI) classification has become a fundamental task in remote sensing due to its wide range of applications, including land cover mapping, environmental monitoring, and precision agriculture [1]-[3]. By capturing hundreds of contiguous spectral bands, hyperspectral sensors provide rich spectral information that enables fine-grained discrimination among different materials. However, the high dimensionality of hyperspectral data, together with limited labeled samples, poses significant challenges to the development of reliable classification models [4].

One critical yet often overlooked challenge in hyperspectral image classification is class imbalance [5]-[6]. In real-world scenarios, the number of samples available for different land-cover classes is typically highly uneven, where minority classes may contain only a small fraction of labeled samples compared with majority classes. This imbalance can severely bias conventional supervised learning algorithms toward majority classes, resulting in poor recognition performance for minority classes. Although various data-level and algorithm-level imbalance-handling techniques have been proposed, many of them rely on extensive labeled data or synthetic sample generation, which may distort the original data distribution and degrade classification performance.

Semi-supervised learning (SSL) has emerged as a promising paradigm for hyperspectral image classification by exploiting abundant unlabeled data in conjunction with limited labeled samples [7]. By leveraging the intrinsic structure of hyperspectral data, SSL methods are able to improve generalization performance without requiring additional manual labeling. Nevertheless, existing SSL-based approaches often remain sensitive to class imbalance and hyperparameter selection. In particular, inappropriate utilization of unlabeled samples may reinforce majority-class dominance, further deteriorating minority-class recognition under imbalanced conditions.

Motivated by these observations, this paper proposes a semi-supervised learning algorithm for hyperspectral image classification with imbalanced data. The proposed method effectively integrates labeled and unlabeled samples through an iterative learning strategy that enhances minority-class representation while preserving class boundary discrimination. Unlike traditional resampling-based approaches, the proposed framework does not generate synthetic samples; instead, it exploits the intrinsic data structure to improve classification robustness. As a result, the proposed method is naturally suited to imbalanced hyperspectral classification scenarios.

This paper follows a methodologically bounded research framework aimed at developing and validating a semi-supervised learning approach for hyperspectral image classification under class imbalance while preserving the intrinsic data distribution. The study includes (i) formulation of a cost-sensitive semi-supervised algorithm, (ii) experimental validation under varying imbalance ratios, (iii) evaluation across multiple commonly used classifiers, and (iv) robustness analysis with respect to key hyperparameters. The methodology and experimental design in Sections III-V are structured to satisfy these validation criteria and to provide a systematic, reproducible assessment of the proposed framework rather than a collection of ad hoc experiments.

The remainder of this paper is organized as follows. Section II reviews related work on hyperspectral image classification, class imbalance learning, and semi-supervised methods. Section III presents the proposed semi-supervised learning framework in detail. Section IV describes the experimental setup and evaluation metrics.

Section V discusses the experimental results and analysis. Finally, Section VI concludes the paper and outlines future research directions.

II. RELATED WORK

A. Hyperspectral Image Classification

Hyperspectral image (HSI) classification has been extensively studied in the remote sensing community due to the rich spectral information provided by hyperspectral sensors. Early approaches primarily relied on conventional supervised classifiers, such as support vector machines, k-nearest neighbors, and random forests, which demonstrated strong performance under sufficient labeled data. More recently, advanced feature extraction and representation learning techniques have been introduced to further improve classification accuracy. Despite these advances, most existing HSI classification methods implicitly assume balanced class distributions and sufficient labeled samples, an assumption that rarely holds in real-world remote sensing applications.

B. Learning with Imbalanced Data

Class imbalance is a well-known challenge in machine learning and has motivated a wide range of imbalance-handling techniques. Data-level methods, such as random oversampling and synthetic sample generation, aim to rebalance class distributions by augmenting minority-class samples. Among them, synthetic oversampling techniques have been widely adopted due to their simplicity and effectiveness in certain scenarios. Algorithm-level approaches, including cost-sensitive learning and decision boundary adjustment, modify learning objectives to penalize misclassification of minority classes. However, in hyperspectral image classification, these methods often suffer from limited labeled data availability or risk altering the original data distribution, which may negatively affect classification performance.

