frontiers | Frontiers in Remote Sensing

TYPE Original Research
PUBLISHED 02 October 2025
DOI 10.3389/frsen.2025.1685140

OPEN ACCESS

EDITED BY
Fulin Luo,
Chongqing University, China

REVIEWED BY
Zhang Ying,
Hunan University of Technology, China
Jing Yao,
Chinese Academy of Sciences (CAS), China
Nabila Chergui,
University Ferhat Abbas of Setif, Algeria
Chuan Fu,
Chongqing University, China

*CORRESPONDENCE
Lingling Li,
lilingling@sdaep.com
Fernando J. Aguilar,
faguilar@ual.es

RECEIVED 19 August 2025
ACCEPTED 15 September 2025
PUBLISHED 02 October 2025

CITATION
Wang Y, Li L, Xu M, Liu S, Yasir M, Aguilar MÁ and Aguilar FJ (2025) AU-super: superpixel scale optimization and training data augmentation strategy for hyperspectral image classification. *Front. Remote Sens.* 6:1685140.
doi: 10.3389/frsen.2025.1685140

AU-super: superpixel scale optimization and training data augmentation strategy for hyperspectral image classification

Yiran Wang[1], Lingling Li[2]*, Mingming Xu[1], Shanwei Liu[1], Muhammad Yasir[1], Manuel Á. Aguilar[3] and Fernando J. Aguilar[3]*

[1]College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China, [2]Shandong Academy for Environmental Planning, Key Laboratory of Land and Sea Ecological Governance and Systematic Regulation, Ministry of Ecology and Environment, Jinan, China, [3]Department of Engineering, Research Centre CIAIMBITAL, University of Almería, Almería, Spain

The spectral information of each pixel in hyperspectral images contains valuable information about object properties, although accurate labeling is required in supervised classification to guide the model in distinguishing different land cover types. However, labeling data for hyperspectral images is difficult to obtain, especially in complex or remote areas. This results in a shortage of labeled samples, which prevents the model from fully learning the features of different classes. To overcome this challenge, this work proposes a hyperspectral image classification method, called AU-Super, that combines adaptive superpixel scale selection, superpixel label expansion, and data augmentation. First, an adaptive method is developed to determine an appropriate superpixel segmentation scale based on feature values, thereby ensuring that superpixel segmentation effectively captures the spatiospectral information of the image. Second, feature extraction is performed at the previously estimated superpixel scale. Third, pixel labels are converted to superpixel labels to reduce the effects of labeling noise during the training process. Furthermore, superpixel-level label-based data augmentation techniques are introduced to mitigate the problem of under-labeled patterns. The comparative results against various state-of-the-art algorithms demonstrate that AU-Super-RF consistently achieves superior performance across multiple accuracy metrics. Under few-shot training scenarios (with only 1–10 samples per class) on the Indian Pines, Salinas, and Pavia University datasets, it improves overall accuracy by 3%–7%, average accuracy by 2%–6%, and the Kappa coefficient by 3%–8%, highlighting the robustness and practical utility of the method.

KEYWORDS

hyperspectral remote sensing, image classification, superpixel segmentation, data augmentation, superpixel labeling

1 Introduction

Hyperspectral refers to an imaging technique that captures highly detailed spectral information over a wide spectral range (Gan et al., 2024). Compared with traditional RGB and multispectral images, hyperspectral images (HSI) scan the scene across hundreds of narrow spectral bands (Lassalle et al., 2023), offering significant advantages in high-precision feature identification and target detection because they provide richer spectral

information than conventional images (Liu et al., 2023; Xie et al., 2024). This unique property gives hyperspectral images a clear advantage in remote sensing classification (O'Shea et al., 2023). Regarding hyperspectral classification, the hyperspectral image is analyzed and processed to assign each pixel or superpixel to a specific class, using the spectral information of each pixel to distinguish between different types of land cover or objects (Jia et al., 2024). However, obtaining labeled data for hyperspectral images can be challenging in complex or remote areas where manual labeling is difficult and sample size is limited (Zhao et al., 2023). Furthermore, annotators may struggle to accurately identify certain land cover types, especially when the spectral characteristics are very similar (Zhang Q. et al., 2022). This label noise and annotation errors can negatively impact classification performance. Researchers' proposed solutions to address these problems fall primarily into two categories: dimension reduction and data augmentation.

Dimensionality reduction methods are often part of feature extraction. Common methods include principal component analysis (PCA) (Maćkiewicz and Ratajczak, 1993), linear discriminant analysis (LDA) (Xanthopoulos et al., 2013), and independent component analysis (ICA) (Lee, 1998). These methods can effectively reduce dimensionality, extract important spectral features, improve class separability, and optimize classification performance. In the case of insufficient samples, dimensionality reduction helps eliminate redundant and noisy features, allowing the classification model to focus more on the key features that distinguish land cover types, thereby improving classification accuracy (Zhang et al., 2022d). Deep learning-based feature extraction methods can also improve the adaptability of classifiers to limited samples (Yasir et al., 2023; Maffei et al., 2020; Liang et al., 2023). More recently, advanced architectures such as graph neural networks (GNNs) and Transformers have been introduced into hyperspectral image analysis (Zhang X.et al., 2023; Scheibenreif et al., 2023; Sun et al., 2024), showing promising performance in few-shot learning and lightweight modeling scenarios. In addition, the introduction of attention mechanisms offers new technical support for extracting and integrating different features (Hu et al., 2022; Zhou et al., 2023; Zhang et al., 2024). However, both traditional dimensionality reduction methods and deep learning-based methods are pixel-level approaches with high computational complexity that are sensitive to noise and may not effectively capture spatial contextual information and inter-pixel relationships, which are crucial factors for improving classification accuracy in complex scenes (Novelli et al., 2016; Aguilar et al., 2018).

On the other hand, data augmentation increases the diversity and number of samples, which alleviates the problem of sample imbalance, especially when minority class samples are missing. Common data augmentation techniques include RCSMOTE (Soltanzadeh and Hashemzadeh, 2021), which balances the class distribution of the dataset by synthesizing minority class samples, and generative adversarial networks (GANs) (Creswell et al., 2018), which generates virtual samples similar to the real data to further increase the number of minority class samples. Recent studies have introduced zero-shot learning to hyperspectral image classification. The SPECIAL framework leverages CLIP to generate pseudo-labels and incorporates noisy label learning to enhance model generalization (Pang et al., 2025). In addition, methods such as generating pseudo-samples (Wang et al., 2020) and data mixing (Zhou et al., 2022) can effectively expand the training set size, thereby enhancing model's generalization ability and stability. These sample expansion methods can effectively address the problem of data imbalance, improve the classifier's ability to recognize land cover types in minority classes, and ultimately increase classification accuracy (Zhang Q. et al., 2023). However, most sample expansion algorithms currently rely on pixel-level labels to generate new samples. If the original labels are affected by labeling errors, noise, or spectral variations, these disturbances will be carried over to the samples generated during the expansion process. It is important to note that noise in the labels can lead to mislabeling of the expanded samples, which hinders model learning and exacerbates classification errors.

To overcome the shortcomings of pixel-based data reduction and enhancement methods, replacing pixels with super-pixels has proven to be an effective solution (Liu et al., 2015). Superpixels, sometimes called objects when focusing on multispectral images (Blaschke, 2010), refer to irregular pixel regions with meaningful visual features composed of adjacent pixels with similar textures, colors, brightness, and other characteristics (Yan et al., 2022). By grouping similar pixels within a local spatial domain, the segmentation algorithm divides the 2D space into multiple subregions that are internally similar, thus effectively reducing computational complexity. Since the publication of the simple linear iterative clustering (SLIC) algorithm (Achanta et al., 2012), superpixel research has entered a period of rapid development due to its simplicity and speed. For example, Zhao (Zhao et al., 2018) proposed an improved SLIC search strategy called the fast linear iterative clustering (FLIC) method. This method achieves rapid convergence through active search based on prior information and improves edge alignment through quick traversal. Ban et al. (2018) proposed a Gaussian mixture model (GMM) based superpixel method that uses a Gaussian distribution model to describe superpixels and applies the expectation-maximization (EM) algorithm to estimate the parameters of the Gaussian distribution through maximum likelihood, finally assigning all pixels to a specific Gaussian model. In recent years, many studies have combined superpixel segmentation with other dimensionality reduction methods to improve classification accuracy (Zhang et al., 2025). In this way, multiscale superpixel principal component analysis (MSuperPCA) (Jiang et al., 2018) performs principal component analysis on superpixels at different scales using a voting mechanism, while the superpixel hybrid discriminant analysis (SHDA) method treats the spectral mean of superpixels as nodes and combines linear discriminant analysis (LDA) with the local linear embedding algorithm, improving classification accuracy (Zhang Q. et al., 2022). Similarly, the unsupervised LDA framework based on Gabor superpixels (Jia et al., 2021) and the decision fusion method based on local binary patterns (LBP) (Huang et al., 2020) also proved to enhance classification performance. Other discriminant analysis methods (Fang et al., 2015a) and kernel methods (Fang et al., 2015b) integrated with superpixels can also effectively extract spatial-spectral features from hyperspectral images. More recently, the band-by-band adaptive multi-scale superpixel feature extraction (BAMS-FE) method (Li et al., 2023) was shown to improve classification accuracy by exploring the

various spatial structural features inherent in hyperspectral images by combining spatial and spectral features. It also introduced the entropy rate superpixel segmentation algorithm (Liu et al., 2011) and explained variation (EV) evaluation metric.

It is worth noting that while superpixel segmentation methods are widely used in hyperspectral image preprocessing, research combining superpixels and data augmentation to address the sampling problem in hyperspectral image classification is still scarce. Furthermore, superpixel segmentation methods still face the challenge of selecting the optimal superpixel scale when processing images at different scales. The selection of superpixel segmentation scales is often based on experience or manual tuning, so continuous parameter adjustment is required in practice. Particularly in hyperspectral image classification, due to image complexity and high-dimensional features, automatically selecting the optimal superpixel scale represents a challenge for current research.

In summary, the method proposed in this work achieves a systematic integration of the following three innovations:

1. An automatic superpixel scale selection strategy based on EV trend analysis, which automatically determines the optimal initial superpixel scale, avoiding dependence on manual parameter settings and improving model adaptability.
2. A superpixel-level label generation mechanism with high spatial consistency, which uses a majority voting strategy to assign pixel-level labels to superpixels, thereby effectively suppressing label noise caused by annotation errors and spectral variability, thus improving stability in weak-label scenarios.
3. A feature-level data augmentation strategy based on superpixel labels, which integrates feature interval swapping, feature smoothing, and Gaussian perturbation to increase the diversity of the training data and improve the modeling of intra-class variation, thereby optimizing generalization under small sample sizes scenarios.

To the best of our knowledge, this is the first study focused on integrating automatic superpixel scale selection, superpixel label construction, and superpixel-based data augmentation into a unified framework for hyperspectral image classification. It systematically addresses three persistent challenges in small-sample hyperspectral classification: i) the lack of automated superpixel scale selection mechanisms; ii) the susceptibility of pixel-level labels to noise and other disturbances; iii) the difficulty of training robust models due to the limited number of annotated samples.

2 Fundamentals

2.1 Entropy rate superpixel segmentation

Superpixel segmentation algorithms treat spatially adjacent pixels as a subregion, thus achieving 2D image segmentation. They employ local adjacency as a strong constraint and measure the spectral similarity between pixels, typically implementing optimal segmentation of 2D space through optimization algorithms. In this work, we build on the entropy rate superpixel segmentation (ERS) algorithm. ERS constructs an entropy rate-based optimization process based on the graph segmentation methodology. It is mainly applied to 2D grayscale images from two fundamental theories: entropy and random walks. Compared with conventional methods such as SLIC or LSC, ERS achieves optimal segmentation by maximizing the graph entropy rate. Consequently, it has demonstrated superior performance in boundary preservation, compactness control, and spatiospectral consistency management in hyperspectral images. Furthermore, the superpixels generated by ERS exhibit improved structural consistency and strong compatibility with the EV-based segmentation quality-assessment criterion adopted in this study, providing a more stable and reliable basis for further work.

2.1.1 Entropy and entropy rate

The definition of entropy was first proposed by Shannon (Shannon, 1948), and it is thus also known as Shannon entropy. In the field of information theory, the entropy of a random variable is the average amount of information or uncertainty contained in that random variable. For a discrete variable X whose values belong to an Alphabet space and follow a 0–1 distribution (p: $X \rightarrow [0,1]$), the entropy of the variable is given by Equation 1.

$$H(X) = -\sum_{x \in X} p(x) log_b(p(x)), \tag{1}$$

Where p(x) represents the probability of the occurrence of each variable value. The choice of the log base in the logarithmic function has different meanings and applications. For example, when the base is 2, the resulting unit of entropy is bits, while when the base is e, the unit is nats. Finally, when the base is 10, the resulting units are called dits, bans, or hartleys. An equivalent definition of entropy is the expected value of the self-information of a variable.