C. Semi-Supervised Learning for Hyperspectral Images

Semi-supervised learning (SSL) has emerged as an effective framework for hyperspectral image classification by leveraging abundant unlabeled samples to compensate for limited labeled data. Existing SSL approaches for HSI include graph-based methods, self-training strategies, and manifold learning techniques, which exploit the intrinsic structure of hyperspectral data to improve classification accuracy. While these methods have shown promising results, most of them are not explicitly designed to address class imbalance. In imbalanced scenarios, unlabeled samples are often dominated by majority classes, which may lead to biased pseudo-label propagation and further degrade minority-class recognition.

D. Motivation and Research Gap

Although hyperspectral image classification, imbalanced learning, and semi-supervised learning have each been studied independently, their integration remains insufficiently explored. Existing imbalance-handling techniques typically rely on supervised settings or synthetic data generation, whereas many SSL-based HSI methods do not explicitly consider the effects of class imbalance or parameter sensitivity. Consequently, there is a clear need for a semi-supervised learning framework that effectively exploits unlabeled data while enhancing minority-class representation and maintaining robustness under varying imbalance conditions. This gap motivates the proposed semi-supervised learning algorithm for hyperspectral image classification with imbalanced data.

III. PROPOSED METHODOLOGY

In this section, we describe the proposed semi-supervised learning (SSL) framework for hyperspectral image (HSI) classification with imbalanced training data. The main goal is to generate pseudo-labeled samples for minority classes from the unlabeled data, thereby balancing the training set and improving classifier performance on underrepresented classes.

A. Problem Formulation

Let $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$ denote a hyperspectral dataset consisting of a small labeled set $\mathcal{D}_l = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_l}$ and a large unlabeled set $\mathcal{D}_u = \{\mathbf{x}_j\}_{j=1}^{N_u}$, where $\mathbf{x} \in \mathbb{R}^d$ represents the spectral feature vector and $y \in \{1, \dots, C\}$ denotes the class label. In practical hyperspectral image classification tasks, the labeled samples are typically severely imbalanced, with minority classes containing significantly fewer samples than majority classes.

The objective of this work is to learn a classifier $f(\cdot)$ that achieves robust classification performance under imbalanced conditions by effectively exploiting unlabeled samples, while avoiding bias toward majority classes.

B. Proposed Semi-Supervised Learning Framework

The proposed method follows an iterative semi-supervised learning framework that alternates between supervised training on labeled data and selective incorporation of unlabeled samples. The core idea is to progressively enhance minority-class representation by leveraging intrinsic data structure rather than generating synthetic samples. The overall procedure consists of the following main stages:

1. Initial supervised training using the available labeled dataset $\mathcal{D}_l$;
2. Neighborhood-based candidate selection from the unlabeled pool $\mathcal{D}_u$;
3. Pseudo-label assignment for selected unlabeled samples;
4. Model update using the expanded labeled set;
5. Iteration until convergence or a predefined stopping criterion is reached.

This design enables the classifier to gradually improve decision boundaries while mitigating the dominance of majority classes.

C. Neighborhood-Based Unlabeled Sample Selection

To exploit the local structure of hyperspectral data, a neighborhood-based strategy is employed to identify informative unlabeled samples. For each labeled sample, a set of K nearest neighbors is identified from the unlabeled pool based on spectral similarity. This neighborhood construction allows the algorithm to focus on samples that are more likely to share the same class characteristics.

To control the influence of unlabeled data, a parameter q is introduced to determine the number of unlabeled samples selected for pseudo-labeling in each iteration. By limiting the number of selected samples, the proposed method avoids overwhelming the training process with potentially noisy pseudo-labels, especially under severe class imbalance.

D. Pseudo-Label Assignment and Model Update

For each selected unlabeled sample, a pseudo-label is assigned using the current classifier. These pseudo-labeled samples are then combined with the original labeled dataset

to form an augmented training set. The classifier is retrained on this expanded dataset to refine its decision boundaries.

Unlike traditional self-training methods that indiscriminately include pseudo-labeled samples, the proposed framework emphasizes controlled and structured integration of unlabeled data. This strategy reduces error accumulation and prevents majority-class samples from dominating the learning process.

E. Robustness to Imbalance and Hyperparameter Selection

A key advantage of the proposed method lies in its robustness to both class imbalance and hyperparameter selection. The neighborhood size K governs local structural exploitation, while the parameter q controls the degree of unlabeled sample participation. As demonstrated in the experimental analysis, the proposed framework exhibits stable performance across a wide range of parameter settings, indicating that it does not rely on fine-grained hyperparameter tuning. Furthermore, by avoiding synthetic data generation and relying on intrinsic data structure, the proposed method preserves the original hyperspectral data distribution, which is critical for reliable classification performance.

F. Proposed Algorithm

The proposed semi-supervised learning algorithm for hyperspectral image classification with imbalanced data can be summarized as follows:

1. Train an initial classifier on the labeled dataset D_l using a cost-sensitive mechanism to emphasize the minority class. The cost matrix is defined as:

$$\text{Cost} = \begin{bmatrix} 0 & \frac{N_{\text{pos}}}{N_{\text{neg}}} \\ \frac{N_{\text{neg}}}{N_{\text{pos}}} & 0 \end{bmatrix}$$

where:

- N_{pos} is the number of minority class samples,
- N_{neg} is the number of majority class samples,
- $\text{Cost}(i, j)$represents the penalty for predicting a sample of true class i as class j.
- While SVM is used by default, other classifiers (e.g., Random Forest, KNN) can also be employed if they can incorporate class weighting or output probabilities.

2. Predict pseudo-labels for the unlabeled pool D_u and compute confidence scores s_i for each sample. The confidence score indicates the likelihood of the sample belonging to the predicted class. For SVM, it can be computed as:

$$s_i = \frac{1}{1 + \exp(-f(x_i))}, f(x_i) = \mathrm{w}^T \mathrm{x}_i + b$$

where:

- $s_i \in [0,1]$is the confidence score of sample x_i,
- $f(x_i)$ is the decision value,
- $\mathrm{w}$and b are the weight and bias learned by the cost-sensitive classifier.
- For other classifiers, s_i corresponds to the predicted class probability or confidence output.

3. Apply adaptive, class-specific thresholds to select reliable pseudo-labeled samples:

$$\tau_{\text{pos}} = \max(\tau_{\text{min}}, \tau_0(1-\alpha \cdot \text{imbalanceRatio}))$$

where:

- τ_{pos} and τ_{neg} are thresholds for minority and majority classes, respectively,
- τ_0 is the base confidence threshold,
- τ_{min} is the minimum allowed threshold for the minority class,
- α is a scaling factor,
- $\text{imbalanceRatio} = N_{\text{pos}}/N_{\text{neg}}$.

4. Select a controlled number of high-confidence samples per iteration that satisfy $s_i \geq \tau_{\text{class}}$, to prevent overfitting or bias amplification.
5. Augment the labeled dataset with the selected pseudo-labeled samples and remove them from the unlabeled pool.
6. Retrain the classifier on the augmented dataset and repeat steps 2-5 iteratively until convergence or until no high-confidence samples remain.

This framework is classifier-agnostic and can be integrated with commonly used supervised classifiers, making it flexible for practical hyperspectral image classification tasks.

IV. EXPERIMENTAL SETUP

A. Dataset

The proposed method is evaluated on a benchmark hyperspectral dataset, PaviaU [8] that is widely used in hyperspectral image classification research. The dataset contains spatial dimensions of 610 × 340 pixels and 103 spectral bands, and covers nine land-cover classes with highly imbalanced sample distributions. To ensure fair evaluation, unlabeled samples are drawn from the same scene while labeled samples are randomly selected according to predefined imbalance ratios.