The conditional entropy between two variables is defined in Equation 2.

$$H(X/Y) = -\sum_{x,y \in X \times Y} p_{X,Y}(x,y) log \frac{p_{X,Y}(x,y)}{p_Y(y)}, \tag{2}$$

Note that $p_{X,Y}(x,y) = P[X = x, Y = y]$ represents the probability of x occurring given that y has occurred, while $P_Y(y) = P[Y = y]$ represents the probability of y occurring. This formula characterizes the probability of the remaining random events occurring, given that a certain random variable has occurred.

2.1.2 Random walk on graphs

A random walk is a stochastic process that describes a path consisting of a series of random steps within a mathematical space. A one-dimensional random walk is also called a Markov chain. Let $X = \{X_t \backslash t \in T, X_t \in V\}$ denote a random walk process defined on a graph $G = (V, E)$. There is non-negative similarity (w) between the vertices of the graph. If a random model is used to represent the transition probability from one vertex to another, this probability is given by Equation 3.

$$p_{i,j} = Pr\left(X_{t+1} = v_j \middle| X_t = v_i\right) = w_{i,j} \big/ w_i \tag{3}$$

This formula represents the probability of reaching vi from vj at the next time step, where $w_i = \sum_{k:\, e_{i,k} \in E} w_{i,k}$ is the sum of the weights

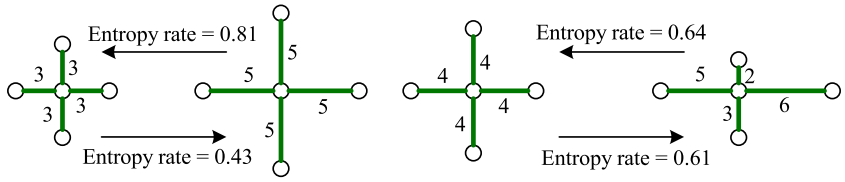

FIGURE 1
Entropy rate change diagram.

between all adjacent vertices of a vertex. The probability distribution across all vertices is given by Equation 4.

$$\mu = \left(\mu_1, \mu_2, ..., \mu_{|V|}\right)^T = \left(\frac{w_1}{w_T}, \frac{w_2}{w_T}, ..., \frac{w_{|V|}}{w_T}\right)^T \tag{4}$$

where $w_T = \sum_{i=1}^{|V|} w_i$ is a constant.

Under these premises, the entropy rate of the random walk on graph G is given by Equation 5.

$$\begin{aligned} H(X) = H(X_2|X_1) &= \sum_i \mu_i H(X_2|X_1 = v_i) = -\sum_i \mu_i \sum_j p_{i,j} log\, p_{i,j} \\ &= -\sum_i \frac{w_i}{w_T} \sum_j \frac{w_{i,j}}{w_i} log \frac{w_{i,j}}{w_i} \\ &= -\sum_i \sum_j \frac{w_{i,j}}{w_T} log \frac{w_{i,j}}{w_T} + \sum_i \frac{w_i}{w_T} log \frac{w_i}{w_T}, \end{aligned} \tag{5}$$

The ERS algorithm treats an image as a graph G = (V, E), with pixels representing the vertices in the graph and the edge weights between adjacent pixels determined by their similarity. The goal of ERS is to obtain a graph that contains k subgraphs. The algorithm uses the entropy rate of the random walk on the graph as the criterion to obtain compact and homogeneous clusters.

This construction explained above maintains the stationary distribution of the random walk unchanged (i.e., $\mu = (\mu_1, \mu_2, ..., \mu_{|V|})^T = (\frac{w_1}{w_T}, \frac{w_2}{w_T}, ..., \frac{w_{|V|}}{w_T})^T$). The transition probabilities within the vertex set are defined as Equation 6.

$$p_{i,j}(A) = \begin{cases} \frac{w_{i,j}}{w_i} & \text{if } i \neq j \text{ and } e_{i,j} \in A \\ 0 & \text{if } i \neq j \text{ and } e_{i,j} \notin A \\ 1 - \frac{\sum_{j:\, e_{i,j} \in A} w_{i,j}}{w_i} & \text{if } i = j, \end{cases} \tag{6}$$

Further, the entropy rate of the random walk on graph G = (V, A) can be redefined as Equation 7.

$$H(A) = -\sum_i \mu_i \sum_j p_{i,j}(A) log\left(p_{i,j}(A)\right) \tag{7}$$

Although adding any edge in set A will increase the entropy rate, the largest increase in entropy occurs when compact and homogeneous edges are selected. The reason for the monotonic increase in entropy with the addition of any edge is that each edge increases the uncertainty of the random walk. Figure 1 shows how graph structure is affected by the addition of edges, influencing the entropy rate and thus the effect of superpixel segmentation. It is important to highlight that the higher the entropy rate, the greater the uncertainty of the information in the graph, making the segmentation more likely to be more accurate.

2.2 Explained variation value

The categorization of evaluation metrics is based on the availability of ground-truth segmentation labels, resulting in two main types: supervised and unsupervised metrics. Supervised metrics include boundary recall (Rec), subsegmentation error (UE), compactness (CO), achievable segmentation accuracy (ASA), and mean boundary distance (MDE). All of these require reference segmentation maps for their computation. In contrast, unsupervised metrics leverage intrinsic properties of superpixels and pixel distributions in the original image, with explained variation (EV) and intra-cluster variation (ICV) being representative examples. ICV quantifies the spectral variation within a superpixel, while EV measures the adherence of the superpixel boundary to the original image. Note that when developing metrics to evaluate superpixel segmentation results, the similarity between neighboring superpixel blocks needs to be considered. In our approach, we choose the EV as the criterion to evaluate the effectiveness of multi-level superpixel optimization. The EV calculation formula is provided in Equation 8.

$$EV(S) = \frac{\sum_i (\mu_i - \mu)^2}{\sum_i (x_i - \mu)^2} \tag{8}$$

Where it is presented the sum over i pixels of an image I, being xi the actual pixel value, μ is the global pixel mean for the image I, and μi is the mean value of the pixels assigned to the superpixel that contains xi. As a result, EV quantifies the image variation explained by superpixels. Higher EV values imply greater explanation of the image's spectral variation.

2.3 Waveband-by-waveband adaptive multi-scale superpixel feature extraction algorithm

The hyperspectral image is represented as I (H × W × C), where its length, width, and number of bands are denoted by H, W, and C. After performing PCA on the hyperspectral image, the dataset is represented as Ipca (H × W × P), where the number of retained principal components is denoted as P. Each pixel of Ipca is represented as xi = 1, 2, ⋯ n, with n = H × W and xi (c) representing the grayscale value of the i pixel in the c principal

component. The initial superpixel segmentation scale is denoted as Ssize, while the relationship between the step and the number of iterations is given by segsizestep = Ssize × step. The number of superpixels to be segmented in each iteration is given by Equation 9.

$$m = \frac{H \times W}{segsize_{step} \times segsize_{step}} \tag{9}$$

The segmented superpixels are denoted as S_i (with i = 1, 2, ⋯m), while $|S_i|$ represents the average grayscale value of the superpixel i.

The adaptive capability of the superpixel segmentation algorithm to spatial and spectral features is crucial for joint spatiospectral feature extraction from hyperspectral images. Conventional algorithms usually apply superpixel segmentation to the first principal component of the PCA-transformed image and then compute the spectral mean of each superpixel, but this approach fails to fully exploit spatial information across bands. To address this, the proposed method applies ERS segmentation to each band of the hyperspectral image, either in the original feature space or in the PCA-transformed space, with input dimensions ranging from one to several bands. The core idea is to extract joint spatiospectral features through multi-scale superpixel segmentation, with the optimal segmentation scale determined by the EV metric to improve feature extraction.

In this process, the hyperspectral image is represented as a three-dimensional structure (length, width, and bands), then reduced by PCA to a few principal components. Superpixel segmentation is performed on the PCA-processed data, iteratively adjusting the segmentation scale to produce multiple segmentation results. Since a single scale can only capture features at one spatial resolution, combining multiple scales allows extraction of more comprehensive spatiospectral features. The EV metric serves as the criterion for selecting the optimal segmentation scale by measuring boundary adhesion between superpixels and the original image. In summary, the algorithm integrates ERS-based multi-scale superpixel segmentation with EV-based scale optimization to efficiently extract joint spatiospectral features. The integration of entropy rate superpixel segmentation and adaptive multi-scale feature extraction for each band forms the base module of this methodology, with implementation details intentionally omitted.

3 Description of the proposed methods

As is well known, obtaining accurate labeling data is difficult in hyperspectral classification, resulting in a shortage of labeled samples and affecting classification performance. To address this problem, this study proposes a comprehensive hyperspectral image classification method that combines adaptive EV-based superpixel scale selection, superpixel label expansion, and data expansion. This section is divided into three parts.

It should be noted that in the feature extraction stage, this study employs the existing BAMS-FE module without any modification, using it solely as a stable baseline to provide spatial–spectral consistent inputs for the subsequent modules of automatic superpixel scale selection, superpixel label construction, and superpixel-based augmentation. The novelty of this work does not lie in improving BAMS-FE, but rather in proposing three new modules within the AU-Super framework built upon it.

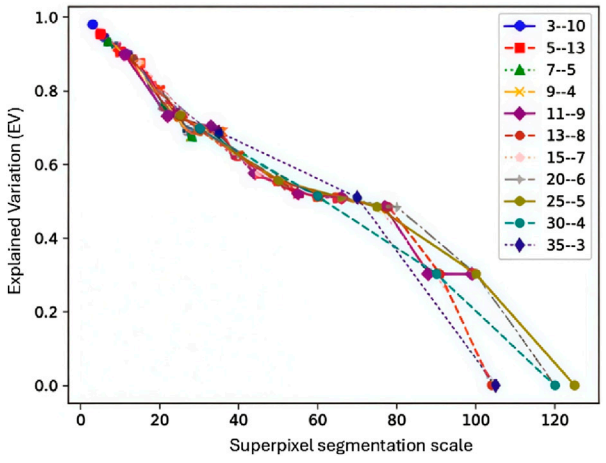

FIGURE 2
Automation diagram to find the optimal initial superpixel scale.

3.1 Automated search for the optimal initial superpixel scale

In hyperspectral image classification, the choice of superpixel scale directly affects the effectiveness of spatial–spectral feature representation. An excessively large scale may mix pixels from different classes into a single superpixel, while an overly small scale can lead to over-segmentation, increased computational cost, and the introduction of noise. Traditional approaches rely on manual parameter tuning, which is subjective and lacks consistency across datasets. To address this issue, we propose an adaptive scale selection method based on explained variation (EV), which enables more robust and generalizable superpixel segmentation.

The ERS algorithm was used to segment the PCA-processed images into superpixels of various scales, as it balances segmentation accuracy with computational efficiency. Note that the scale range of the superpixels can be adjusted to explore different levels of detail in the images. To determine the optimal superpixel scale, the EV value was calculated for each superpixel scale to assess the captured variance. EV serves as a robust metric for feature information retained at each scale; a higher EV indicates that the superpixel segmentation explains more variance. To avoid potential overfitting to specific datasets, the convergence trend of EV is used as the selection criterion rather than the absolute EV value, ensuring the selected scale reflects generalizable spatiospectral structures. In addition, the proposed method was validated across three heterogeneous datasets (Indian Pines, Salinas, and Pavia University), and the consistent improvements demonstrate that the automatic scale search does not overfit to a single dataset but provides stable segmentation quality in diverse scenarios. We then compared the EV values across successive iterations and selected the optimal scale at the point where the EV value converges, indicating that further increasing the number of superpixels yields diminishing

returns in explained variance, thus achieving a balance between detail preservation and computational efficiency. This stabilization point was identified by monitoring changes in EV as the scale increased and determining the scale at which further increases yielded the least additional information. Finally, we selected the scale that retained the most feature information as the initial optimal superpixel scale.

As an example, suppose the candidate superpixel scales are 3, 5, 7, 9, 11, 13, 15, 20, 25, 30, and 35. As shown in Figure 2, the first number in the notation (e.g., "3–10" or "5–13") denotes the initial scale, while the second indicates the number of iterations required for the EV value to converge. The convergence is defined as the point where the EV value approaches zero or stabilizes. For example, the blue curve in Figure 2 corresponds to an initial scale of 3, with its first point on the x-axis located at 3. This visualization allows for a straightforward comparison of EV convergence across different initial scales.

Continuing with the example depicted in Figure 2, we would select an initial superpixel scale of 5 (13 iterations) based on the following reasoning. The initial scale requiring more iterations means that, at this scale, the image contains more complex structural information and richer details during the superpixel segmentation process. That is, the fact that the algorithm requires more steps to reach stability (i.e., for the EV curve to converge) indicates that this scale can capture more spatial-spectral details of the image. It is worth noting that, although the initial scale of 3 produces a higher final EV value, it requires fewer iterations. This suggests that the information extracted at this scale is relatively simple or too focused on fine details, lacking hierarchical variation. In contrast, the scale of five can provide clearer feature layering and more complex structural information throughout the processing, making it more suitable as an initial scale for further analysis.

In summary, the proposed method for automatic superpixel scale selection, based on the EV criterion, efficiently determines the optimal initial scale for superpixel segmentation, thereby avoiding the subjectivity and instability brought about by manual parameter tuning. This module provides a stable and reliable structural foundation for subsequent superpixel labeling and feature enhancement strategies. The next section describes how to construct robust superpixel labels based on segmentation results to further mitigate the impact of label noise in small-sample classification tasks.

3.2 Superpixel enhanced training data

Pixel-level annotations often suffer from noise and inconsistency, especially in hyperspectral data where mixed pixels and boundary pixels significantly reduce classification reliability. Moreover, the limited number of pixel-level samples can make the training process unstable. To overcome these challenges, we elevate labels from the pixel level to the superpixel level, leveraging spatial consistency to suppress annotation noise and construct a more balanced and reliable training sample set.

In this section, a majority voting mechanism is used to map the original pixel-level labels to spatially consistent superpixel-level labels, significantly reducing labeling errors and noise interference. These structurally consistent superpixel labels provide a more stable supervisory signal for training with small samples. Based on these labels, we develop a set of diversity enhancement strategies to expand the distribution of limited training samples, as presented in the next section. In this sense, and after reducing the dimensionality of the hyperspectral image using PCA and preserving the first principal component, the ERS segmentation algorithm is applied to the previously determined optimal superpixel scale (see Section 3.1). Superpixel segmentation efficiently clusters spatially adjacent pixels with similar spectral properties into coherent regions. Based on the segmentation results, pixel-level ground-truth labels are mapped to superpixel-level labels. Each superpixel is assigned the label of the majority class within its region. This majority voting method ensures that the superpixel label represents the dominant class, thereby reducing the influence of noise and mislabeled pixels. Equation 10 defines this mapping scheme.

$$L_{superpixel}(i) = \arg\max_{c} \sum_{j \in S_i} \coprod \left(L_{pixel}(j) = c \right) \tag{10}$$

where $L_{superpixel}(i)$ is the label of superpixel i, S_i represents the set of pixels in superpixel i, $L_{pixel}(j)$ is the pixel-level label, c denotes the class, and $\coprod$ is the indicator function.

To address class imbalance and ensure diverse training data, a fixed number of superpixels are sampled for each class. This sampling process is controlled by parameter N, which specifies the number of superpixels to be selected per class. The balanced sampling ensures that each class is equally represented regardless of its prevalence in the original ground truth data. The sampled superpixel-level labels are then projected back to pixel-level labels, creating an enriched and denoised training dataset that preserves both spatial and spectral consistency.

Once the superpixel-level labels have been generated, the corresponding features are extracted from the hyperspectral image using the BAMS-FE algorithm. The extracted features include spatial and spectral information, providing a robust representation of each superpixel. These features are then scaled and used as input for classifier training.

3.3 Data augmentation

Under small-sample conditions, models are often prone to underfitting or over-reliance on a few specific samples. Traditional pixel-level augmentation may disrupt the consistency between spectral and spatial information. To better preserve spatiospectral integrity, we propose superpixel-level feature augmentation strategies, including feature interval swapping, feature smoothing, and Gaussian perturbation. These strategies not only increase intra-class diversity but also enhance the robustness of the model under limited training data. To further improve sample diversity and the classifier's generalization ability, multiple augmentation strategies are integrated while maintaining the inherent spatiospectral relationships of hyperspectral data.

3.3.1 Feature interval swapping

This technique increases intra-class variability by randomly exchanging a subset of features between two superpixels of the same class. First, two distinct samples (sample1 and sample2) are

selected randomly for each iteration. Second, half of the features are randomly chosen and swapped between the two samples, ensuring that the augmented data reflect plausible variations within the same class.

The swapped features for each sample are mathematically defined as given in Equation 11.

$$F_{augmentation} = \left\{ F_1^{orig}, F_2^{orig} \,\middle|\, swap(F_1, F_2) \right\} \tag{11}$$

where F1 and F2 represent feature subsets of the two samples, and F_1^{orig} and F_2^{orig} denote d-dimensional feature vectors of the selected samples, being d the number of extracted spectral-spatial features after preprocessing and dimensionality reduction.

3.3.2 Feature smoothing

To simulate local continuity and reduce noise in the training data, smoothing is applied to the selected features. For each feature, its value is replaced with the average of its neighboring features, if available. This operation is applied with a 50% probability for randomly selected features in the sample. This process ensures that the augmented samples maintain spatial coherence, which is crucial in hyperspectral image analysis.

3.3.3 Feature noise addition

Gaussian noise is added to the features that are not smoothed during the augmentation process. This introduces subtle variability, helping the model become more robust to noise in real-world data. The noise follows a Gaussian distribution as expressed in Equation 12.

$$N_{augmentation} = F_{orig} + n\,N\left(0, \sigma^2\right) \tag{12}$$

where σ is the standard deviation that controls the intensity of the Gaussian noise. In this study, σ is set as a fixed value (σ = 0.05) based on empirical evaluation to ensure a balance between feature stability and variability during augmentation.

By integrating feature interval swapping, smoothing operations, and Gaussian perturbation, a multi-strategy augmentation framework based on superpixel labels is constructed. This framework effectively expands the spectral-spatial distribution of training samples and improves the classifier's ability to model intra-class variability. At this point, the overall methodological framework of this study is fully established.

As shown in the ablation experiment (Table 10), the combination of these three types of enhancement methods has a better effect, further verifying their synergistic role in improving the generalization ability of the model.

3.3.4 Augmentation workflow

The augmentation process begins by pooling superpixel-level samples by class. For each class, the algorithm iteratively applies the above techniques to generate a specified number of augmented samples. The steps include: i) randomly selecting two samples from the class; ii) applying feature interval swapping, followed by feature smoothing or feature noise addition with a 50% probability; iii) repeating the process until the desired number of augmented samples is reached.

The augmented samples are then combined with the original training data, creating a richer and more diverse dataset. The processed superpixel-enhanced training data are fed into a random forest classifier. The classifier is trained on the augmented data, and predictions are validated against ground-truth labels. In summary, this study provides a robust framework for hyperspectral image classification, termed AU-Super, that effectively addresses the challenges of label sparsity and manual parameter tuning. The specific flowchart is illustrated in Figure 3.

4 Data sets and experimental design

All experiments were conducted using three public datasets: the Indian Pines dataset, the Pavia University dataset, and the Salinas dataset. These datasets have been widely used in the field of hyperspectral image classification in recent years (Xie et al., 2019). The spatial resolution of these datasets ranges from 5 cm to 20 m, and the number of spectral bands varies from 100 to 270. The land cover classes in these datasets include vegetation, buildings and other categories.

Indian Pines is one of the first publicly available hyperspectral datasets used for classification tasks. It was acquired in 1992 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over an agricultural area in Indiana, United States. The original image consists of 220 contiguous spectral bands covering the wavelength range 0.4–2.5 μm. Since bands 104–108, 150–163, and band 220 are severely affected by water absorption and cannot effectively reflect surface information, these bands were removed in this study. The remaining 200 bands were used for experiments. The spatial resolution of the image is 20 m. The dataset contains 16 land cover classes, with an unbalanced distribution of labeled samples among classes.

The Pavia University dataset was acquired in 2003 by the Reflective Optics Spectrographic Imaging System (ROSIS-03) at the University of Pavia, Italy. This spectrometer captures 115 contiguous spectral bands in the wavelength range 0.43–0.86 μm, with a spatial resolution of 1.3 m. In this study, 13 noise-affected bands were removed, resulting in a hyperspectral image composed of 103 spectral bands. The image contains a total of 42,776 labeled pixels, classified into nine land cover categories.

The Salinas dataset, similar to the Indian Pines dataset, was also acquired by the AVIRIS sensor over the Salinas Valley in California, United States. The spatial resolution of this dataset is 3.7 m. After removing bands 108–112, 154–167, and band 224 due to low signal-to-noise ratios, 204 bands were retained. The image contains a total of 54,129 labeled pixels and includes 16 different land cover classes.

The Indian Pines and Salinas datasets were provided by the California Institute of Technology, USA, while the Pavia University dataset was collected by the University of Pavia, Italy. The three datasets are publicly accessible at the following link:https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

Tables 1–3 summarize the land cover types within the three aforementioned datasets.

The experiments were carried out with a limited number of samples. For each class, 1 to 10 samples were selected as the training set. Each configuration was repeated ten times, recording the average and standard deviation of the performance metrics used to carry out the accuracy evaluation. In each run, training samples were randomly selected to ensure the robustness and reliability of the

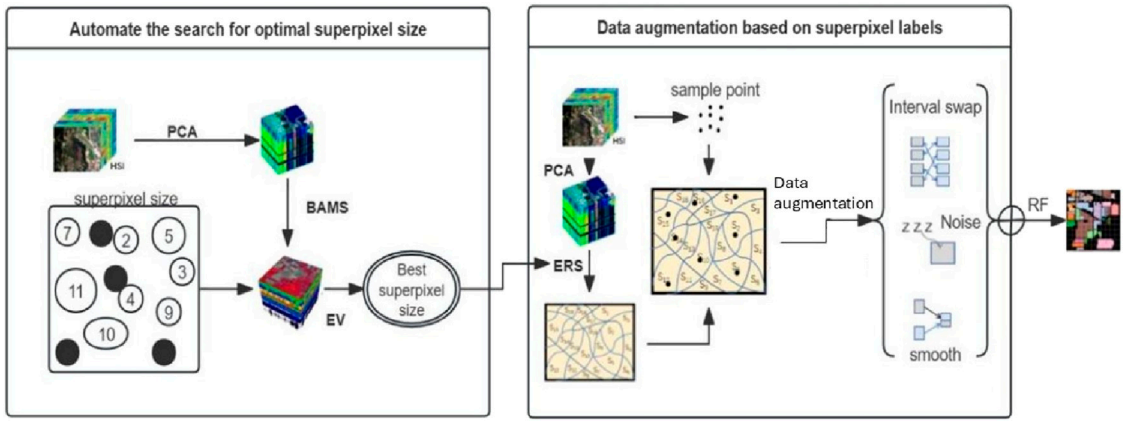

FIGURE 3
Flowchart corresponding to the proposed method.

TABLE 1 Class labels information for the Indian Pines dataset.

Id	Class names	Number of pixels	Label colors
1	Alfalfa	46	
2	Corn-no till	1,428	
3	Corn-min till	830	
4	Corn	237	
5	Grass-pasture	483	
6	Grass-trees	730	
7	Grass-pasture-mowed	28	
8	Hay-windrowed	478	
9	Oat	20	
10	Soybean-no till	972	
11	Soybean-min till	2,455	
12	Soybean-clean	683	
13	Wheat	205	
14	Woods	1,265	
15	Buildings-Grass-Trees-Drives	386	
16	Stone-Steel-Towers	93	

experiment, extracting N (1:10) training samples without replacement (i.e., the same sample is not selected more than once in a single run). Per-class classification accuracy, overall precision (OA), average precision (AA), and kappa coefficient were used as performance metrics for the experiments. Per-class classification accuracy, also known as recall, is calculated as the number of correctly classified pixels in each class divided by the total number of pixels in that class. OA is defined as the ratio of accurately classified pixels to the entire dataset, while the kappa coefficient is used to assess the consistency between the ground truth and the classification result. AA is the average of the recall values for each class.

Several state-of-the-art (SOTA) methods were used as a benchmark to evaluate the performance of the proposed method. These included SuperPCA and MSuperPCA (Jiang et al., 2018), superpixel-adaptive singular spectral analysis (SpaSSA) (Sun et al., 2022), superpixel-based Brownian descriptor (SBD) (Zhang et al., 2022c), superpixel-level hybrid discriminant analysis (SHDA) (Zhang et al., 2022b), and band-by-band adaptive multi-scale superpixel feature extraction system (BAMS-FE) (Li et al., 2023). These methods cover key directions in current hyperspectral image classification, such as superpixel segmentation, multi-scale processing, small sample learning, and feature fusion. These reference algorithms represent some of the most representative

TABLE 2 Class labels information for the Pavia University dataset.

Id	Class names	Number of pixels	Label colors
1	Asphalt	6,631	
2	Meadows	3,682	
3	Gravel	1,345	
4	Trees	947	
5	Painted metal sheets	1,330	
6	Bare soil	18,649	
7	Bitumen	2099	
8	Self-blocking bricks	3,064	
9	Shadows	3,682	

TABLE 3 Class labels information for the Salinas dataset.

Id	Class names	Number of pixels	Label colors
1	Broccoli green weeds 1	2,099	
2	Broccoli green weeds 2	3,726	
3	Fallow	1976	
4	Fallow rough plow	1,394	
5	Fallow smooth	2,678	
6	Stubble	3,959	
7	Celery	3,579	
8	Grapes untrained	11,271	
9	Soil vineyard develop	6,203	
10	Corn senesced green weeds	3,278	
11	Lettuce romaine 4weeks	1,068	
12	Lettuce romaine 5weeks	1,927	
13	Lettuce romaine 6weeks	916	
14	Lettuce romaine 7weeks	1,070	
15	Vinyard untrained	7,268	
16	Vinyard vertical trellis	1,807	

techniques developed in recent years in the field of hyperspectral image preprocessing for classification.