B. Experimental Design

Three groups of experiments are conducted to comprehensively evaluate the proposed semi-supervised learning framework:

1. Imbalance Variation Analysis: To investigate the effectiveness of the proposed method under different imbalance conditions, multiple training sets with varying imbalance ratios are generated. The performance of the proposed method is compared with commonly used imbalance-handling baseline methods across these scenarios, and paired *t*-tests are conducted to statistically assess the significance of the observed performance differences.
2. Classifier Impact Analysis: To examine the generality of the proposed framework, multiple supervised classifiers are employed as base learners. This experiment evaluates whether the proposed semi-supervised strategy consistently improves performance regardless of classifier choice.
3. Hyperparameter Sensitivity Analysis: The robustness of the proposed method with respect to hyperparameter selection is analyzed by varying the neighborhood size K and the unlabeled sample selection parameter q. Mean F1-score is used to visualize performance trends under different parameter combinations.

C. Baseline Methods and Classifiers

The proposed method is compared with representative baseline approaches designed to handle imbalanced data. Data-level imbalance-handling techniques are adopted as baseline methods to rebalance the training set prior to classifier learning. For classifier impact analysis, commonly used supervised classifiers are employed as base learners, including margin-based and tree-based models. All methods are implemented using the same feature representations and evaluation protocols to ensure fair comparison.

D. Evaluation Metrics

Given the imbalanced nature of the classification task, overall accuracy alone is insufficient to reflect classification performance. Therefore, multiple evaluation metrics are adopted, with a particular emphasis on minority-class performance. Specifically, recall and F1-score are used as the primary evaluation metrics, as they provide more informative assessment under imbalanced conditions. All reported results are averaged over multiple independent runs to reduce the impact of random sampling.

E. Implementation Details

All experiments are conducted using a consistent experimental environment. For each experimental setting, labeled samples are randomly selected according to the specified imbalance ratios, while the remaining samples are treated as unlabeled data. The same training and testing splits are used for all compared methods. Hyperparameters are set consistently across experiments, and default parameter values are adopted unless otherwise specified. For robustness analysis, the neighborhood size K and the parameter q are varied within predefined ranges.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. Imbalance Variation Analysis

This experiment validates the proposed framework under varying class imbalance conditions to assess its effectiveness and robustness. The first group of experiments evaluates the performance of the proposed semi-supervised learning method under varying class imbalance conditions. Figs. 1-4 present the classification performance of the proposed semi-supervised learning method and the SMOTE-based baseline under varying imbalance ratios. Four commonly used metrics for imbalanced classification, recall, G-mean, F1-score, and AUC, are reported to provide a comprehensive evaluation.

As shown in Fig. 1, the proposed method consistently achieves higher average recall than the baseline method across all imbalance ratios. The performance gap is particularly evident under severe imbalance conditions, indicating that the proposed framework is more effective in enhancing minority-class recognition. Fig. 2 further confirms this observation through the G-mean metric, which jointly reflects the classification performance of both majority and minority classes. The proposed method maintains a higher G-mean across all imbalance ratios, suggesting a better balance between class-wise accuracies. In contrast, the SMOTE-based baseline exhibits larger performance fluctuations, especially under severe imbalance, which may be attributed to the limitations of synthetic sample generation in complex hyperspectral feature spaces.

Similar trends can be observed in Fig. 3, where the proposed method consistently outperforms the baseline in terms of F1-score. The improvement in F1-score indicates that the proposed framework effectively balances precision and recall for minority classes, rather than achieving gains at the expense of increased false positives. Moreover, the smaller variance observed for the proposed method suggests more stable performance across different experimental runs. Finally, Fig. 4 reports the average AUC results, which further demonstrate the robustness of the proposed approach. Across all imbalance ratios, the proposed method achieves higher AUC values than the SMOTE-based baseline, indicating improved overall discriminative capability. Notably, the performance gap remains consistent as the imbalance ratio increases, highlighting the scalability of the proposed method under varying degrees of class imbalance.