Two more non-superpixel-based methods were included to be used as benchmarks. One of them is a novel multi-scale 2-D singular spectrum analysis (2-D-SSA) fusion method based on PCA and segmented PCA (SPCA). It is based on obtaining multi-scale spatial features by applying multi-scale 2-D-SSA on the SPCA-reduced dimensional images to be subsequently fused with the PCA-derived global spectral features to form multi-scale spectral-spatial feature principal components (MSF-PCs) (Fu et al., 2022). The other was multi-hop attention graph and multi-scale convolutional fusion network (AMGCFN) (Zhou et al., 2023), a recent neural network-based method that applies an end-to-end deep learning model that usually requires large amounts of labeled samples. It is important to note that we have intentionally avoided including more end-to-end deep learning methods (e.g., those based on GNNs or Transformers) for two reasons. First, they require a large number of labeled samples, which is inconsistent with the focus of this study, namely, small-sample and superpixel-based enhancement. In this sense, including these models may lead to unfair comparisons. Second, this work focuses on traditional classifiers (e.g., Random Forest) combined with preprocessing and data augmentation strategies, aiming to improve learning performance on small samples rather than designing new classifiers.

All experiments in this study were conducted on a computer system equipped with an Intel i7-13700K processor. Feature extraction procedures for algorithms such as SuperPCA, MSuperPCA, SpaSSA, SBD, SHDA, and MSF-PCs were performed using MATLAB 2022b platform, while BAMS-FE algorithm was implemented by integrating Python 3.10 and MATLAB 2022b. Finally, AMGCFN was executed in Python 3.10 environment with the help of PyTorch 1.11 open-source library. The classification tasks were performed in Python 3.10 using the scikit-learn library 1.3.0, employing Random Forest (RF) as the classifier for training and prediction. The hyperparameters of all comparative algorithms were strictly configured according to the predefined configurations in their original publications or published source codes, ensuring the fairness and consistency of the experimental comparisons.

5 Results

5.1 AU-super performance. General quantitative analysis

Experiments were conducted under multisampling configurations with 1, 3, 5, and 10 randomly drawn samples per class for training. For reasons of space and clarity, the results are partially presented in Tables 4–6. Indeed, while Tables 4 and 6 show the results for five training samples per class, Table 5 shows the results for a single training sample per class. There are two main reasons for doing this:

i. This work seeks to compare different preprocessing methods for classifying HIS data with different numbers of training samples. Therefore, the three tables cover two typical scenarios (with one and five samples per class) to demonstrate the stability and adaptability of the compared methods in their performance under different conditions.
ii. The Salinas dataset was selected to represent the results of the one-sample-per-class comparison due to its high spectral separability, which generally produces exceptional classification results. In this sense, with the standard five-sample configuration, classification accuracy often approaches saturation, which can obscure performance differences between methods and make fair comparisons difficult. To improve sensitivity and more accurately assess the performance of each method under limited sample conditions, we specifically selected the one-sample-per-class

TABLE 4 Per-class classification accuracy for different feature extraction algorithms using random forest as a classifier on the Indian Pines dataset with five training samples per class. The values following the symbol ± represent the standard deviation computed over ten independent runs. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Class no.	AU-super	MSuperPCA	SpaSSA	SHDA	SBD	BAMS-FE	AMGCFN	SuperPCA
1	98.22 ± 0.10	**100.00 ± 0.00**	77.80 ± 0.13	97.80 ± 0.01	97.80 ± 0.01	99.02 ± 0.01	98.29 ± 0.03	**100.00 ± 0.00**
2	56.21 ± 0.10	64.79 ± 0.10	36.44 ± 0.12	**73.72 ± 0.10**	46.05 ± 0.10	62.46 ± 0.12	58.48 ± 0.09	50.71 ± 0.10
3	**93.06 ± 0.10**	65.84 ± 0.15	41.74 ± 0.11	72.78 ± 0.17	68.39 ± 0.15	70.18 ± 0.18	66.48 ± 0.12	60.77 ± 0.16
4	81.13 ± 0.11	87.56 ± 0.11	71.14 ± 0.15	84.07 ± 0.13	80.82 ± 0.22	**90.95 ± 0.09**	93.81 ± 0.08	68.32 ± 0.16
5	**100.00 ± 0.00**	85.60 ± 0.08	56.74 ± 0.14	80.08 ± 0.07	80.56 ± 0.16	78.13 ± 0.07	78.52 ± 0.09	**88.77 ± 0.09**
6	**99.06 ± 0.13**	92.64 ± 0.15	50.88 ± 0.08	93.66 ± 0.06	90.73 ± 0.08	94.52 ± 0.09	88.37 ± 0.09	79.18 ± 0.15
7	96.63 ± 0.07	96.30 ± 0.02	94.78 ± 0.10	96.52 ± 0.02	92.17 ± 0.21	**99.78 ± 0.01**	100.00 ± 0.00	96.30 ± 0.02
8	**100.00 ± 0.00**	**100.00 ± 0.00**	85.17 ± 0.07	94.13 ± 0.11	85.00 ± 0.17	**100.00 ± 0.0**	98.11 ± 0.03	93.15 ± 0.12
9	**100.00 ± 0.00**	**100.00 ± 0.00**	86.67 ± 0.12	**100.00 ± 0.0**	95.00 ± 0.22	**100.00 ± 0.0**	97.67 ± 0.08	**100.00 ± 0.00**
10	**88.37 ± 0.13**	83.82 ± 0.05	38.46 ± 0.12	72.43 ± 0.15	65.45 ± 0.15	81.61 ± 0.06	76.57 ± 0.08	70.32 ± 0.14
11	**88.02 ± 0.10**	77.49 ± 0.12	47.56 ± 0.15	66.90 ± 0.10	59.00 ± 0.13	79.49 ± 0.10	70.24 ± 0.10	45.21 ± 0.10
12	**83.11 ± 0.08**	60.98 ± 0.12	36.52 ± 0.09	80.53 ± 0.18	61.85 ± 0.17	72.24 ± 0.11	60.01 ± 0.13	46.72 ± 0.11
13	**100.00 ± 0.00**	99.52 ± 0.00	85.55 ± 0.09	99.52 ± 0.00	99.52 ± 0.00	99.88 ± 0.00	97.47 ± 0.03	99.55 ± 0.00
14	89.81 ± 0.14	89.38 ± 0.10	74.22 ± 0.12	**94.21 ± 0.06**	86.42 ± 0.10	93.94 ± 0.07	85.05 ± 0.10	62.16 ± 0.09
15	**99.55 ± 0.00**	93.40 ± 0.10	69.57 ± 0.10	86.75 ± 0.14	61.15 ± 0.12	91.81 ± 0.09	95.91 ± 0.06	89.10 ± 0.14
16	**100.00 ± 0.00**	96.42 ± 0.05	94.09 ± 0.06	**100.00 ± 0.0**	98.01 ± 0.02	**100.00 ± 0.0**	97.95 ± 0.03	89.32 ± 0.12
OA	**86.50 ± 3.28**	79.99 ± 2.92	52.55 ± 4.08	79.13 ± 2.91	68.61 ± 4.24	81.34 ± 3.39	75.61 ± 3.40	62.55 ± 2.60
AA	**92.07 ± 4.93**	87.11 ± 7.14	65.46 ± 10.83	87.07 ± 8.06	79.25 ± 12.51	88.38 ± 6.34	85.18 ± 7.10	77.47 ± 9.41

configuration for the accuracy evaluation performed on the Salinas dataset (Table 5).

Overall, and when integrated with the RF classifier, the proposed AU-Super method showed solid performance, consistently outperforming existing techniques to achieve remarkable OA and AA scores across all datasets.

Regarding the Indian Pine dataset (Table 4), AU-Super ranked first in terms of per-class classification accuracy in up to 11 classes out of 16, followed by BAMS-FE and MSuperPCA, with a notable difference of more than five percentage points. AU-Super's performance was particularly excellent (100% recall) in detecting agricultural crops such as grass-pasture, hay-windrowed, oat and wheat, indicating its promising application in agriculture. The only class that showed a very low recall rate (below 60%) was no-till corn, although the other benchmark methods also performed poorly for this class.

In the case of the Salinas dataset (Table 5), it is worth noting that the algorithm proposed in this work achieved even better performance (OA = 99.65%) only working with one training sample per class, reaching first place in up to 14 of the 16 classes, albeit tied seven times with BAMS-FE. However, its performance only surpassed the BAMS-FE method by a difference of 1.52 percentage points in terms of overall accuracy for the 16 classes involved. This dataset focuses more on field crops such as broccoli, celery, lettuce, and vineyards, also including agricultural practices such as fallows and stubble, all of them showing a great spectral separability. Therefore, considering the excellent performance of AU-Super on this dataset, its application could be recommended for agricultural monitoring using HIS data.

The Pavia University dataset contains land covers quite different to those contained in the Indian Pine or Salinas datasets. Actually, this dataset is more focused on addressing urban land covers such as asphalt, gravel, shadows, or painted metal sheets, which can be very useful for segmenting roads and buildings within the framework of a road infrastructure safety assessment (Brkić et al., 2023). AU-Super performed best in five of the nine classes, with recall rates above 90% with just five training samples across up to six classes, including meadows, painted metal sheets, bare ground, bitumen, self-blocking bricks, and shadows, and only partially failed to correctly detect trees (recall = 79.33%). BAMS-FE came in second, but very close to AU-Super, with a difference of just 0.15 percentage points in overall accuracy. The other benchmark methods performed significantly worse, failing to even reach 80% overall accuracy. It is important to highlight the ability of BAMS-FE to discriminate between asphalt and bare soil (recall rates of 92.24% and 99.40%, respectively), which suggests that this method can be promisingly used in tasks related to road segmentation for pavement distress monitoring (Chen et al., 2024).

Regarding the stability of the proposed method's performance when modifying the randomly drawn training sample set to configure ten repetitions, the results in Tables 4–6 demonstrate that the standard deviations of per-class classification accuracy values are quite low. This finding can be applied to the rest of

TABLE 5 Per-class classification accuracy for different feature extraction algorithms using random forest as a classifier on the Salinas dataset with only one training sample per class. The values following the symbol ± represent the standard deviation computed over ten independent runs. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Class no.	AU-super	MSuperPCA	SpaSSA	SHDA	SBD	BAMS-FE	AMGCFN	SuperPCA
1	**100.00 ± 0.00**	**100.00 ± 0.0**	51.87 ± 0.19	99.40 ± 0.02	**100.00 ± 0.0**	**100.00 ± 0.00**	97.81 ± 0.03	55.29 ± 0.13
2	**100.00 ± 0.00**	99.94 ± 0.00	77.67 ± 0.18	80.60 ± 0.16	99.04 ± 0.04	**100.00 ± 0.00**	99.54 ± 0.01	31.15 ± 0.08
3	**100.00 ± 0.00**	99.99 ± 0.00	48.72 ± 0.18	78.07 ± 0.18	46.69 ± 0.11	**100.00 ± 0.00**	95.93 ± 0.08	69.03 ± 0.19
4	**100.00 ± 0.00**	90.81 ± 0.09	80.49 ± 0.24	97.33 ± 0.06	99.93 ± 0.00	99.83 ± 0.00	92.97 ± 0.08	55.33 ± 0.22
5	**99.83 ± 0.10**	65.37 ± 0.20	45.20 ± 0.19	85.63 ± 0.19	97.52 ± 0.09	99.06 ± 0.00	82.51 ± 0.19	46.99 ± 0.19
6	**100.00 ± 0.00**	93.75 ± 0.10	90.86 ± 0.03	88.04 ± 0.14	99.92 ± 0.00	99.89 ± 0.00	95.21 ± 0.04	40.94 ± 0.17
7	99.91 ± 0.13	53.00 ± 0.08	68.17 ± 0.22	88.32 ± 0.12	99.01 ± 0.00	**100.00 ± 0.00**	97.45 ± 0.04	29.21 ± 0.07
8	**99.40 ± 0.23**	42.89 ± 0.15	40.90 ± 0.22	59.85 ± 0.22	55.17 ± 0.33	93.14 ± 0.13	57.20 ± 0.28	16.44 ± 0.07
9	**100.00 ± 0.00**	82.68 ± 0.16	79.68 ± 0.25	94.59 ± 0.05	99.49 ± 0.01	99.98 ± 0.00	**100.00 ± 0.00**	37.50 ± 0.12
10	**100.00 ± 0.00**	53.62 ± 0.21	31.47 ± 0.15	81.18 ± 0.16	57.95 ± 0.24	96.61 ± 0.02	84.51 ± 0.14	38.59 ± 0.16
11	**100.00 ± 0.00**	58.78 ± 0.22	44.43 ± 0.24	96.41 ± 0.09	97.15 ± 0.01	97.99 ± 0.02	95.76 ± 0.06	49.33 ± 0.20
12	**100.00 ± 0.00**	72.61 ± 0.14	75.03 ± 0.17	89.77 ± 0.13	98.68 ± 0.06	**100.00 ± 0.00**	95.38 ± 0.10	40.75 ± 0.12
13	**98.36 ± 0.01**	98.24 ± 0.00	86.06 ± 0.20	92.24 ± 0.14	97.82 ± 0.00	98.03 ± 0.00	96.92 ± 0.04	98.25 ± 0.00
14	97.43 ± 0.05	96.31 ± 0.03	85.88 ± 0.18	80.80 ± 0.26	94.20 ± 0.08	**97.95 ± 0.01**	94.23 ± 0.07	54.59 ± 0.13
15	**99.53 ± 0.23**	89.86 ± 0.14	67.05 ± 0.28	72.40 ± 0.19	72.82 ± 0.30	99.52 ± 0.00	89.87 ± 0.15	31.90 ± 0.12
16	**100.00 ± 0.00**	99.07 ± 0.00	52.64 ± 0.24	68.65 ± 0.21	80.45 ± 0.29	**100.00 ± 0.00**	94.64 ± 0.06	39.36 ± 0.12
OA	**99.65 ± 3.34**	74.51 ± 5.55	61.53 ± 4.18	79.59 ± 4.30	81.29 ± 6.64	98.13 ± 2.66	86.36 ± 4.81	36.21 ± 3.32
AA	**99.50 ± 4.09**	81.06 ± 9.40	64.13 ± 19.8	84.58 ± 14.5	87.24 ± 9.69	98.88 ± 1.14	91.87 ± 8.58	45.92 ± 13.0