Overall, these results demonstrate that the proposed semi-supervised learning framework consistently outperforms traditional imbalance-handling techniques across multiple evaluation metrics. By exploiting intrinsic data structure through unlabeled samples rather than relying on synthetic data generation, the proposed method provides a more robust and reliable solution for hyperspectral image classification under imbalanced conditions.

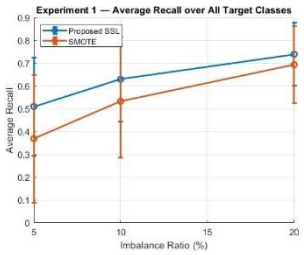

Fig. 1. Average recall over all target classes under different imbalance ratios.

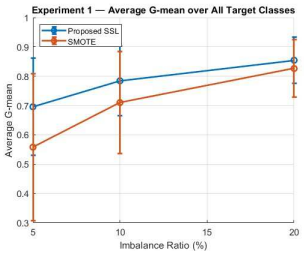

Fig. 2. Average G-mean over all target classes under different imbalance ratios.

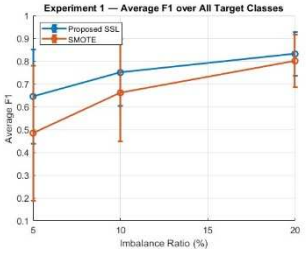

Fig. 3. Average F1-score over all target classes under different imbalance ratios.

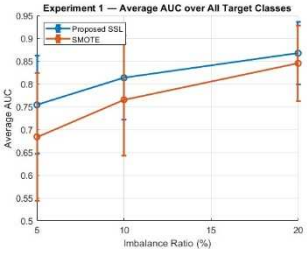

Fig. 4. Average AUC over all target classes under different imbalance ratios.

To further validate whether the observed improvements are statistically significant, a paired t-test is conducted on the Recall metric between the proposed semi-supervised learning method and the SMOTE baseline across different imbalance ratios. The results are illustrated in Fig. 5.

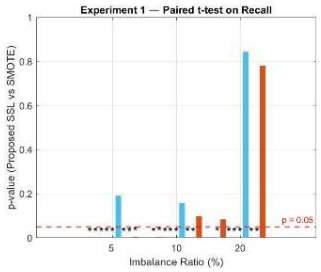

Fig. 5. Paired t-test results on Recall for the proposed SSL method versus the SMOTE baseline. Each bar shows the p-value for a specific target class at a given imbalance ratio (5%, 10%, 20%). The dashed red line indicates the significance threshold (p = 0.05). Asterisks (*) denote p-values below this threshold, representing statistically significant improvements for that class.

As shown in Fig. 5, the paired *t*-test on Recall demonstrates a clear and statistically significant advantage of the proposed SSL method over the SMOTE baseline, particularly under severe class imbalance. At imbalance ratios of 5% and 10%, the vast majority of class-wise comparisons yield *p*-values well below the 0.05 threshold, with many approaching zero. This consistent statistical significance across classes provides strong evidence that the observed Recall improvements are systematic rather than due to random variation. These results confirm the robustness of the proposed SSL framework in learning discriminative representations when minority-class samples are extremely scarce. In contrast, when the imbalance ratio increases to 20%, corresponding to a relatively less skewed class distribution, the *p*-values rise above the significance threshold for most classes. This indicates that the performance difference between the proposed SSL method and SMOTE becomes smaller and less consistent as the dataset approaches a more balanced regime. Such behavior is expected, as the reliance on sophisticated imbalance-handling strategies diminishes when sufficient labeled samples are available.