TABLE 6 Per-class classification accuracy for different feature extraction algorithms using random forest as a classifier on the Pavia University dataset with five training samples per class. The values following the symbol ± represent the standard deviation computed over ten independent runs. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Class no.	AU-super	MSuperPCA	SpaSSA	SHDA	SBD	BAMS-FE	AMGCFN	SuperPCA
1	89.14 ± 0.03	44.39 ± 0.08	43.86 ± 0.14	74.10 ± 0.01	70.28 ± 0.10	**92.24 ± 0.06**	76.75 ± 0.12	39.50 ± 0.08
2	**94.53 ± 0.06**	80.24 ± 0.08	48.53 ± 0.11	73.41 ± 0.10	76.65 ± 0.13	86.15 ± 0.10	80.55 ± 0.11	81.46 ± 0.08
3	**89.13 ± 0.04**	78.85 ± 0.10	39.37 ± 0.10	84.60 ± 0.17	84.18 ± 0.11	88.28 ± 0.10	86.03 ± 0.10	78.55 ± 0.08
4	79.33 ± 0.02	60.57 ± 0.07	65.71 ± 0.14	68.34 ± 0.13	59.46 ± 0.11	**85.50 ± 0.09**	76.39 ± 0.12	32.29 ± 0.07
5	**100.00 ± 0.00**	98.25 ± 0.02	96.21 ± 0.04	97.12 ± 0.07	94.31 ± 0.04	99.69 ± 0.00	99.75 ± 0.00	98.65 ± 0.01
6	90.15 ± 0.15	82.40 ± 0.10	64.35 ± 0.13	84.26 ± 0.06	72.79 ± 0.22	**99.40 ± 0.02**	94.16 ± 0.08	84.26 ± 0.07
7	**100.00 ± 0.00**	98.40 ± 0.06	52.63 ± 0.10	96.97 ± 0.02	86.82 ± 0.03	**100.00 ± 0.00**	94.68 ± 0.10	89.80 ± 0.06
8	96.27 ± 0.10	93.39 ± 0.06	57.91 ± 0.12	74.03 ± 0.11	82.00 ± 0.15	**97.21 ± 0.04**	82.56 ± 0.13	63.55 ± 0.09
9	**100.00 ± 0.00**	99.29 ± 0.01	70.67 ± 0.08	90.61 ± 0.06	76.51 ± 0.05	92.23 ± 0.08	86.60 ± 0.09	80.57 ± 0.08
OA	**90.80 ± 2.28**	76.11 ± 3.78	53.37 ± 5.75	75.76 ± 5.07	75.62 ± 4.15	90.65 ± 4.25	82.88 ± 4.87	70.86 ± 2.99
AA	**93.61 ± 4.53**	81.31 ± 6.26	59.47 ± 11.01	82.49 ± 8.82	78.89 ± 10.61	93.41 ± 5.39	86.39 ± 9.37	72.07 ± 6.77

the benchmark methods. Thus, all the tested methods can extract valuable information for the RF classifier from a small number of training samples, regardless of the composition of this sample set. In this way, even the end-to-end deep learning model AMGCFN provided competitive results despite the small number of labeled samples available to train the network, achieving OA values of 75.61%, 86.36% and 82.88% for the Indian Pines, Salinas, and Pavia University datasets, respectively.

Overall, when integrated with the RF classifier, the proposed AU-Super method demonstrates stable performance and consistently outperforms existing techniques. To provide a clearer comparison, Figure 4 presents line plots of OA and AA under

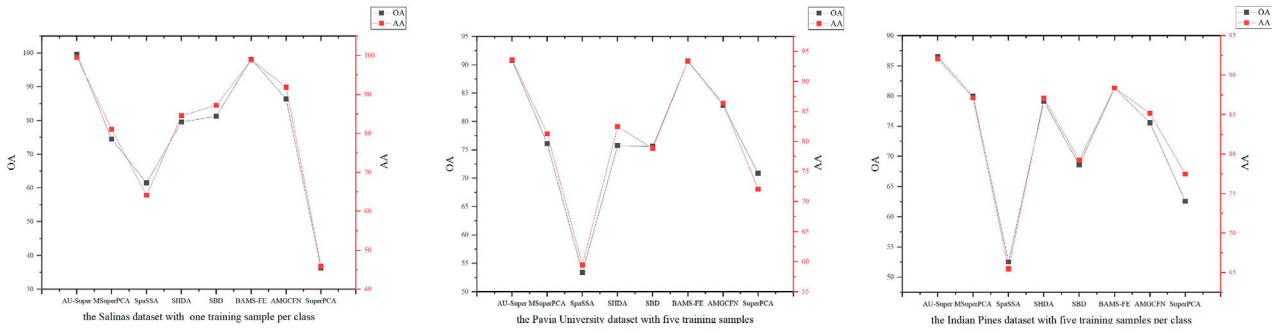

FIGURE 4
Overall accuracy (OA) and average accuracy (AA) of different methods under 10 training samples per class.

different numbers of training samples, showing that classification accuracy exhibits a steady growth trend as the number of samples increases. In addition, the cross-dataset performance differences reveal meaningful insights. On the Indian Pines dataset, AU-Super achieves significantly higher overall accuracy than other methods, with an OA improvement of more than 10% compared to traditional approaches such as SSGSA and SSEDA. This indicates that the method can better preserve spatial–spectral consistency in agricultural scenes with severe spectral mixing and subtle inter-class differences, thereby producing more reliable classification results. On the Salinas dataset, AU-Super outperforms the best baseline method, BAMS-FE, by approximately 1.5% in OA, suggesting that the method is particularly effective when class spectral separability is high. This result also demonstrates that superpixel-based label expansion and feature augmentation strategies can further amplify inter-class differences, allowing the method to maintain high accuracy even under extreme small-sample conditions (e.g., only one training sample per class). On the Pavia University dataset, the OA difference between AU-Super and BAMS-FE is less than 0.2%, indicating that the two methods achieve very similar performance. This observation suggests that in urban scenes with highly heterogeneous class boundaries, the complex spatial structures may offset part of the advantages of the enhancement strategies, making the classification capabilities of different methods comparable.

5.2 AU-super performance vs. number of training samples per class

To validate the effectiveness of the proposed method under few-shot training scenarios, this section compares AU-Super with three recently proposed SOTA algorithms for HIS classification based on a small number of training samples: MSuperPCA [30], BAMS-FE [36], and MSF-PCs [42]. Note that, as discussed above, the MSF-PC method is a non-superpixel-based algorithm, although it was specifically designed to work with a very small number of training samples. To perform a comprehensive evaluation across different training sample sizes, experiments were conducted using a range of training samples per class between 1 and 10. Each configuration was repeated ten times, and the average classification accuracy metrics (OA, AA and kappa) were recorded as the final results presented in Tables 7–9.

These results indicate that AU-Super consistently outperforms other benchmark methods, especially on the University of Pavia dataset, demonstrating its superiority when working with small datasets. Furthermore, they reveal a clear pattern indicating that the model's classification performance improves significantly as the number of training samples increases, thereby achieving higher accuracy.

There is a consensus that the effectiveness of a model is positively correlated with the number of labeled samples when applying machine learning approaches. However, manual labeling of HSI data is time-consuming and expensive due to the limited spatial resolution, which makes it difficult to have a large amount of labeled data to train the classifier. In this context, current research is clearly focused on the development of preprocessing and classification methods capable of producing good results from a small number of labeled training samples (Wang et al., 2023). In this sense, AU-Super has proven to be very efficient in obtaining excellent classification results using a relatively small training sample size, reaching OA and AA values above 80% with only three or four samples per class.

5.3 Ablation experiment

Since the developed algorithm consists of two main components, an ablation experiment was performed to evaluate their complementarity with the full method. Indeed, the first part of the algorithm performs automatic initialization for optimal superpixel scaling, while the second part is dedicated to data augmentation. The results of the ablation experiment were obtained using the Indian Pines dataset. In Table 10, "Super" refers to the case where the optimal superpixel size was set manually, i.e., without implementing the automation module of the full algorithm. Similarly, in Table 10, "AU-Super-0" refers to the method without applying the data augmentation module. It should be noted that the training patterns were the same as those used in Section 5.4.

Furthermore, the results in Table 10 show that AU-Super exhibits stable and superior classification performance even with extremely limited training samples. When only one sample per class is used, AU-Super achieves an overall accuracy (OA) of 59.91%, which is significantly higher than the 40.20% obtained when the

TABLE 7 Accuracy assessment of the four methods tested as a function of the number of training samples per class for the Random Forest classifier on the Indian Pines dataset. Ns means the number of samples per class used for training. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Ns	AU-super			BAMS-FE			MSF-PCs			MsuperPCA		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
1	**59.91**	**74.41**	**55.77**	55.23	71.93	50.12	47.03	70.61	42.47	50.95	67.94	46.15
2	**76.59**	**84.95**	**73.59**	73.76	83.41	70.14	58.45	73.76	53.94	64.47	78.17	60.70
3	**79.57**	84.83	**76.76**	73.84	**85.02**	70.73	60.51	79.31	56.00	77.44	82.83	74.42
4	**83.27**	**90.76**	**81.02**	79.32	86.40	77.46	72.85	84.17	69.39	79.99	86.59	77.46
5	**86.50**	**91.95**	**84.70**	80.69	90.11	78.60	75.75	83.59	72.47	85.16	89.84	82.14
6	**88.65**	**94.03**	**87.10**	83.48	89.46	81.23	77.64	87.79	74.56	84.62	90.85	81.66
7	**90.43**	**94.39**	**89.14**	85.42	92.23	83.56	75.33	86.37	71.92	88.45	89.71	86.98
8	**90.39**	**95.40**	**89.13**	84.91	91.50	82.93	83.05	91.21	80.83	84.62	90.85	82.66
9	**92.16**	**95.85**	**91.09**	89.12	93.66	87.61	81.99	88.83	79.49	89.67	92.82	88.19
10	**93.60**	**96.62**	**92.75**	91.55	94.78	90.35	87.06	92.78	92.87	86.32	89.28	84.44

TABLE 8 Accuracy assessment of the four methods tested as a function of the number of training samples per class for the Random Forest classifier on the Salinas dataset. Ns means the number of samples per class used for training. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Ns	AU-super			BAMS-FE			MSF-PCs			MsuperPCA		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
1	95.14	**98.15**	**94.60**	87.65	90.03	85.35	**95.18**	92.87	94.58	61.23	74.33	58.31
2	95.93	**98.11**	**95.47**	91.51	94.52	88.91	**96.08**	94.15	95.58	83.44	86.87	81.81
3	**99.02**	**98.89**	**98.91**	93.72	95.57	90.85	97.07	95.88	9,671	90.87	91.10	89.89
4	**99.63**	**9,937**	**99.59**	95.32	96.67	91.84	97.97	97.64	97.72	95.66	96.01	95.09
5	**99.55**	**99.46**	**99.50**	98.17	98.65	92.11	96.98	96.32	96.60	95.64	96.54	95.17
6	98.59	**99.5**	**98.43**	**98.67**	98.84	94.64	98.76	98.41	98.60	94.96	96.49	94.4
7	**99.37**	**99.15**	**99.30**	98.71	99.85	96.13	98.73	97.95	98.57	97.31	97.28	97.01
8	**99.44**	**98.82**	**99.38**	98.88	97.40	93.22	98.47	98.43	98.28	97.87	97.88	97.63
9	**99.55**	**99.51**	**99.49**	99.32	98.14	99.45	98.02	97.06	97.77	0.98	98.49	97.77
10	**99.60**	**99.44**	**99.56**	99.08	98.42	99.53	98.46	98.66	98.27	99.2	98.61	99.11

automatic scale selection module (Super) is removed. This indicates that the automated superpixel scale selection plays a crucial role in improving the model performance. Similarly, removing the data augmentation module (AU-Super-0) also leads to a decrease in classification accuracy, further confirming that the superpixel-based augmentation strategy effectively improves model training. As the number of training samples gradually increases to 10, the overall accuracy of AU-Super steadily increases to 93.60%, notably outperforming the comparative methods. These findings demonstrate that the two core modules of AU-Super offer significant complementary advantages in small sample sizes, effectively mitigating performance limitations caused by insufficient samples and improving the accuracy and robustness of hyperspectral image classification.

5.4 Summary of the experimental results

This work develops an automated method to select the optimal HSI superpixel scale based on the BAMS feature extraction framework (Li et al., 2023), using the EV metric (Moore et al., 2008) to measure the color homogeneity of a superpixel. This method has been devised to automatically determine the best superpixel segmentation scale based on the spectral features and spatial structure of the HSI image, avoiding tedious manual parameter tuning. Furthermore, considering that pixel-level labeling is much more difficult than image-level labeling, an unsupervised method is proposed to expand superpixel labels. In this sense, when segmenting the image with superpixels, the pixel-level labels become superpixel-level labels, effectively reducing the

TABLE 9 Accuracy assessment of the four methods tested as a function of the number of training samples per class for the Random Forest classifier on the Pavia University dataset. Ns means the number of samples per class used for training. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Ns	AU-super			BAMS-FE			MSF-PCs			MsuperPCA		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
1	**61.06**	64.06	**51.25**	56.80	**69.09**	48.36	38.80	56.32	28.89	30.11	33.37	20.19
2	71.24	**83.18**	**65.17**	69.68	75.82	62.30	**72.21**	71.47	63.36	50.76	60.11	41.32
3	**82.77**	**87.51**	**78.09**	71.30	80.24	68.35	65.23	75.07	57.47	58.85	68.16	49.73
4	**82.75**	**91.10**	**78.41**	75.59	85.01	76.81	79.62	86.32	74.32	71.02	78.42	63.74
5	**88.05**	**92.35**	**83.52**	82.85	87.55	74.99	82.9	87.91	78.48	66.32	75.66	58.31
6	**88.19**	**93.46**	**84.92**	84.92	88.50	76.01	85.19	90.55	80.87	76.14	81.91	69.91
7	**90.73**	**95.33**	**88.07**	86.87	90.57	84.19	85.17	88.64	80.90	75.59	82.43	69.22
8	**92.57**	**94.63**	**90.29**	88.05	**91.57**	87.61	86.47	89.08	82.43	75.31	80.80	68.55
9	**93.39**	**93.09**	**90.08**	92.73	91.72	89.66	87.97	91.95	84.59	81.94	86.89	76.88
10	**95.49**	**95.33**	**94.06**	94.23	92.37	91.43	88.52	91.56	85.08	83.4	87.38	78.69

TABLE 10 Accuracy assessment metrics corresponding to the complete (AU-Super) and the ablated two components (Super, without automatically calculated superpixel scaling, and AU-Super-0, without augmentation) of the proposed algorithm as a function of the number of training samples per class (Ns) for the Random Forest classifier on the Indian Pines dataset. Metrics are presented as a percentage (%). The best results are highlighted in bold.

Ns	AU-super			Super			AU-super-0		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
1	59.91	**74.41**	55.77	40.20	63.10	42.37	**61.02**	68.87	**56.14**
2	**76.59**	**84.95**	**73.59**	63.25	78.50	60.88	74.75	81.62	68.41
3	**79.57**	84.83	**76.76**	76.42	**85.12**	73.90	76.6	85.00	73.80
4	**83.27**	**90.76**	**81.02**	80.07	86.29	77.18	76.89	85.87	74.21
5	**86.50**	**91.95**	**84.70**	81.88	88.31	80.71	80.11	87.87	77.7
6	**88.65**	**94.03**	**87.10**	85.65	90.52	81.88	85.99	91.22	80.65
7	**90.43**	**94.39**	**89.14**	89.29	91.63	83.15	87.69	92.44	87.11
8	90.39	**95.40**	**89.13**	**90.72**	91.46	87.22	89.85	91.87	89.00
9	**92.16**	**95.85**	**91.09**	91.22	93.79	88.09	91.09	93.77	90.08
10	**93.60**	**96.62**	**92.75**	92.74	94.77	91.71	89.37	94.43	88.01

effects of noise and false labels when working at the pixel scale (Yi et al., 2022). One of the strengths of the proposed method lies in the introduction of data augmentation techniques, which has resulted in further improved superpixel labels, as demonstrated in the ablation experiment. Indeed, not only is more complete information on spatiospectral features preserved, but the overall characteristics and spatial distribution of land covers are also better reflected. In addition, the idea of applying a superpixel label expansion method has proven useful for summarizing information from multiple similar pixels, thereby reducing the uncertainty of individual pixel labels and increasing the stability of the classification model regardless of the number of training samples used.

What are the reasons why the proposed method outperformed other benchmark methods in providing suitable spectral and spatial features for supervised classification? First, by adaptively selecting the optimal superpixel scale, the method avoids tedious manual parameter tuning and effectively solves the problem of superpixel scale selection in hyperspectral image classification. Moreover, superpixel labeling expansion and data augmentation techniques (Table 10) increase the diversity of the training samples, which mitigates the problem of small patterns and improves the model's ability to capture spatial context.

Figures 5–7 show the classification results for the three different datasets when ten samples per class were selected for training. In each figure, (a) and (b) show the pseudo color image and the ground truth, while (c) to (f) present the classification results using different algorithms on the same dataset: (c) AU-Super, (d) BAMS-FE, (e) MSF-PCs, (f) MsuperPCA. By comparing the classification images

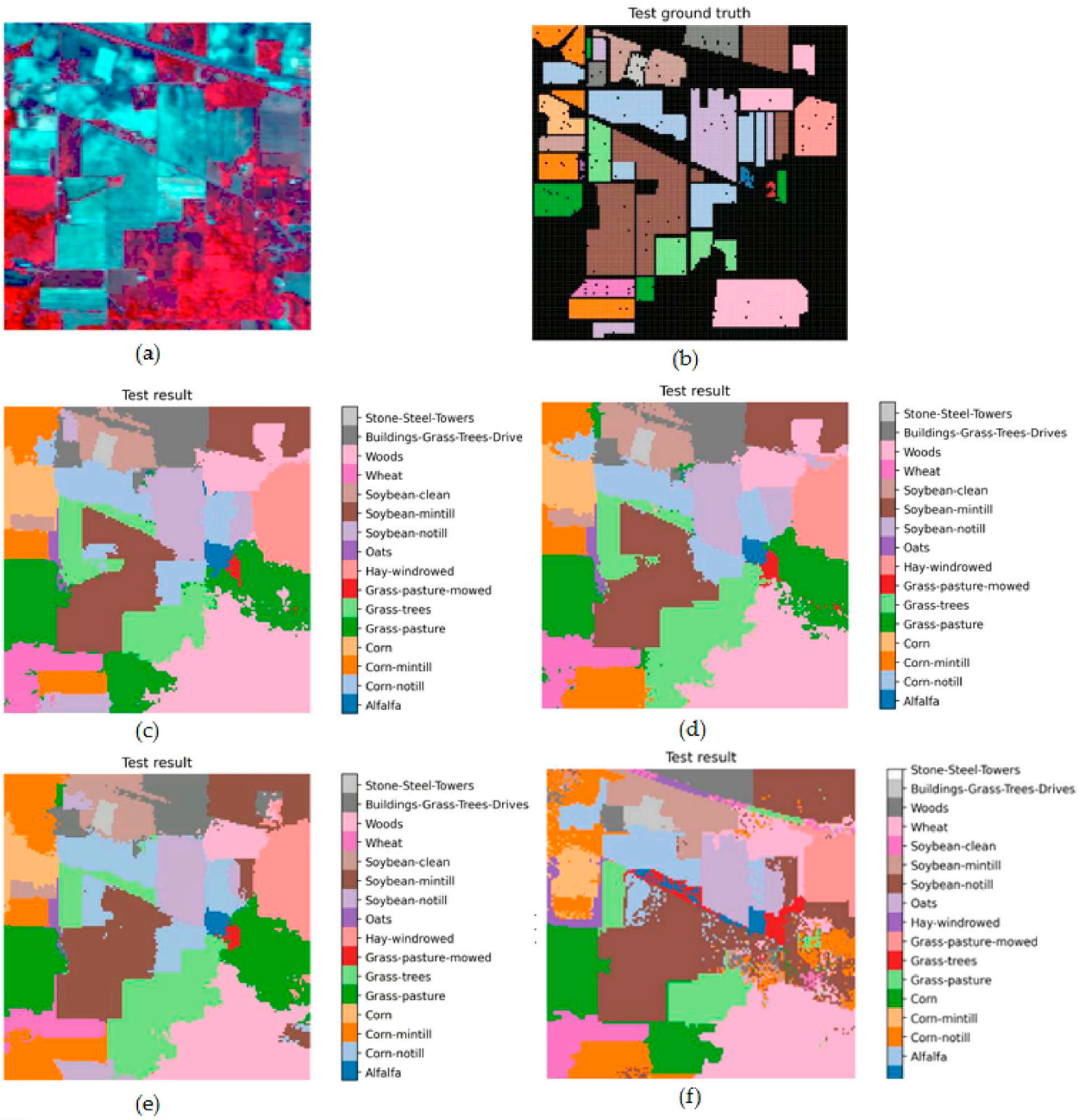

FIGURE 5
Results corresponding to ten samples randomly extracted from each class to train the RF classifier for the Indian Pines dataset. **(a)** Pseudo-color image, **(b)** ground truth, **(c)** AU-Super classification, **(d)** BAMS-FE classification, **(e)** MSF-PCs classification, **(f)** MsuperPCA classification. The quantitative classification results are represented by the OA values as follows: **(c)** 0.936, **(d)** 0.9155, **(e)** 0.8706, and **(f)** 0.8632.

of these four methods on the Indian Pines, Salinas datasets and Pavia University, it can be observed that AU-Super demonstrates stronger boundary preservation capability and spatial consistency across most classes. Particularly, in regions with vague land cover boundaries or limited samples, AU-Super accurately reconstructs true class shapes and significantly reduces salt-and-pepper noise. For example, in forest areas (Figure 4), road regions (Figure 5), and vegetable plots (Figure 6), AU-Super exhibits higher classification consistency and more natural spatial distribution. These results fully validate its superior generalization capability and spatial feature representation under limited sample conditions. As shown in Figures 5–7, AU-Super achieves the highest overall accuracy (OA) on all three datasets (Indian Pines: 0.936 > 0.9155/0.8706/0.8632; Salinas: 0.9960 > 0.9908/0.9846/0.9920; Pavia University: 0.9549 > 0.9423/0.8852/0.8340). Visually, in areas such as forests, roads, and vegetable plots, AU-Super classification results demonstrate clearer class boundaries, more continuous spatial distribution, and significantly reduced salt-and-pepper noise. These observations are consistent with its highest OA values, and together, the quantitative and qualitative results validate its excellent

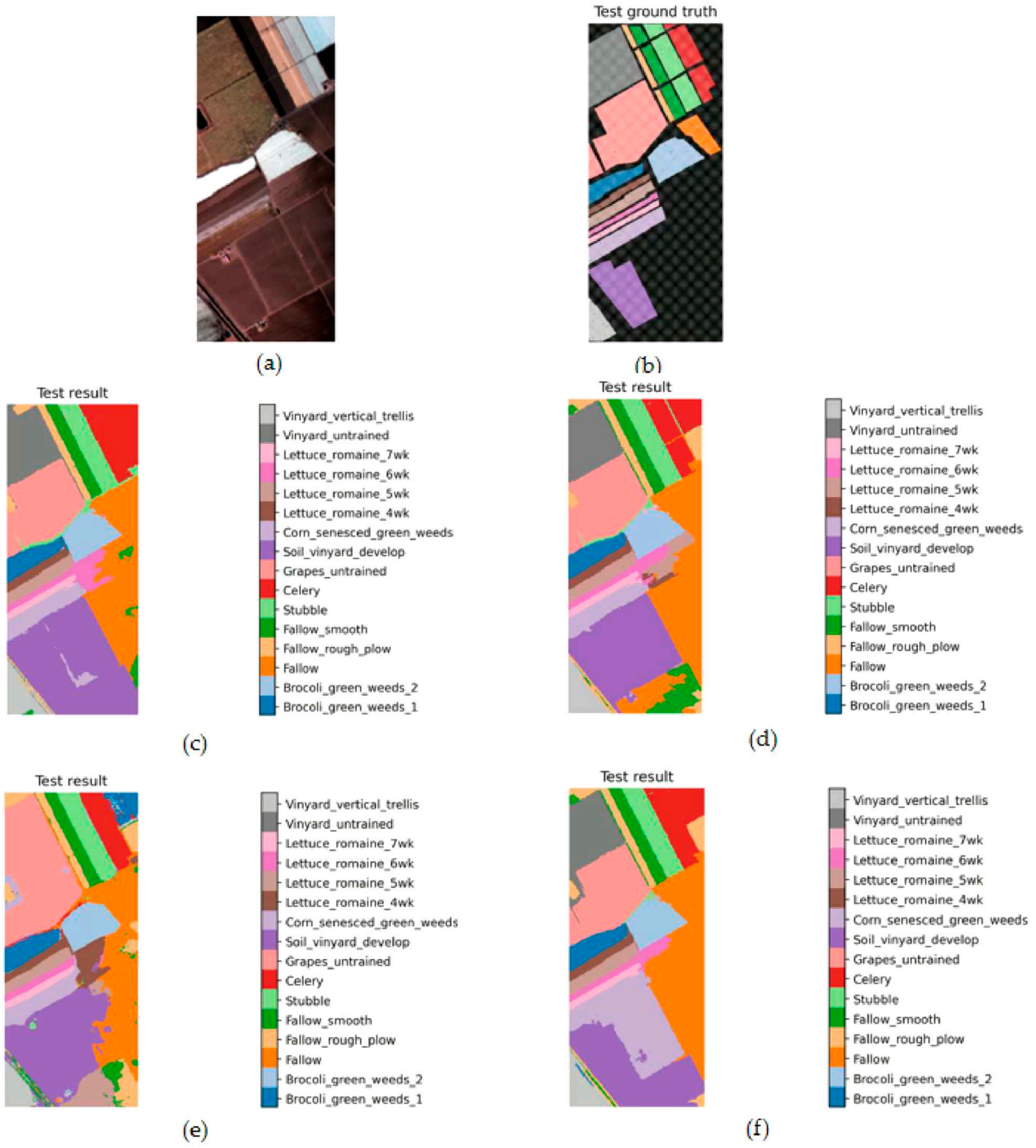

FIGURE 6
Results corresponding to ten samples randomly extracted from each class to train the RF classifier for the Salinas dataset. **(a)** Pseudo-color image, **(b)** ground truth, **(c)** AU-Super classification, **(d)** BAMS-FE classification, **(e)** MSF-PCs classification, **(f)** MsuperPCA classification. The quantitative classification results are represented by the OA values as follows: **(c)** 0.9960, **(d)** 0.9908, **(e)** 0.9846, and **(f)** 0.9920.

generalization capability and spatial feature representation under limited sample conditions.

Overall, experimental results on the Indian Pines, Pavia University, and Salinas datasets show that AU-Super significantly outperforms several SOTA methods addressing deep learning-based feature extraction and dimensionality reduction techniques in terms of classification accuracy. In particular, the proposed method shows excellent performance in OA, AA, and kappa coefficient when working with a very limited number of labeled samples. For example, as can be seen in Table 7, with the expansion of the training set size, and as expected, the overall classification accuracy (OA) of the Indian Pines dataset gradually improves, increasing from an initial value of 0.5991 (one training sample per class) to 0.9360 (ten training samples per class). Of course, increasing the

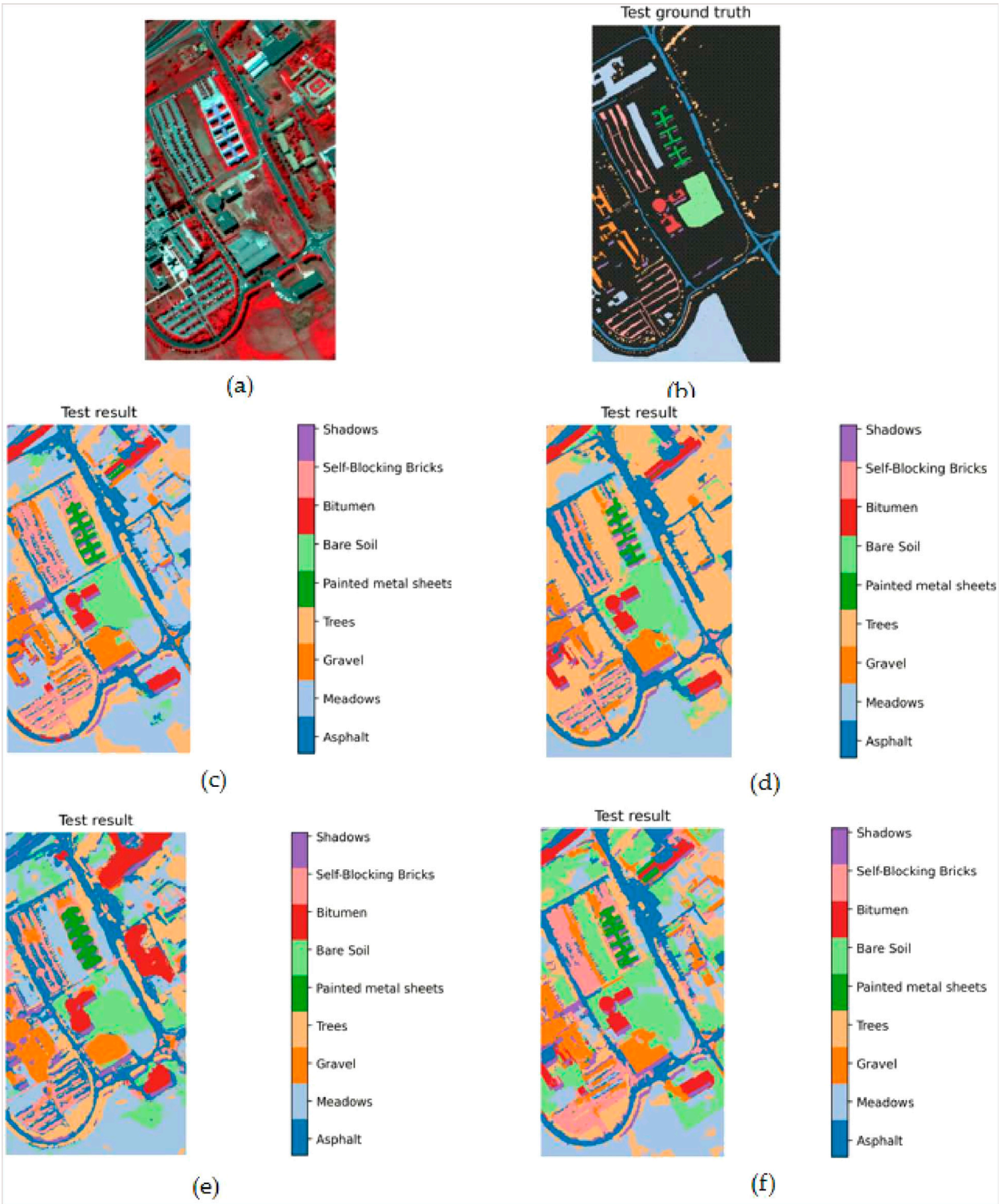

FIGURE 7
Results corresponding to ten samples randomly extracted from each class to train the RF classifier for the Pavia University dataset. **(a)** Pseudo-color image, **(b)** ground truth, **(c)** AU-Super classification, **(d)** BAMS-FE classification, **(e)** MSF-PCs classification, **(f)** MsuperPCA classification. The quantitative classification results are represented by the OA values as follows: **(c)** 0.9579, **(d)** 0.9423, **(e)** 0.8852, **(f)** 0.8340.

number of training samples has a significant positive impact on model performance. But AU-Super is especially efficient, compared to other baseline methods in cases with a limited number of training samples (e.g., from 1 to 5 samples). This fact is also evident in the case of the datasets from the Universities of Salinas and Pavia (Tables 8 and 9), where it is important to highlight that the

improvement in classification accuracy is particularly noticeable, indicating that the model can better learn the features of each category by increasing the training data when samples are scarce. These results not only confirm the positive impact of increasing training samples on model performance but also highlight the superiority of the proposed method in scenarios with limited samples.

When the number of randomly drawn training samples increases to its maximum value of ten, AU-Super achieves overall classification accuracies of 0.9360, 0.9960, and 0.9549 for the Indian Pines, Salinas, and University of Pavia datasets, respectively, demonstrating that the proposed method can achieve very high classification accuracies across different land covers. The algorithm's performance in terms of overall accuracy is particularly notable in the case of urban scenarios (i.e., University of Pavia dataset), where AU-Super outperformed its best competitor by more than 5 percentage points. This result validates the effectiveness of the proposed method for hyperspectral image classification in complex urban scenarios.

Using the University of Pavia dataset as an example for further analysis, it is clear that certain categories are prone to confusion when the training samples per class are between 1 and 4, especially when their sample size is small. However, as the number of training samples increases, the confusion phenomenon is significantly mitigated, and the model's classification ability improves. For example, when the number of training samples increased to 9, the overall accuracy reached 0.9239 and the kappa coefficient 0.9008 (Table 9), indicating a high degree of agreement between the model's predictions and the actual classifications, with the classification performance stabilizing. Nevertheless, some misclassifications still occurred, such as the misclassification of meadows as gravel (misclassification rate of 0.06) and the misclassification of trees as bare soil (misclassification rate of 0.012) (confusion matrix not shown). These misclassifications can be related to spectral similarity between categories, uneven distribution of training samples, or spatial overlap between categories. By selecting ten training samples for each category, the misclassification rate decreases significantly, and the classification accuracy for each category improves significantly, further optimizing the model's overall performance (see Figure 7). In this study, AU-Super-RF demonstrates clear advantages under small-sample conditions, which aligns with the motivation of addressing limited training data. As the number of available samples increases, the relative performance gap compared with conventional methods may gradually narrow. Nevertheless, the proposed method maintains higher stability and robustness across datasets. Future work could further investigate its applicability across a broader range of training sample sizes to provide more comprehensive evidence.

In summary, experimental results on several benchmark hyperspectral datasets (Indian Pines, Salinas, and Pavia University) have shown that AU-Super significantly outperforms existing conventional methods in terms of classification accuracy and stability, especially in complex remote sensing classification scenarios characterized by small samples and high noise. Nevertheless, the generality of the method warrants further consideration. Although AU-Super-RF has been validated on three benchmark datasets, its applicability to large-scale or heterogeneous remote sensing imagery deserves additional attention. When applied to regional-scale data or to sensors with varying spectral and spatial resolutions, adjustments in superpixel search ranges and parameter fine-tuning may be required to ensure optimal performance. Future work will therefore focus on extending the applicability of AU-Super-RF to more diverse remote sensing contexts, which will further demonstrate its practical value.

6 Conclusion

In this work, a novel preprocessing method, called AU-Super, is proposed to improve hyperspectral image classification based on random forest classifier. It includes three complementary stages: i) adaptive superpixel scale selection, ii) superpixel label expansion, and iii) data augmentation. Experimental results, conducted on three widely known hyperspectral datasets, have shown that the proposed method achieved higher classification accuracy compared to SOTA methods. The excellent performance of AU-Super was especially notable when applied with limited labeled samples. This demonstrates that the method can effectively overcome the challenges of hyperspectral image classification, such as high dimensionality, sparse and limited labeling, and spatiospectral variability. Future research could focus on further optimizing the superpixel scale selection strategy to improve the method's adaptability to different types of hyperspectral images. In addition, more advanced spatial-spectral fusion techniques could be explored to improve the model's ability to capture complex spatial relationships. Finally, the method could be extended to classify multitemporal hyperspectral data to enable the classification of dynamic land cover changes. Although the proposed strategy is designed for hyperspectral classification, its modular structure suggests potential applicability to other spatial-spectral analysis tasks, which will be explored in future work. This study is currently limited to public datasets with single-date imagery. Future research will consider extending the framework to multi-temporal data, active learning settings, or integrating deep learning modules for enhanced spectral feature representation. Future research will consider extending the framework to multi-temporal cross-modal data fusion (Yao et al., 2023), active learning for partially observed modalities, and integrating deep cross-modal representation learning modules to enhance feature extraction and alignment.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YW: Investigation, Software, Writing – original draft. LL: Data curation, Funding acquisition, Investigation, Supervision, Writing – review and editing. MX: Methodology, Visualization, Writing – review and editing. SL: Formal Analysis, Project administration, Validation, Writing – review and editing. MY: Formal Analysis, Validation, Writing – review and editing. MA: Conceptualization, Visualization, Writing – review and editing. FA: Conceptualization, Formal Analysis, Validation, Visualization, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work has been supported by the SOTER research project, funded by the Andalusian Technology Corporation through the consortium formed by CONACON S.A., 3D Geospace and the University of Almería, Spain. It also takes part of the general research lines promoted by the Agrifood Campus of International Excellence ceiA3 (http://www.ceia3.es/en).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Analysis Mach. Intell.* 34, 2274–2282. doi:10.1109/tpami.2012.120

Aguilar, M. A., Novelli, A., Nemmaoui, A., Aguilar, F. J., García-Lorca, A. G., and González-Yebra, Ó. (2018). Optimizing multiresolution segmentation for extracting plastic greenhouses from worldview-3 imagery. *Intelligent Interact. Multimedia Syst. Serv. Smart Innovation, Syst. Technol.* 76, 31–40. doi:10.1007/978-3-319-59480-4_4

Ban, Z., Liu, J., and Cao, L. (2018). Superpixel segmentation using Gaussian mixture model. *IEEE Trans. Image Process.* 27, 4105–4117. doi:10.1109/tip.2018.2836306

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS J. Photogrammetry Remote Sens.* 65, 2–16. doi:10.1016/j.isprsjprs.2009.06.004

Brkić, I., Ševrović, M., Medak, D., and Miler, M. (2023). Utilizing high resolution satellite imagery for automated road infrastructure safety assessments. *Sensors* 23, 4405. doi:10.3390/s23094405

Chen, X., Zhang, X., Ren, M., Zhou, B., Sun, M., Feng, Z., et al. (2024). A multiscale enhanced pavement crack segmentation network coupling spectral and spatial information of UAV hyperspectral imagery. *Int. J. Appl. Earth Observation Geoinformation* 128, 103772. doi:10.1016/j.jag.2024.103772

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: an overview. *IEEE Signal Process. Mag.* 35, 53–65. doi:10.1109/msp.2017.2765202

Fang, L., Li, S., Kang, X., and Benediktsson, J. A. (2015a). Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Trans. Geoscience Remote Sens.* 53, 4186–4201. doi:10.1109/tgrs.2015.2392755

Fang, L., Li, S., Duan, W., Ren, J., and Benediktsson, J. A. (2015b). Classification of hyperspectral images by exploiting spectral-spatial information of superpixel *via* multiple kernels. *IEEE Trans. Geoscience Remote Sens.* 53, 6663–6674. doi:10.1109/tgrs.2015.2445767

Fu, H., Sun, G., Ren, J., Zhang, A., and Jia, X. (2022). Fusion of PCA and segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery. *IEEE Trans. Geoscience Remote Sens.* 60, 1–14. doi:10.1109/tgrs.2020.3034656

Gan, Y., Zhang, H., Liu, W., Ma, J., Luo, Y., and Pan, Y. (2024). Local-global feature fusion network for hyperspectral image classification. *Int. J. Remote Sens.* 45, 8548–8575. doi:10.1080/01431161.2024.2403622

Hu, X., Wang, X., Zhong, Y., and Zhang, L. (2022). S3ANet: spectral-spatial-scale attention network for end-to-end precise crop classification based on UAV-borne H2 imagery. *ISPRS J. Photogrammetry Remote Sens.* 183, 147–163. doi:10.1016/j.isprsjprs.2021.10.014

Huang, W., Huang, Y., Wang, H., Liu, Y., and Shim, H. J. (2020). Local binary patterns and superpixel-based multiple kernels for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 13, 4550–4563. doi:10.1109/jstars.2020.3014492

Jia, S., Zhao, Q., Zhuang, J., Tang, D., Long, Y., Xu, M., et al. (2021). Flexible Gabor-based superpixel-level unsupervised LDA for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 59, 10394–10409. doi:10.1109/tgrs.2020.3048994

Jia, W., Pang, Y., and Tortini, R. (2024). The influence of BRDF effects and representativeness of training data on tree species classification using multi-flightline airborne hyperspectral imagery. *ISPRS J. Photogrammetry Remote Sens.* 207, 245–263. doi:10.1016/j.isprsjprs.2023.11.025

Jiang, J., Ma, J., Chen, C., Wang, Z., Cai, Z., and Wang, L. (2018). SuperPCA: a superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans. Geoscience Remote Sens.* 56, 4581–4593. doi:10.1109/tgrs.2018.2828029

Lassalle, G., Ferreira, M. P., Cué La Rosa, L. E., Del'Papa Moreira Scafutto, R., and de Souza Filho, C. R. (2023). Advances in multi- and hyperspectral remote sensing of mangrove species: a synthesis and study case on airborne and multisource spaceborne imagery. *ISPRS J. Photogrammetry Remote Sens.* 195, 298–312. doi:10.1016/j.isprsjprs.2022.12.003

Lee, T.-W. (1998). "Independent component analysis," in *Independent component analysis: theory and applications* (US, Boston: Springer), 27–66.

Li, J., Sheng, H., Xu, M., Liu, S., and Zeng, Z. (2023). BAMS-FE: band-by-band adaptive multiscale superpixel feature extraction for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 61, 1–15. doi:10.1109/tgrs.2023.3294227

Liang, L., Zhang, Y., Zhang, S., Li, J., Plaza, A., and Kang, X. (2023). Fast hyperspectral image classification combining transformers and SimAM-based CNNs. *IEEE Trans. Geoscience Remote Sens.* 61, 1–19. doi:10.1109/tgrs.2023.3309245

Liu, M. Y., Tuzel, O., Ramalingam, S., and Chellappa, R. (2011). "Entropy rate superpixel segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Colorado Springs, CO: IEEE) 2097–2104. doi:10.1109/CVPR.2011.5995323

Liu, Y., Cao, G., Sun, Q., and Siegel, M. (2015). Hyperspectral classification *via* deep networks and superpixel segmentation. *Int. J. Remote Sens.* 36, 3459–3482. doi:10.1080/01431161.2015.1055607

Liu, N., Wagner Hokanson, E., Hansen, N., and Townsend, P. A. (2023). Multi-year hyperspectral remote sensing of a comprehensive set of crop foliar nutrients in cranberries. *ISPRS J. Photogrammetry Remote Sens.* 2. doi:10.1016/j.isprsjprs.2023.10.003

Maćkiewicz, A., and Ratajczak, W. (1993). Principal components analysis (PCA). *Comput. Geosciences* 19, 303–342. doi:10.1016/0098-3004(93)90090-r

Maffei, A., Haut, J. M., Paoletti, M. E., Plaza, J., Bruzzone, L., and Plaza, A. (2020). A single model CNN for hyperspectral image denoising. *IEEE Trans. Geoscience Remote Sens.* 58, 2516–2529. doi:10.1109/tgrs.2019.2952062

Moore, A. P., Prince, S. J. D., Warrell, J., Mohammed, U., and Jones, G. (2008). "Superpixel lattices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Anchorage, AK: IEEE) 1–8. doi:10.1109/CVPR.2008.4587471

Novelli, A., Aguilar, M. A., Nemmaoui, A., Aguilar, F. J., and Tarantino, E. (2016). Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: a case study from Almería (Spain). *Int. J. Appl. Earth Observation Geoinformation* 52, 403–411. doi:10.1016/j.jag.2016.07.011

O'Shea, R. E., Pahlevan, N., Smith, B., Boss, E., Gurlin, D., Alikas, K., et al. (2023). A hyperspectral inversion framework for estimating absorbing inherent optical properties and biogeochemical parameters in inland and coastal waters. *Remote Sens. Environ.* 295, *113706*. doi:10.1016/j.rse.2023.113706

Pang, L., Yao, J., Li, K., and Cao, X. (2025). SPECIAL: zero-shot hyperspectral image classification with CLIP. arXiv Prepr. *arXiv:2501.16222*. doi:10.48550/arXiv.2501.16222

Scheibenreif, L., Mommert, M., and Borth, D. (2023). "Masked vision transformers for hyperspectral image classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops*, (Vancouver, BC: IEEE) 2166–2176.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Soltanzadeh, P., and Hashemzadeh, M. (2021). RCSMOTE: range-controlled synthetic minority over-sampling technique for handling the class imbalance problem. *Inf. Sci.* 542, 92–111. doi:10.1016/j.ins.2020.07.014

Sun, G., Fu, H., Ren, J., Zhang, A., Zabalza, J., Jia, X., et al. (2022). SpaSSA: superpixelwise adaptive SSA for unsupervised spatial-spectral feature extraction in hyperspectral image. *IEEE Trans. Cybern.* 52, 6158–6169. doi:10.1109/tcyb.2021.3104100

Sun, D., Yao, J., Zhou, C., Cao, X., and Ghamisi, P. (2024). Mask approximation net: a novel diffusion model approach for remote sensing change captioning. *arXiv Prepr. arXiv:2412.19179*. doi:10.1109/TGRS.2025.3587261

Wang, C., Zhang, L., Wei, W., and Zhang, Y. (2020). Hyperspectral image classification with data augmentation and classifier fusion. *IEEE Geoscience Remote Sens. Lett.* 17, 1420–1424. doi:10.1109/lgrs.2019.2945848

Wang, X., Liu, J., Chi, W., Wang, W., and Ni, Y. (2023). Advances in hyperspectral image classification methods with small samples: a review. *Remote Sens.* 15, 3795. doi:10.3390/rs15153795

Xanthopoulos, P., Pardalos, P. M., and Trafalis, T. B. (2013). "Linear discriminant analysis," in *Robust data mining* (New York, NY: Springer), 27–33.

Xie, F., Lei, C., Yang, J., and Jin, C. (2019). An effective classification scheme for hyperspectral image based on superpixel and discontinuity preserving relaxation. *Remote Sens.* 11, 1149. doi:10.3390/rs11101149

Xie, F., Wang, R., Jin, C., and Wang, G. (2024). Hyperspectral image classification based on superpixel merging and broad learning system. *Photogrammetric Rec.* 39, 435–456. doi:10.1111/phor.12493

Yan, Q., Jiang, X., Zhang, Y., and Cai, Z. (2022). "Superpixel correction based label propagation for hyperspectral images classification," in *Proceedings of the IEEE international geoscience and remote sensing symposium (IGARSS)*. (Kuala Lumpur, Malaysia; IEEE) 3616–3619. doi:10.1109/IGARSS46834.2022.9884197

Yao, J., Hong, D., Wang, H., Liu, H., and Chanussot, J. (2023). UCSL: toward unsupervised common subspace learning for cross-modal image classification. *IEEE Trans. Geoscience Remote Sens.* 61, 1–12. doi:10.1109/tgrs.2023.3282951

Yasir, M., Jianhua, W., Shanwei, L., Sheng, H., Mingming, X., and Hossain, M. (2023). Coupling of deep learning and remote sensing: a comprehensive systematic literature review. *Int. J. Remote Sens.* 44, 157–193. doi:10.1080/01431161.2022.2161856

Yi, R., Huang, Y., Guan, Q., Pu, M., and Zhang, R. (2022). Learning from pixel-level label noise: a new perspective for semi-supervised semantic segmentation. *IEEE Trans. Image Process.* 31, 623–635. doi:10.1109/tip.2021.3134142

Zhang, Q., Yuan, Q., Song, M., Yu, H., and Zhang, L. (2022a). Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising. *IEEE Trans. Image Process.* 31, 6356–6368. doi:10.1109/tip.2022.3211471

Zhang, S., Lu, T., Fu, W., and Li, S. (2022b). Superpixel-level hybrid discriminant analysis for hyperspectral image feature extraction. *IEEE Trans. Geoscience Remote Sens.* 60, 1–12. doi:10.1109/TGRS.2022.3214523

Zhang, S., Lu, T., Li, S., and Fu, W. (2022c). Superpixel-based Brownian descriptor for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 60, 1–12.

Zhang, S., Lu, T., Li, S., and Fu, W. (2022d). Superpixel-based Brownian descriptor for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 60, 1–12. doi:10.1109/tgrs.2021.3133878

Zhang, Q., Dong, Y., Zheng, Y., Yu, H., Song, M., Zhang, L., et al. (2024). Three-dimension spatial–spectral attention transformer for hyperspectral image denoising. *IEEE Trans. Geoscience Remote Sens.* 62, 1–13. doi:10.1109/tgrs.2024.3458174

Zhang, Y., Duan, P., Liang, L., Kang, X., Li, J., and Plaza, A. (2025). PFS3F: probabilistic fusion of superpixel-wise and semantic-aware structural features for hyperspectral image classification. *IEEE Trans. Circuits Syst. Video Technol.* 35, 8723–8737. doi:10.1109/tcsvt.2025.3556548

Zhang, Q., Zheng, Y., Yuan, Q., Song, M., Yu, H., and Xiao, Y. (2023). Hyperspectral image denoising: from model-driven, data-driven, to model-data-driven. *IEEE Trans. Neural Netw. Learn. Syst.* 35, 13143–13163. doi:10.1109/tnnls.2023.3278866

Zhang, X., Su, Y., Gao, L., Bruzzone, L., Gu, X., and Tian, Q. (2023). A lightweight transformer network for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 61, 1–17. doi:10.1109/tgrs.2023.3297858

Zhao, J., Bo, R., Hou, Q., Cheng, M. M., and Rosin, P. (2018). FLIC: fast linear iterative clustering with active search. *Comput. Vis. Media* 4, 333–348. doi:10.1007/s41095-018-0123-y

Zhao, X., Zhang, S., Yan, W., and Pan, X. (2023). Multi-temporal grass hyperspectral classification *via* full pixel decomposition spectral manifold projection and boosting active learning model. *Int. J. Remote Sens.* 44, 469–491. doi:10.1080/01431161.2023.2165422

Zhou, L., Zhu, J., Yang, J., and Geng, J. (2022). "Data augmentation and spatial-spectral residual framework for hyperspectral image classification using limited samples," in *Proceedings of the IEEE international conference on unmanned systems (ICUS)*, (Guangzhou, China: IEEE) 490–495. doi:10.1109/ICUS55513.2022.9986968

Zhou, H., Luo, F., Zhuang, H., Weng, Z., Gong, X., and Lin, Z. (2023). Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification. *IEEE Trans. Geoscience Remote Sens.* 61, 1–14. doi:10.1109/tgrs.2023.3265879