Overall, these results highlight that the proposed SSL method is particularly well suited for highly imbalanced learning scenarios, where it delivers statistically significant and reliable gains in Recall over conventional oversampling techniques. While its advantage naturally decreases as class imbalance is alleviated, the method's strong performance under the most challenging conditions underscores its practical value for real-world imbalanced classification tasks.

B. Classifier Assessment

This experiment evaluates the generality of the proposed framework by assessing its performance across multiple widely used classifiers. To evaluate the generality of the proposed framework, experiments are conducted using different base classifiers, including support vector machines (SVM), random forests (RF), and k-nearest neighbors (KNN). Figs. 6 and 7 present the minority-class recall before and after applying the proposed semi-supervised learning framework, along with the corresponding performance improvements.

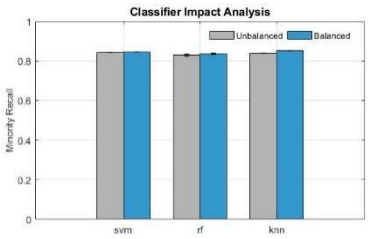

Fig. 6. Minority-class recall obtained with different base classifiers before and after applying the proposed semi-supervised learning framework.

Fig. 7. Improvement in minority-class recall achieved by the proposed semi-supervised learning framework across different base classifiers.

As shown in Fig. 6, incorporating the proposed semi-supervised learning strategy consistently improves minority-class recall across all evaluated classifiers. Although the absolute recall values vary among different classifiers, the proposed method provides noticeable performance gains in all cases. This indicates that the proposed framework effectively enhances minority-class representation regardless of the underlying classifier.

Fig. 7 further quantifies the recall improvement achieved by the proposed method. Among the evaluated classifiers, KNN exhibits the largest performance gain, followed by RF and SVM. This observation suggests that classifiers relying more heavily on local neighborhood information may benefit more from the proposed semi-supervised strategy. Nevertheless, the consistent improvement across all classifiers demonstrates that the effectiveness of the proposed framework is not restricted to a specific learning model.

Overall, these results confirm that the proposed semi-supervised learning framework is classifier-agnostic and can be seamlessly integrated with different supervised classifiers to improve minority-class recognition in hyperspectral image classification tasks.

C. Hyperparameter Optimization

This experiment examines the robustness of the proposed framework with respect to key hyperparameters to determine its sensitivity and stability across different parameter settings. To evaluate the robustness of the proposed semi-supervised learning framework, we analyze the sensitivity of two key hyperparameters: the neighborhood size $K$and the number of selected unlabeled samples q.

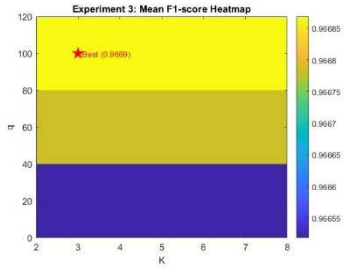

Fig. 8. Mean F1-score heatmap of the proposed semi-supervised learning method with respect to the neighborhood size K and the unlabeled sample selection size q. The best performance is highlighted.

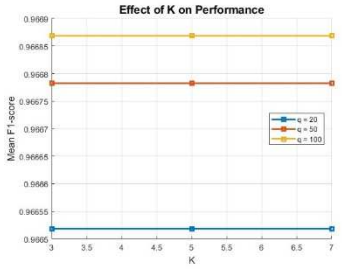

Fig. 9. Effect of neighborhood size K on the mean F1-score under different values of q.

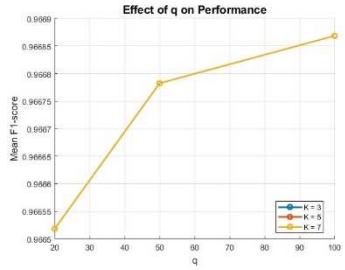

Fig. 10. Effect of unlabeled sample selection size q on the mean F1-score under different values of K.

Fig. 8 presents the mean F1-score heatmap over a wide range of parameter combinations. As shown in Fig. 8, the proposed method achieves consistently high performance across most parameter settings, indicating strong robustness to hyperparameter variations. The best performance of 0.9669 is observed around $K = 3$ and $q = 100$; however, the performance differences across neighboring configurations remain marginal, suggesting that the method does not rely on fine-grained parameter tuning.

The effect of the neighborhood size K is further illustrated in Fig. 9. For different values of q, the mean F1-score remains nearly constant as K increases from 3 to 7. This indicates that the proposed method is relatively insensitive to the choice of neighborhood size, provided that a reasonable local structure is preserved.

Fig. 10 analyzes the influence of the unlabeled sample selection size q. A slight performance improvement is observed as q increases, reflecting the benefit of incorporating more informative unlabeled samples into the training process. Nevertheless, the performance gain gradually saturates, demonstrating that the proposed framework can effectively exploit unlabeled data without requiring excessively large unlabeled pools. Overall, these results confirm that the proposed semi-supervised learning approach exhibits stable and reliable performance across a broad range of hyperparameter settings, making it practical for real-world hyperspectral image classification tasks where parameter tuning is often challenging.

VI. CONCLUSIONS

This paper presented a semi-supervised learning framework for imbalanced hyperspectral image classification that enhances minority-class representation by selectively exploiting unlabeled samples and intrinsic data structure, without relying on synthetic data generation.

Extensive experimental evaluations demonstrate that the proposed method consistently outperforms widely used baseline approaches across multiple imbalance ratios and base classifiers, with particularly pronounced gains in minority-class Recall and F1-score. Paired statistical significance tests further confirm that these improvements are not due to random variation, especially under severe class imbalance, highlighting the reliability and effectiveness of the proposed framework. In addition, hyperparameter sensitivity analysis shows that the method maintains stable performance over a broad range of parameter settings, substantially reducing the need for careful manual tuning. Future work will focus on extending the proposed framework to more complex and large-scale hyperspectral scenes, as well as incorporating advanced feature representation and deep learning techniques to further enhance performance under extreme imbalance and limited-label scenarios.

ACKNOWLEDGMENT

This work was supported and conducted under National Science Foundation Award #2401880.

REFERENCES

[1] X. Cui and L. Zhang, “DEMUNet: Dual encoder mamba U-Net for hyperspectral image classification,” in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 99-113, 2026, doi: 10.1109/JSTARS.2025.3631311.

[2] Z. Khan, N. Dey, K. Kathiravan, S. -C. Yoon and S. M. Bhandarkar, "Spatial-Spectral Transformer With Patch-Local Mixed-Axis 2-D Rotary Position Embedding for Hyperspectral Image Classification," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 1910-1925, 2026.

[3] Y. Cui, M. Jia, L. Wang, S. Gao, L. Chen and C. Zhao, "Hyperspectral Image Classification Based on Normalized Grouping Fused Comprehensive Relative Position Transformer Network," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 19, pp. 1515-1532, 2026.

[4] Y. Wang *et al*., “Bidirectional stacking ensemble curriculum learning for hyperspectral image imbalanced classification with noisy labels,” in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-15, 2025.

[5] P. Cotae, N. Zhang, and O. Obioha-Val, “A decision-making framework for addressing the imbalanced learning problem in standoff detection,” *2025 American Society for Engineering Education Northeast Section (ASEE-NE) Conference*, Bridgeport, CT, March 22, 2025.

[6] N. Zhang, S. Rouamba, and P. Cotae, “Imbalanced data classification using transfer learning and oversampling,” *13th International Conference on Intelligent Control and Information Processing (ICICIP 2025)*, Muscat, Oman, February 6-11, 2025.

[7] H. Sun *et al*., “Class-aware consistency learning for open-set semi-supervised hyperspectral image classification,” in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-15, 2025.

[8] Hyperspectral Remote Sensing Scenes: https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes