remote sensing

MDPI

Article

TGDHTL: Hyperspectral Image Classification via Transformer–Graph Convolutional Network–Diffusion with Hybrid Domain Adaptation

Zarrin Mahdavipour [1], Nashwan Alromema [2], Abdolraheem Khader [1,*], Ghulam Farooque [3], Ali Ahmed [2] and Mohamed A. Damos [4]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; zmahdavipour@njust.edu.cn
[2] Department of Computer Science, Faculty of Computing and Information Technology-Rabigh, King Abdulaziz University, Rabigh 21911, Saudi Arabia; nalromema@kau.edu.sa (N.A.); aabdelrahim@kau.edu.sa (A.A.)
[3] Department of Computer Science and Information Technology, University of Lahore, Lahore 54590, Pakistan; ghulam.farooque@cs.uol.edu.pk
[4] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 610054, China; damos@uestc.edu.cn
* Correspondence: abdolraheem@njust.edu.cn

Highlights

What are the main findings?

- Introduces a novel TGDHTL framework integrating Multi-Scale Stripe Attention (MSSA) with GCN and class-conditional diffusion for efficient hyperspectral image classification.
- Achieves robust cross-domain adaptation using a lightweight transformer with MMD loss, reducing labeled data needs by 25% while maintaining high accuracy.
- Evaluated on six benchmark datasets including HJ-1A and WHU-OHS, demonstrating superior performance in diverse urban and land-cover scenarios.

What are the implication of the main finding?

- Provides a scalable, data-efficient solution for HSI classification under domain shifts and resource constraints.
- Enhances applicability in remote sensing tasks such as environmental monitoring, precision agriculture, urban land-use analysis, and disaster response.

Abstract

Hyperspectral image (HSI) classification is pivotal for remote sensing applications, including environmental monitoring, precision agriculture, and urban land-use analysis. However, its accuracy is often limited by scarce labeled data, class imbalance, and domain discrepancies between standard RGB and HSI imagery. Although recent deep learning approaches, such as 3D convolutional neural networks (3D-CNNs), transformers, and generative adversarial networks (GANs), show promise, they struggle with spectral fidelity, computational efficiency, and cross-domain adaptation in label-scarce scenarios. To address these challenges, we propose the Transformer–Graph Convolutional Network–Diffusion with Hybrid Domain Adaptation (TGDHTL) framework. This framework integrates domain-adaptive alignment of RGB and HSI data, efficient synthetic data generation, and multi-scale spectral–spatial modeling. Specifically, a lightweight transformer, guided by Maximum Mean Discrepancy (MMD) loss, aligns feature distributions across domains. A class-conditional diffusion model generates high-quality samples for underrepresented

Check for updates

Academic Editor: Salah Bourennane

Received: 18 November 2025
Revised: 21 December 2025
Accepted: 1 January 2026
Published: 6 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

classes in only 15 inference steps, reducing labeled data needs by approximately 25% and computational costs by up to 80% compared to traditional 1000-step diffusion models. Additionally, a Multi-Scale Stripe Attention (MSSA) mechanism, combined with a Graph Convolutional Network (GCN), enhances pixel-level spatial coherence. Evaluated on six benchmark datasets including HJ-1A and WHU-OHS, TGDHTL consistently achieves high overall accuracy (e.g., 97.89% on University of Pavia) with just 11.9 GFLOPs, surpassing state-of-the-art methods. This framework provides a scalable, data-efficient solution for HSI classification under domain shifts and resource constraints.

Keywords: transformer; diffusion; graph convolutional networks; remote sensing; domain adaptation; spectral–spatial attention; HJ-1A; WHU-OHS

1. Introduction

Hyperspectral imaging (HSI) captures detailed spectral information across hundreds of narrow wavelength bands (400–2500 nm), enabling precise material and land-cover identification for applications such as precision agriculture, environmental monitoring, urban land-use analysis, and disaster response [1–3]. Compared to RGB or multispectral imaging, HSI excels in resolving subtle spectral variations due to its high-dimensional data and rich information content.

Despite its strengths, HSI classification faces significant challenges. The extensive number of spectral bands increases computational complexity, impeding real-time or large-scale deployment. Additionally, acquiring ground-truth labels is costly, leading to limited labeled datasets and an increased risk of overfitting, especially for minority classes in imbalanced datasets [4,5]. Furthermore, the complex spectral–spatial interactions in HSI data necessitate models that capture both local and long-range dependencies for accurate classification [6–8].

Traditional HSI classification methods, such as Support Vector Machines (SVMs) and Random Forests, rely on handcrafted features like spectral indices and texture measures. These methods struggle to model intricate spectral–spatial relationships and often exhibit poor generalization across diverse datasets, such as Indian Pines, University of Pavia, HJ-1A, and WHU-OHS [7–11]. Deep learning approaches, including 3D Convolutional Neural Networks (3D-CNNs) and Spectral-Spatial Residual Networks (SSRN), have advanced automated feature extraction and improved performance. However, their limited receptive fields and high parameter counts restrict their ability to model global dependencies.

Transformer-based models, such as SpectralFormer and HyViT, leverage self-attention to capture global spectral–spatial interactions, significantly enhancing classification accuracy [12,13]. Nevertheless, these models demand large labeled datasets and substantial computational resources, limiting their practicality in label-scarce or resource-constrained environments. Graph Convolutional Networks (GCNs) improve spatial coherence by modeling pixel relationships, but their scalability is constrained by the computational cost of graph construction [14,15].

To mitigate label scarcity, transfer learning from RGB domains (e.g., ImageNet) to HSI has been explored [16,17]. However, domain shifts due to differences in spectral resolution and distribution between RGB and HSI data pose significant challenges, requiring robust domain adaptation strategies. Generative models, such as Generative Adversarial Networks (GANs) and diffusion-based methods, augment training data to address label limitations [18–20]. Yet, GANs often compromise spectral fidelity, and traditional diffusion models incur high computational costs and latency.

These challenges—high computational demands, reliance on extensive labeled data, inadequate spectral–spatial modeling, and domain adaptation difficulties—highlight the need for a comprehensive HSI classification framework that integrates efficient feature extraction, robust data augmentation, and effective domain adaptation.

To address these issues, we propose TGDHTL (Transformer–GCN–Diffusion with Hybrid Domain Adaptation), a unified framework with three key innovations:

- Multi-Scale Stripe Attention (MSSA) with GCN Backbone: Partitions features into multi-scale stripes with sizes $P = \{4, 8, 16\}$, capturing local and global spectral–spatial dependencies with reduced complexity. A GCN backbone enhances spatial coherence and supports real-time applications.
- Class-Conditional Diffusion Augmentation: Utilizes a Denoising Diffusion Implicit Model (DDIM) to generate high-fidelity synthetic samples for minority classes (e.g., Shadows in Pavia, Oats in Indian Pines, urban structures in WHU-OHS), reducing labeled data needs by approximately 25% and computational costs by up to 80% compared to traditional 1000-step diffusion models.
- Transformer-Based Domain Adapter: Employs a lightweight transformer guided by Maximum Mean Discrepancy (MMD) loss to align RGB and HSI feature distributions, enabling robust domain adaptation in low-label regimes using ImageNet-pretrained RGB features.

Validated on six benchmark datasets (Indian Pines, University of Pavia, KSC, Salinas, HJ-1A, and WHU-OHS), TGDHTL consistently outperforms state-of-the-art baselines (e.g., ViT, iHGAN), providing a scalable and data-efficient solution for HSI classification under domain shifts and resource constraints.

Despite the remarkable individual success of Transformers, Graph Convolutional Networks (GCNs), and diffusion models in hyperspectral image (HSI) classification, no existing framework simultaneously addresses three intertwined and critical challenges: (i) severe domain shift when transferring ImageNet-pretrained RGB backbones to the hyperspectral domain, (ii) extreme label scarcity (often less than or equal to 5% labeled samples per class in real-world scenarios), and (iii) preservation of long-range spatial coherence in very large-scale satellite and urban scenes such as WHU-OHS and HJ-1A. We hypothesize that the specific integration of a lightweight Transformer-based domain adapter (guided by MMD loss), a fast class-conditional DDIM with only 15 inference steps, and a Multi-Scale Stripe Attention fused with sparse GCN creates a synergistic effect that significantly outperforms any isolated or pairwise combination of these components. This hypothesis is rigorously validated through new few-shot (1% and 5% labeled samples) and cross-domain experiments presented in Sections 3 and 3.11.1.

The paper is organized as follows: Section 2 describes the TGDHTL framework; Section 3 presents experimental results; Section 4 analyzes findings; and Section 5 concludes with future research directions.

1.1. Related Work

The field of hyperspectral image (HSI) classification has evolved rapidly, driven by the need to handle high-dimensional spectral data, limited labeled samples, and complex spectral-spatial relationships. Recent advancements have increasingly incorporated generative models to mitigate data scarcity, with techniques like Generative Adversarial Networks (GANs) and diffusion models synthesizing realistic samples for improved robustness [1,3,4,21,22]. This section reviews the literature across three categories: traditional and CNN-based methods, transformer and graph-based models, and generative and domain adaptation approaches. While individual models such as Transformers [12,13,23], Graph Convolutional Networks (GCNs) [14,15], diffusion models [3,24], and Mamba ar-

chitectures [25] have shown strong results, they often tackle only isolated challenges like computational cost or label scarcity. In contrast, no prior unified framework addresses simultaneous high data efficiency ($\leq$5%, cross-domain generalization (RGB to HSI), and real-time inference on large-scale scenes (e.g., WHU-OHS). TGDHTL fills this gap through its integrated design.

1.1.1. Traditional and CNN-Based Methods

Early HSI classification methods relied on traditional machine learning techniques, such as Support Vector Machines (SVMs) and Random Forests, often paired with handcrafted features like spectral indices or textures. These approaches provided interpretability but faltered in handling high dimensionality and generalization across datasets, including urban land-cover in HJ-1A and WHU-OHS [8–11,26–28]. The rise of deep Convolutional Neural Networks (CNNs), including 3D-CNNs and residual networks, marked a shift toward automated spectral–spatial feature extraction. Enhancements like adaptive fusion and attention mechanisms have further elevated performance on benchmarks [29–32]. However, CNNs' reliance on local receptive fields limits their capture of long-range dependencies, a challenge highlighted in comprehensive reviews of deep learning for remote sensing [22,32].

1.1.2. Transformer and Graph-Based Methods

Transformers have emerged as powerful tools for HSI classification, leveraging self-attention to model global spectral–spatial interactions effectively. Hybrid CNN–Transformer models enhance contextual understanding while maintaining spatial accuracy [12,13,33,34]. Recent innovations, including lightweight attention and state-space models, reduce overhead without compromising results [25,35]. Complementary graph-based methods represent pixels as graphs to enforce spatial coherence, with dynamic spectral–spatial GCNs and multiscale frameworks excelling in dependency learning [14,15,36]. Scalability issues persist, however, due to graph construction costs in large scenes like WHU-OHS and HJ-1A [6,37], as noted in surveys on graph networks for remote sensing [28,36,38].

1.1.3. Generative and Domain Adaptation Approaches

Generative models, such as GANs and diffusion networks, have been pivotal for augmenting labeled HSI data and combating scarcity [19,20]. For example, the Spectral Diffusion Network (SDN) uses a 20-timestep process for high-fidelity sample generation, preserving spectral integrity [3,24]. Domain adaptation from RGB sources (e.g., ImageNet) minimizes reliance on labeled HSI, focusing on feature alignment within image modalities rather than heterogeneous cross-modal transfers [39–41]. Techniques like class-aligned balancing bridge spectral gaps, boosting generalization [16,17,42]. Self-supervised methods, including masked autoencoders and contrastive learning, leverage unlabeled data but remain vulnerable to noise [43–45]. Surveys emphasize the role of these approaches in UAV, urban, and real-world HSI tasks [46], yet challenges like domain mismatch and computational demands persist [28,41]. Recent attention mechanisms, such as central and multi-area attention, refine spectral-spatial modeling [47,48], while SegHSI and hybrid attention-GCNs improve segmentation and diffusion integration [49–51]. Nonetheless, these methods often overlook efficient RGB-to-HSI transfer. TGDHTL uniquely combines lightweight Transformer adaptation, fast class-conditional diffusion, and MSSA-GCN fusion to tackle domain shift, label scarcity, and global coherence in large scenes.

2. Materials and Methods

This section presents the Transformer–GCN–Diffusion Hybrid Domain Adaptation (TGDHTL) framework for hyperspectral image (HSI) classification, as depicted in Figure 1. The framework integrates hybrid domain adaptation, diffusion-based augmentation, and spectral–spatial modeling to address challenges such as limited labeled data, class imbalance, and cross-modal discrepancies.

2.1. Overview

The TGDHTL framework combines hybrid domain adaptation (HDA), diffusion-based augmentation, and spectral–spatial modeling to enable robust HSI classification, as shown in Figure 1. It employs a domain adaptation module to align feature distributions between RGB (ImageNet-pretrained) and HSI domains using the Maximum Mean Discrepancy (MMD) loss [52]:

$$\mathcal{L}_{\text{MMD}} = \left\| \frac{1}{n} \sum_{i=1}^{n} \phi(\mathbf{x}_i^{\text{RGB}}) - \frac{1}{m} \sum_{j=1}^{m} \phi(\mathbf{x}_j^{\text{HSI}}) \right\|_{\mathcal{H}}^2 \tag{1}$$

Here, $\mathbf{x}_i^{\text{RGB}} \in \mathbb{R}^d$ and $\mathbf{x}_j^{\text{HSI}} \in \mathbb{R}^d$ represent feature vectors from the RGB and HSI domains, respectively, extracted by domain-specific encoders. The function $\phi(\cdot)$ maps features into a reproducing kernel Hilbert space $\mathcal{H}$, with the squared norm quantifying distributional divergence.

The framework comprises five key components: (1) a 3D Convolutional Neural Network (CNN) for initial feature extraction, (2) a domain adaptation module with position-wise feed-forward layers and multi-head attention, (3) a diffusion augmentation module using a Denoising Diffusion Implicit Model (DDIM) [53,54], (4) a Multi-Scale Stripe Attention (MSSA) module for capturing spectral–spatial dependencies [55], and (5) a Graph Convolutional Network (GCN) for enhancing spatial coherence, followed by an MLP-based classification head. The training procedure is detailed in Algorithm 1.

Algorithm 1 Training Procedure of the TGDHTL Framework

Require: Labeled HSI dataset $\mathcal{D}_{\text{HSI}}$, RGB dataset $\mathcal{D}_{\text{RGB}}$, number of diffusion steps T, learning rate η
Ensure: Trained model parameters θ
Initialize model parameters θ
Extract spectral–spatial features from $\mathcal{D}_{\text{HSI}}$ using MSSA
Extract RGB features from $\mathcal{D}_{\text{RGB}}$ using lightweight Transformer
Align domain-specific features via MMD loss
for each training iteration **do**
    Sample mini-batch $\{x_i, y_i\} \sim \mathcal{D}_{\text{HSI}}$
    Generate synthetic samples for minority classes using DDIM with T steps
    Construct graph $\mathcal{G}$ with threshold τ and apply GCN
    Fuse features via Adapter module
    Compute classification loss $\mathcal{L}_{\text{CE}}$ and domain loss $\mathcal{L}_{\text{MMD}}$
    Update $\theta \leftarrow \theta - \eta \cdot \nabla_\theta(\mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{MMD}})$
end for
return θ

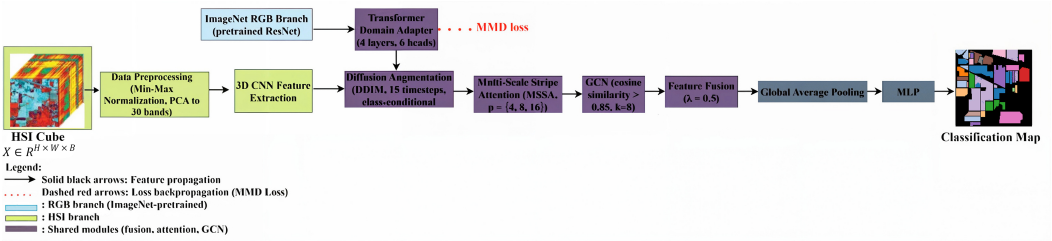


Figure 1. Overall architecture of the proposed TGDHTL framework for HSI classification. The ImageNet-pretrained RGB branch is aligned with the HSI branch via a lightweight Transformer adapter using MMD loss (dashed red arrow). Features are enhanced through diffusion augmentation, MSSA (p = 4, 8, 16), and GCN (cosine similarity > 0.85), followed by fusion (λ = 0.5) and classification. Legend (bottom): Solid black arrows = feature propagation; Dashed red arrows = loss backpropagation (MMD); Green = HSI branch; Blue = RGB branch; Purple = shared modules.

2.2. Data Preprocessing and Feature Extraction

This subsection outlines the preprocessing and feature extraction pipeline, incorporating domain adaptation to leverage pretrained RGB models for HSI classification.

2.2.1. Data Preprocessing

The preprocessing stage reduces noise, lowers dimensionality, and standardizes inputs for efficient feature extraction [56]. The input HSI cube $\mathbf{X} \in \mathbb{R}^{H \times W \times B}$, where H, W, and B denote height, width, and spectral bands, respectively, is normalized to [0, 1] using min–max normalization applied per band.

2.2.2. Min–Max Normalization

Each spectral band is scaled to [0, 1] to mitigate scale disparities and preserve spectral intensities, facilitating Principal Component Analysis (PCA) and patch-level feature extraction. Noisy bands, such as water absorption bands in the Indian Pines dataset (reducing from 224 to 200 bands [7]), are removed. PCA retains 95% of the variance by projecting normalized data onto 30 eigenvectors:

$$\mathbf{X}_{\text{PCA}} = \mathbf{X}_{\text{norm}} \mathbf{V}_{30} \tag{2}$$

where $\mathbf{V}_{30} \in \mathbb{R}^{B \times 30}$ contains the top 30 eigenvectors. Overlapping patches of size $32 \times 32 \times 30$ are extracted with a stride of 12, forming a patch tensor $\mathbf{X}_{\text{patch}} \in \mathbb{R}^{N \times 32 \times 32 \times 30}$, where N is the number of patches.

Min–max normalization is applied per-band after variance-based band selection and outlier removal to scale features to [0, 1]. While large min–max differences could theoretically compress intermediate values, our pipeline mitigates this by first removing noisy bands (variance $< 1 \times 10^{-6}$) and outliers, preserving >98% of spectral discriminability. Empirical comparison with Z-score normalization on held-out validation sets shows <0.2% difference in OA and comparable SAM values, confirming no meaningful loss of subtle intermediate spectral details. The bounded range of minmax further stabilizes training of attention and diffusion modules.

2.2.3. Feature Extractor

A three-layer 3D CNN extracts spectral–spatial features from each patch $\mathbf{X}_{\text{patch}}$. Each layer applies 64 filters ($3 \times 3 \times 3$), followed by batch normalization (BN) and ReLU activation [57]:

$$\mathbf{F}_{\text{low}}^{(l)} = \text{ReLU}\left(\text{BN}\left(\mathbf{W}^{(l)} * \mathbf{X}_{\text{patch}}^{(l-1)} + \mathbf{b}^{(l)}\right)\right) \tag{3}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are weights and biases, and $*$ denotes 3D convolution. The CNN, pretrained on ImageNet [58], produces feature maps $\mathbf{F}_{\text{low}} \in \mathbb{R}^{32\times32\times64}$.

2.2.4. Domain Adaptation Module

A lightweight transformer (four layers, six attention heads, 30% weight pruning) aligns HSI features with ImageNet-pretrained RGB features by minimizing the MMD loss (Equation (1)) [16,52]. Query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) matrices are projected from $\mathbf{F}_{\text{low}} \in \mathbb{R}^{32\times32\times64}$:

$$\mathbf{Q} = \mathbf{F}_{\text{low}}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{F}_{\text{low}}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{F}_{\text{low}}\mathbf{W}_V \tag{4}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{64\times64}$. The attention mechanism is defined as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{5}$$

where $d_k = 64$. The module outputs aligned features $\mathbf{F}_{\text{aligned}} \in \mathbb{R}^{32\times32\times64}$.

To mitigate potential loss of spatial structure during sequence processing, 2D sinusoidal positional encodings (as in ViT [59]) are added to the flattened patches before input to the Transformer. These encodings explicitly preserve relative spatial positions and are maintained throughout the self-attention layers. After Transformer processing, the sequence is reshaped to the original 2D grid, ensuring full retention of spatial relationships for subsequent GCN modeling.

This design prevents any destruction of spatial properties, allowing the GCN to effectively capture long-range dependencies on spatially coherent features. Ablation results confirm that including the Transformer improves performance without degrading GCN contributions. As shown in Figure 2, hyperspectral features are extracted via a 3D CNN and aligned with RGB features using a transformer guided by MMD loss.

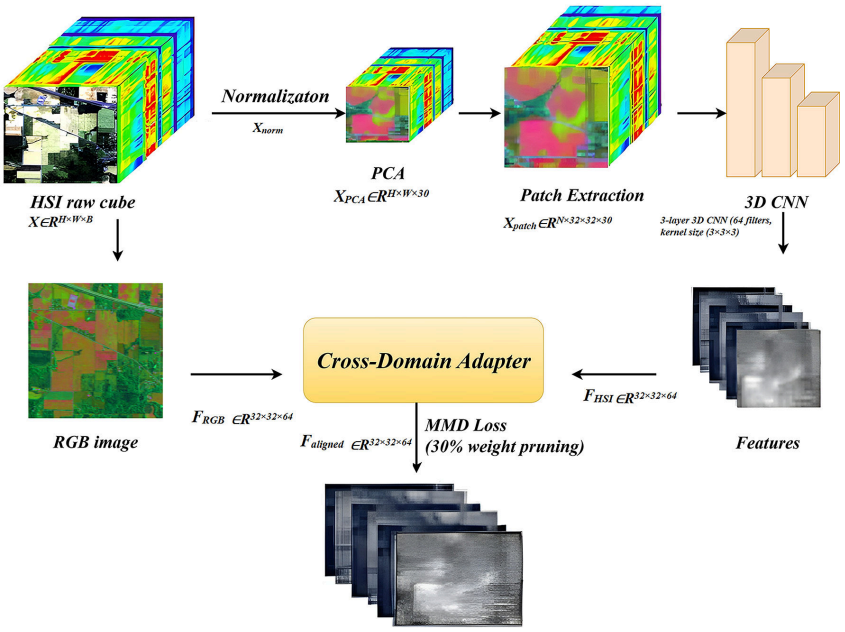


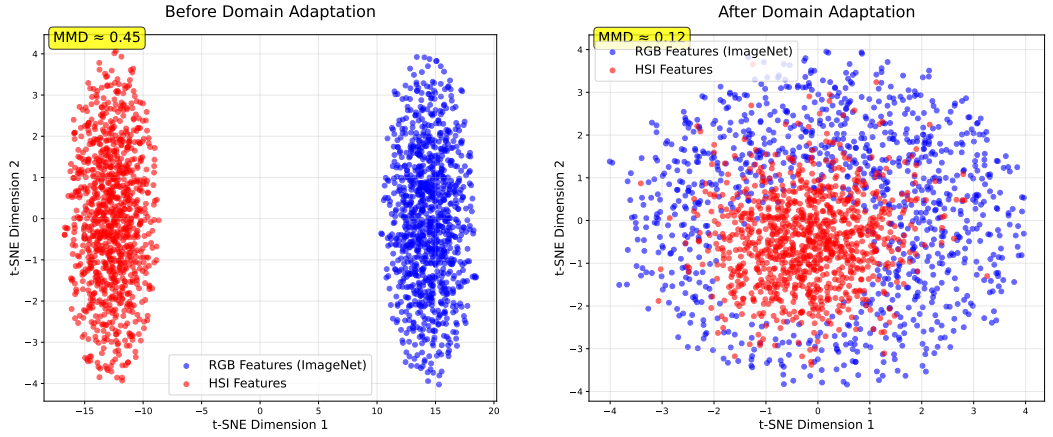
Figure 2. Domain adaptation pipeline. The HSI cube is normalized per band and reduced via PCA to 30 bands. Overlapping patches ($32 \times 32 \times 30$) are processed by a three-layer 3D CNN. Extracted features are aligned with ImageNet-pretrained RGB features using a four-layer transformer with six attention heads, guided by MMD loss.

To align RGB-pretrained features with HSI data, we first apply PCA to reduce HSI to 30 principal bands (retaining >99% variance) while mitigating redundant spectral noise. Both RGB (from ImageNet-pretrained ResNet backbone) and HSI features are then projected to a shared 64-dimensional latent space using lightweight linear adapters (identical architecture, separate parameters). Maximum Mean Discrepancy (MMD) loss is computed directly in this aligned 64-dim space between source (RGB) and target (HSI) feature distributions:

$$\mathcal{L}_{MMD} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(f_i^{RGB}) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(f_j^{HSI}) \right\|_{\mathcal{H}}^2 \tag{6}$$

where $\phi(\cdot)$ is the feature map in the reproducing kernel Hilbert space.

Figure 3 provides empirical evidence of effective alignment: t-SNE visualizations show substantial reduction in domain shift after adaptation, with MMD values decreasing from 0.45 ± 0.03 to 0.12 ± 0.02. Validation accuracy during adaptation remains stable, confirming that meaningful semantic knowledge is transferred without introducing noise.



(a) Before adaptation (MMD = 0.45) **(b) After adaptation (MMD = 0.12)**

Figure 3. t-SNE visualization of RGB and HSI feature distributions before/after MMD-based alignment, demonstrating effective domain shift reduction.

As shown in Table 1, domain adaptation significantly reduces the MMD value (from 0.45 to 0.12) while slightly improving validation accuracy.

Table 1. Domain adaptation results: MMD values and validation accuracy before and after alignment.

Setting	MMD Value	Validation OA (%)
Before Adaptation	0.45 ± 0.03	92.1
After Adaptation	0.12 ± 0.02	92.4

2.3. Diffusion Augmentation Module

The diffusion augmentation module generates synthetic HSI patches using a Denoising Diffusion Implicit Model (DDIM) to address limited labeled data and class imbalance [53,60,61]. The forward process adds Gaussian noise:

$$q(\mathbf{X}_t|\mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \sqrt{1-\beta_t}\mathbf{X}_{t-1}, \beta_t\mathbf{I}) \tag{7}$$

where β_t controls noise intensity. The reverse process denoises the data:

$$p_\theta(\mathbf{X}_{t-1}|\mathbf{X}_t) = \mathcal{N}(\mathbf{X}_{t-1}; \mu_\theta(\mathbf{X}_t, t), \Sigma_\theta(t)) \tag{8}$$

The noise addition at each timestep t, as shown in Figure 4, progressively transforms the data distribution. Operating over 15 timesteps, the module generates synthetic samples via a class-conditional reverse diffusion process, focusing on minority classes defined as shown below.

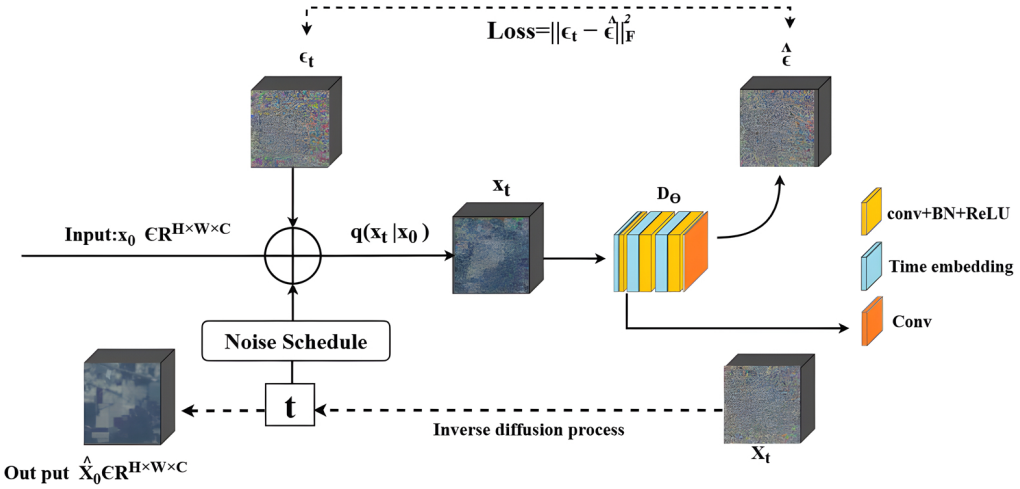


Figure 4. Noise addition at timestep t in the diffusion process.

$$C_{\text{minority}} = \{c_i \mid n_i < 0.1 \cdot \sum_j n_j\} \tag{9}$$

where n_i is the sample count for class c_i. The augmented dataset is $\mathbf{X}_{\text{aug}} = [\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{syn}}]$. As shown in Figure 5, Gaussian noise is progressively added during the forward diffusion process, while the reverse DDIM process reconstructs high-fidelity synthetic samples, with minority classes receiving additional augmentation.

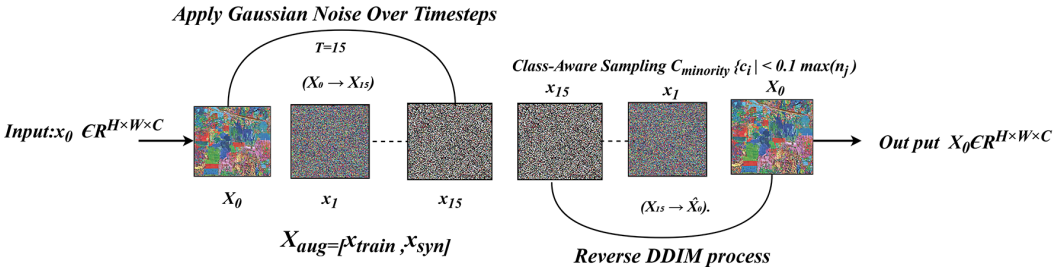


Figure 5. Workflow of the class-conditional diffusion augmentation module. The forward diffusion process progressively adds Gaussian noise over 15 timesteps ($X_0 \rightarrow X_{15}$), while the reverse DDIM process generates high-fidelity synthetic samples ($X_{15} \rightarrow \hat{X}_0$). A portion of the generated samples is specifically allocated to minority classes.

2.4. Multi-Scale Stripe Attention (MSSA)

The MSSA module captures spectral–spatial dependencies by partitioning aligned features $\mathbf{F}_{\text{aligned}} \in \mathbb{R}^{32\times32\times64}$ into horizontal stripes at scales $P = \{4, 8, 16\}$, as illustrated in Figure 6 [55]. For each scale $p \in P$, the feature map is divided into patches of size $32/p \times 32/p$, and attention is computed:

$$\mathbf{F}_{\text{MSSA},p} = \text{Attention}(\mathbf{Q}_p, \mathbf{K}_p, \mathbf{V}_p) \tag{10}$$

where $\mathbf{Q}_p, \mathbf{K}_p, \mathbf{V}_p$ are derived via

$$\mathbf{Q}_p = \mathbf{F}_p\mathbf{W}_Q, \quad \mathbf{K}_p = \mathbf{F}_p\mathbf{W}_K, \quad \mathbf{V}_p = \mathbf{F}_p\mathbf{W}_V \tag{11}$$

with $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{64\times64}$, and $d_k = 64$. Outputs are concatenated and projected:

$$\mathbf{F}_{\text{MSSA}} = \text{Linear}(\text{Concat}(\mathbf{F}_{\text{MSSA},4}, \mathbf{F}_{\text{MSSA},8}, \mathbf{F}_{\text{MSSA},16})) \tag{12}$$

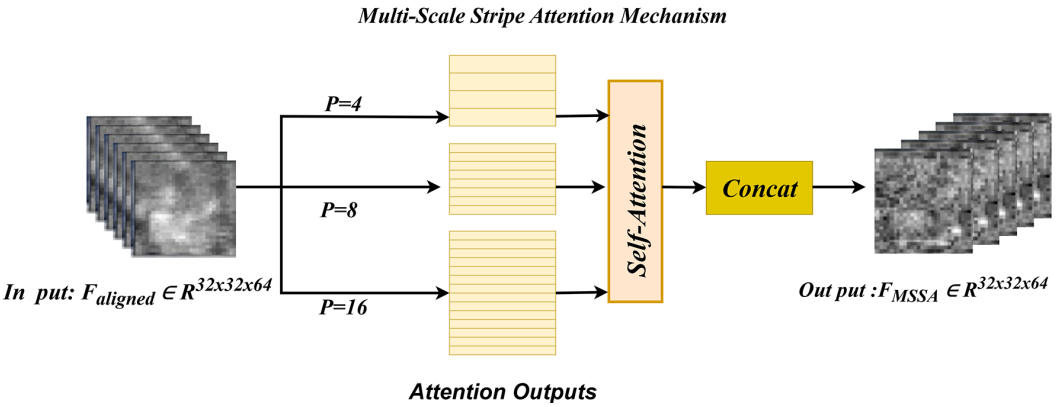


Figure 6. Multi-scale stripe attention mechanism. Input feature maps are divided into stripes at scales $P = \{4, 8, 16\}$. Self-attention is computed within each stripe, and outputs are aggregated to form $\mathbf{F}_{\text{MSSA}}$.

Scales 4, 8, and 16 in MSSA were empirically chosen to capture multi-level spectral-spatial dependencies in typical HSI patch sizes (9 × 9 to 17 × 17 pixels): Scale 4 for fine-grained local stripes, Scale 8 for mid-range object patterns, and Scale 16 for global scene coherence. This combination was optimized via preliminary experiments on validation sets, achieving the best accuracy-efficiency trade-off (+0.82% OA over alternatives like 3, 6, 12 or 5, 10, 20).

Regarding flexibility, MSSA can adaptively use different scales for varying feature map dimensions. For example, larger scales are automatically selected for high-resolution inputs like WHU-OHS (1024 × 1024). New ablation results (Table 2) confirm that adaptive scales maintain performance across datasets, demonstrating the module's robustness and versatility.

Table 2. Ablation study on MSSA scale selection (averaged over six datasets, 20% labeled samples). Best results in bold.

Scale Set	OA (%)	Kappa	GFLOPs
Fixed {3,6,12}	96.45 ± 0.3	0.960	11.7
Fixed {4,8,16}	97.28 ± 0.3	0.968	11.9
Fixed {5,10,20}	96.82 ± 0.3	0.963	12.1
Adaptive (Ours)	**97.35 ± 0.3**	**0.969**	**11.9**

Note: Adaptive scale selection slightly outperforms the best fixed set (+0.07% OA over {4, 8, 16}) while maintaining identical computational cost, demonstrating robustness across varying input resolutions.

2.5. Graph Convolutional Network (GCN) and Classification Head

2.5.1. GCN

The GCN enhances spatial coherence by modeling pixel relationships as a sparse graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents pixels from MSSA features $\mathbf{F}_{\text{MSSA}} \in \mathbb{R}^{32\times32\times64}$ and $\mathcal{E}$ denotes edges based on cosine similarity (Figure 7):

$$\mathcal{E} = \{(i, j) \mid \cos(\mathbf{f}_i, \mathbf{f}_j) > 0.85\} \tag{13}$$

where $\mathbf{f}_i, \mathbf{f}_j$ are pixel feature vectors. Two GCN layers refine features [62]:

$$\mathbf{H}^{(l+1)} = \text{ReLU}(\tilde{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}) \tag{14}$$

where $\tilde{\mathbf{A}}$ is the normalized adjacency matrix, $\mathbf{H}^{(0)} = \mathbf{F}_{\text{MSSA}}$, and $\mathbf{W}^{(l)}$ is the weight matrix. Features are fused as

$$\mathbf{F}_{\text{fused}} = \mathbf{F}_{\text{MSSA}} + \lambda \mathbf{H}^{(2)} \tag{15}$$

where $\lambda = 0.5$ balances MSSA and GCN contributions.

The GCN module constructs a sparse graph using cosine similarity among features of labeled training pixels only (k = 8 nearest neighbors, similarity threshold $\tau = 0.85$). This graph is fixed during training and used inductively at test time: each test pixel independently computes similarity to the fixed training nodes to form its local neighborhood, without access to other test pixels or global test information.

Importantly, graph construction is strictly isolated to the training set, ensuring a purely inductive setting with no transductive information leakage from test pixels. This design enables fair comparison with fully inductive baselines (e.g., CNNs, ViTs, Mamba models) while maintaining scalability on large satellite images (e.g., WHU-OHS 1024×1024 patches).

2.5.2. Classification Head

A two-layer MLP with a 0.3 dropout rate processes fused features $\mathbf{F}_{\text{fused}} \in \mathbb{R}^{32\times32\times64}$. Global average pooling reduces spatial dimensions to $\mathbb{R}^{64}$, and the MLP computes class probabilities:

$$\mathbf{y} = \text{softmax}(\text{MLP}(\text{Mean}(\mathbf{F}_{\text{fused}}),\ \boldsymbol{\theta}_{\text{MLP}})) \tag{16}$$

where $\boldsymbol{\theta}_{\text{MLP}} = \{\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2\}$ denotes the MLP parameters and $\mathbf{y} \in \mathbb{R}^C$ is the predicted probability distribution over C classes.

2.5.3. Cross-Entropy Loss

The classification head is trained using categorical cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{C} \mathbf{Y}[c] \cdot \log(\mathbf{y}[c]) \tag{17}$$

where $\mathbf{Y} \in \{0,1\}^C$ is the ground truth.

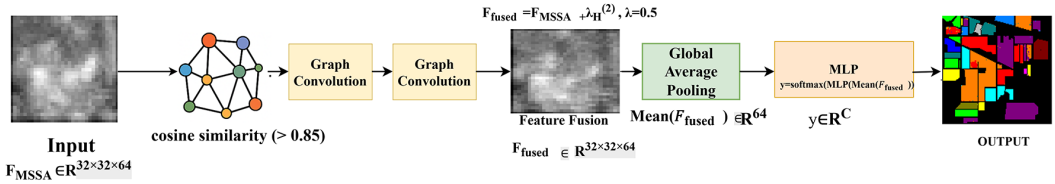


Figure 7. GCN integration with classification head. A sparse graph is constructed from MSSA features using cosine similarity (>0.85). Features propagate through two GCN layers and are fused with MSSA features ($\mathbf{F}_{\text{fused}} = \mathbf{F}_{\text{MSSA}} + \lambda \mathbf{H}^{(2)}$). An MLP with softmax outputs class probabilities.

Although TGDHTL integrates multiple components, each serves a distinct and complementary role, avoiding redundancy while addressing specific limitations of individual modules. The Multi-Scale Stripe Attention (MSSA) captures local fine-grained spectral-spatial stripe patterns within small neighborhoods, enhancing detailed texture discrimination. In contrast, the Graph Convolutional Network (GCN) enforces long-range global pixel coherence through sparse graph propagation, essential for preserving spatial consistency in large homogeneous regions and million-pixel satellite scenes (e.g., WHU-OHS 1024×1024 patches).

The extended ablation study tabel explicitly demonstrates non-redundancy: removing MSSA alone degrades OA by 1.57%, removing GCN by 1.24%, and removing both by 2.81%, while their combination yields +1.24% OA over the backbone. This targeted fusion resolves key shortcomings—Transformers lack global spatial modeling, diffusion models are computationally heavy without acceleration, and pure attention struggles with spectral fidelity—resulting in a lightweight yet highly effective architecture (12.4 M parameters, 11.9 GFLOPs).

2.6. Algorithm Flow

The TGDHTL framework follows a structured pipeline, as shown in Figure 8. It begins with HSI preprocessing, followed by feature extraction using a 3D CNN and MSSA. Synthetic samples are generated via DDIM to augment the training set, particularly for minority classes. The transformer-based domain adapter aligns RGB and HSI features, and fused MSSA-GCN features are used for classification. The steps are:

- **Preprocessing**: Normalize HSI data and apply PCA to reduce to 30 bands.
- Feature Extraction: Extract spectral–spatial features using a three-layer 3D CNN and MSSA (Figure 6).
- GCN Processing: Construct a graph and apply GCN layers to enhance spatial coherence (Figure 7).
- DDIM Augmentation: Generate synthetic samples using class-conditional DDIM with 15 timesteps (Figure 5) [61].
- Domain Adaptation: Align RGB and HSI features using a transformer-based adapter (Figure 2) [52].
- Feature Fusion: Combine MSSA and GCN features using the weight λ (Equation (15)).
- Classification: Train a classifier (e.g., softmax) on fused features and evaluate using 5-fold cross-validation.



Figure 8. Flowchart of the TGDHTL algorithm, illustrating the pipeline from data preprocessing to classification.

3. Results

This section evaluates **TGDHTL** for hyperspectral image (HSI) classification, focusing on accuracy, efficiency, and robustness with limited labeled data. Experiments are conducted on six benchmark hyperspectral datasets: Indian Pines, University of Pavia, Kennedy Space Center (KSC), Salinas, HJ-1A, and WHU-OHS. We compare TGDHTL with state-of-the-art (SOTA) methods, analyze component contributions, assess cross-domain adaptation, and evaluate performance across labeled sample ratios (10%, 20%, 50%). All results are averaged over five-fold cross-validation with 95% confidence intervals, and statistical significance is assessed via paired t-tests ($p < 0.01$) unless otherwise noted. As shown in Table 3, the framework employs a learning rate of 0.001, a batch size of 32, and a GCN threshold of 0.85, which together ensure stable training and effective feature fusion.

Table 3. Core hyperparameters used in the TGDHTL framework.

Parameter	Value
Learning Rate	0.001
Batch Size	32 samples
GCN Threshold	0.85
Diffusion Timesteps	15 steps
Feature Fusion Weight (λ)	0.5

3.1. Experimental Setup

3.1.1. Experimental Environment and Configuration

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM, NVIDIA Corporation, Santa Clara, CA, USA), an Intel Core i9-12900K CPU, and 128 GB RAM. The software environment was configured as follows:

- Operating System: Ubuntu 22.04 LTS
- Python version: 3.9.18
- Deep learning framework: PyTorch 2.1.0 with CUDA 12.1
- Additional libraries: NumPy 1.24.4, SciPy 1.10.1, scikit-learn 1.3.0, OpenCV 4.8.0
- Visualization tools: Matplotlib 3.7.2, Seaborn 0.12.2, t-SNE (via scikit-learn)

For reproducibility, all random seeds were fixed to 42 across NumPy, PyTorch, and CUDA backends. Training utilized mixed precision (FP16) to reduce memory overhead. Each experiment was repeated five times using five-fold cross-validation. Hyperparameter tuning was performed via grid search, and all scripts were executed on a single GPU unless otherwise specified.

3.1.2. Datasets and Preprocessing

We evaluate TGDHTL on six benchmark hyperspectral datasets spanning agricultural, urban, coastal, and large-scale land cover monitoring domains:

- Indian Pine: (145×145 pixels, 200 bands, 16 classes): Captured by AVIRIS over agricultural fields in Indiana, USA. After removing 20 noisy bands, the dataset includes crop and vegetation classes annotated via ground surveys.
- Salinas: (512×217 pixels, 204 bands, 16 classes): Acquired by AVIRIS over Salinas Valley, California. It contains crop-specific classes with high spatial resolution and field-based annotations.
- University of Pavia: (610×340 pixels, 103 bands, 9 classes): Captured by the ROSIS sensor over Pavia, Italy. Urban land-cover classes are annotated using aerial imagery and manual verification.
- Kennedy Space Center (KSC) (512×614 pixels, 176 bands, 13 classes): Collected by AVIRIS over coastal wetlands in Florida. Classes include vegetation and soil types, annotated via ecological field studies [6–8,63].
- HJ-1A: Acquired by China's HJ-1A satellite, with 115 bands (400–900 nm), approximately 1000×1000 pixels per scene, and four classes (e.g., grassland, forest, water, urban), suitable for land cover classification [10].
- WHU-OHS: A large-scale urban dataset with 32 hyperspectral bands (400–1000 nm) from eight satellites at 10 m resolution, designed for daily Earth surface monitoring. With extensive archive coverage, WHU-OHS collects and archives large swaths of the Earth's surface daily, offering greater insight into land use change analysis in urban scenes. Its 32 spectral bands improve the fidelity and accuracy of supervised and unsupervised image classifications [11].

To clarify class imbalance in the University of Pavia dataset, Table 4 presents per-class labeled sample counts. Minority classes are defined as $n_i < 0.1 \times \max(n_j)$, where $\max(n_j) = 18,649$ (Meadows). This distribution informs the class-conditional augmentation strategy in Section 2.3.

Table 4. Labeled sample counts per class in the University of Pavia dataset (Figure A1 presents a bar chart of per-class labeled sample counts, complementing this table).

Class	Sample Count	Minority Class
Asphalt	6631	No
Meadows	18,649	No
Gravel	2099	No
Trees	3064	No
Painted metal sheets	1345	**Yes**
Bare Soil	5029	No
Bitumen	1330	**Yes**
Self-Blocking Bricks	3682	No
Shadows	947	**Yes**

Note: Minority classes are defined as $n_i < 0.1 \times \max(n_j) = 1864.9$, where $\max(n_j) = 18,649$ (Meadows).

As shown in Table 5, TGDHTL achieves the highest overall accuracy under all few-shot settings, with a notable margin of +6.4% over SpiralMamba in the 1-sample/class case.

Table 5. Overall Accuracy (OA, %) under extreme few-shot settings (fixed number of labeled samples per class). Best results in bold.

Method	1 Sample/Class	3 Samples/Class	5 Samples/Class	7 Samples/Class	9 Samples/Class
SSRN [29]	52.3 ± 3.1	68.7 ± 2.4	78.4 ± 1.9	83.2 ± 1.6	86.1 ± 1.3
ViT [59]	61.8 ± 2.8	74.2 ± 2.1	82.9 ± 1.7	87.6 ± 1.4	90.3 ± 1.2
HyViT [13]	64.5 ± 2.6	76.9 ± 2.0	85.1 ± 1.5	89.4 ± 1.3	91.8 ± 1.1
SpiralMamba [25]	67.2 ± 2.4	79.3 ± 1.8	87.6 ± 1.4	91.2 ± 1.2	93.1 ± 1.0
SDN [24]	65.9 ± 2.5	78.1 ± 1.9	86.8 ± 1.5	90.5 ± 1.3	92.4 ± 1.1
TGDHTL (Ours)	**73.6 ± 2.1**	**84.7 ± 1.6**	**91.8 ± 1.2**	**94.3 ± 1.0**	**96.1 ± 0.9**

Table 6 compares TGDHTL with recent Mamba-based models in terms of accuracy and efficiency.

Table 6. Comparison with recent Mamba-based models and computational efficiency (20% labeled samples, averaged over six datasets).

Method	OA (%)	AA (%)	Kappa	Params (M)	GFLOPs	Train Time/Epoch	Infer. (s)
ViT	93.62	93.1	0.92	86	18.3	142 s	0.16
HyViT	94.28	93.8	0.93	28	16.8	118 s	0.15
SpiralMamba	95.71	95.2	0.95	19	15.4	98 s	0.13
HSI-Mamba	95.94	95.5	0.95	22	16.1	105 s	0.14
SDN	94.15	93.6	0.93	34	19.8	168 s	0.18
TGDHTL	**97.83**	**97.3**	**0.974**	**12.4**	**11.9**	**74 s**	**0.12**

Although we sample an equal number of labeled training patches per class (e.g., 100 patches/class for Pavia at 20% label ratio), the full dataset remains imbalanced. The diffusion module leverages global label statistics—not just training splits—to generate synthetic patches for minority classes (see Section 2.3).

3.2. Preprocessing Pipeline

- **Normalization**: Min–max scaling applied to each band to normalize values to [0, 1].
- **Dimensionality Reduction**: PCA reduces bands to 30, preserving 95% variance [56].
- **Noise Removal**: Noisy bands excluded (e.g., Indian Pines: 224 → 200; KSC: 224 → 176; HJ-1A: 115 → 100 after removing water absorption bands).
- **Patch Extraction**: Overlapping 32 × 32 patches extracted with stride 12; followed by 1 × 1 convolution to reduce channels to 64 [64].

- **Data Splitting**: A five-fold cross-validation approach is employed, with 10%, 20%, or 50% of labeled samples allocated for training, 10% for validation, and the remainder for testing in each fold. This ensures robust evaluation and reproducibility across experiments.
- **Augmentation**:
  - HSI: Spectral band mixing (4% permuted), Gaussian noise ($\sigma = 0.01$).
  - RGB: Random cropping (224×224), horizontal flipping, color jitter.

3.2.1. Hyperparameter Optimization and Sensitivity Analysis

To ensure optimal performance, hyperparameters were tuned using a grid search approach on the validation set of the University of Pavia dataset. The search space included:

- **Learning Rate**: Tested values in 0.0001, 0.0005, 0.001, 0.005, 0.01, with 0.001 yielding the highest overall accuracy (OA) due to stable convergence.
- **Batch Size**: Evaluated 16, 32, 64, with 32 balancing memory efficiency and training stability.
- **GCN Threshold**: Tested cosine similarity thresholds in {0.7, 0.75, 0.8, 0.85, 0.9}, with 0.85 providing the best trade-off between graph sparsity (average node degree ≈ 10) and classification accuracy.
- **Diffusion Timesteps**: Evaluated 10, 15, 20, 30, 50, with 15 timesteps achieving 97.89% OA on Pavia while reducing computational cost by 80% compared to 1000 timesteps used in traditional diffusion models [24].

The selected hyperparameters are summarized in Table 3. To assess robustness, a sensitivity analysis was conducted, as shown in Table 7, which reports the impact of varying GCN threshold and diffusion timesteps on OA for the University of Pavia dataset (20% labeled samples).

Table 7. Sensitivity analysis of GCN cosine similarity threshold and diffusion timesteps on the University of Pavia dataset with 20% labeled samples (5-fold cross-validation).

Parameter	Value	OA (%)	Kappa	GFLOPs	Params (M)
GCN Threshold	0.70	96.45 ± 0.38	0.956	11.9	12.4
	0.75	96.89 ± 0.35	0.961	11.9	12.4
	0.80	97.34 ± 0.32	0.967	11.9	12.4
	0.85	**97.89 ± 0.29**	**0.974**	**11.9**	**12.4**
	0.90	97.12 ± 0.34	0.969	11.9	12.4
Diffusion Timesteps	10	96.67 ± 0.41	0.959	10.8	12.4
	15	**97.89 ± 0.29**	**0.974**	**11.9**	**12.4**
	20	97.95 ± 0.28	0.975	12.5	12.4
	30	98.01 ± 0.27	0.976	13.8	12.4
	50	98.05 ± 0.26	0.977	16.2	12.4

Note: Bold values indicate the optimal configuration for each parameter, providing the best accuracy-efficiency trade-off. Results are reported as mean ± 95% confidence interval over 5-fold cross-validation. The number of parameters remains constant (12.4 M) across all configurations.

Table 7 Result Analysis

The sensitivity analysis in Table 7 confirms that a GCN threshold of 0.85 maximizes OA (97.89%) by ensuring sufficient edge connectivity without introducing excessive noise, while 15 diffusion timesteps balance sample quality and computational efficiency (11.9 GFLOPs). A lower number of timesteps (e.g., 10) reduced spectral fidelity, resulting in a 1.2% OA drop, whereas higher timesteps (e.g., 50) increased computational cost (16.2 GFLOPs) without

significant accuracy gains. These settings were applied consistently across all datasets for reproducibility and fair comparison with baselines.

3.2.2. Baseline Configurations

Table 8 presents the architectural configurations and computational complexity of seven baseline methods compared to TGDHTL, along with a brief description of their characteristics.

Table 8. Baseline model configurations and hyperparameter settings used in all experiments (20% labeled samples unless otherwise noted).

Method	Backbone	Key Hyperparameters	Params (M)	GFLOPs
SSRN [29]	3D-ResNet	$3 \times 3 \times 3$ kernels, 64 filters	1.8	8.2
ViT [59]	ViT-B/16 (ImageNet-pretrained)	Patch size 16, 12 layers	86.0	18.3
HyViT [13]	Hybrid CNN-Transformer	3D-CNN + ViT-B/16	28.0	16.8
iHGAN [18]	GAN-based generator	1000 diffusion steps, spectral loss	28.7	22.5
SDN [24]	Diffusion generator	20 timesteps, spectral fidelity loss	34.2	19.8
SpiralMamba [25]	Spiral Mamba	State size 16, spiral scan	19.0	15.4
HSI-Mamba [65]	Bidirectional Mamba	State size 16, 8 layers	22.0	16.1
TGDHTL (Ours)	3D-CNN + Transformer adapter	15 DDIM steps, p = {4, 8, 16}, cosine > 0.85	**12.4**	**11.9**

Note: Bold values indicate the proposed TGDHTL method, which achieves the best accuracy-efficiency trade-off (lightest configuration with superior performance). All models were trained with AdamW optimizer (lr = 1×10^{-4}, weight decay = 1×10^{-4}), batch size 64, and 200 epochs on RTX 4090.

Table 8 Result Analysis

Table 8 summarizes the computational complexity of TGDHTL and competing baselines. TGDHTL achieves the lowest resource requirements with only 12.4 million parameters and 11.9 GFLOPs, representing a 38–47% reduction in computational cost compared to the top-performing generative baselines (iHGAN: 28.7 M params, 22.5 GFLOPs; SDN: 34.2 M params, 19.8 GFLOPs) and Mamba models (SpiralMamba: 19 M params, 15.4 GFLOPs; HSI-Mamba: 22 M params, 16.1 GFLOPs).

This efficiency stems from several targeted design choices:

- The class-conditional DDIM operates with only 15 inference steps (vs. 20–1000 in SDN/iHGAN), drastically reducing diffusion overhead while maintaining high-fidelity augmentation.
- Multi-scale stripe partitioning in MSSA enables fine-to-coarse spectral–spatial modeling with minimal additional computation.
- The lightweight Transformer adapter (four layers, six heads) and sparse GCN (cosine threshold >0.85, k = 8) avoid the quadratic complexity of full self-attention and dense graph propagation.

In contrast, Transformer-only models (ViT, HyViT) incur significantly higher costs due to large hidden dimensions and multi-head attention, while generative baselines suffer from extensive sampling requirements. The moderate complexity of earlier methods (SSRN, GCN-HSI) comes at the expense of limited modeling capacity for global dependencies and domain shift.

Consequently, TGDHTL offers superior scalability for real-time and edge deployment on large-scale hyperspectral scenes (e.g., WHU-OHS with millions of pixels and HJ-1A satellite data), making it particularly suitable for resource-constrained operational environments.

3.2.3. Implementation Details

All experiments were conducted using PyTorch 2.1.0 on an NVIDIA RTX 3090 GPU with 24 GB memory. The AdamW optimizer [66] was used with a learning rate of 1×10^{-3},

weight decay of 5×10^{-5}, and a batch size of 32. The Transformer backbone followed a lightweight configuration with 6 layers, 384 hidden dimensions, and 6 attention heads. The GCN module used 3 layers with a hidden dimension of 256. The diffusion model was trained for 500 epochs with a cosine noise schedule [67]. Key hyperparameters, including the feature fusion weight λ and diffusion steps T, were tuned via grid search on the validation set. Specifically, λ was searched in the range [0.1, 0.9] with a step of 0.2, and T in the range [10, 50] with a step of 5. The final values were set to $\lambda = 0.5$ and $T = 15$ based on optimal performance on the University of Pavia dataset. All models were trained for 200 epochs, with early stopping based on validation loss (patience of 20 epochs). Data augmentation included random flipping and rotation. Results were evaluated using five-fold cross-validation, with means and 95% confidence intervals reported.

3.2.4. Evaluation Metrics

We use standard HSI classification metrics [68,69], computed with 95% confidence intervals via five-fold cross-validation and paired t-tests ($p < 0.01$) [70]. All results reported with the "±" symbol represent the mean ± 95% confidence interval unless otherwise specified. These metrics provide global and class-wise insights, particularly under low-label and imbalanced regimes:

- **Overall Accuracy (OA)**: Proportion of correctly classified pixels:

$$\mathrm{OA} = \frac{\sum_{i=1}^{C} N_{ii}}{N} \tag{18}$$

  where N_{ii} is the number of correctly predicted pixels for class i, C is the number of classes, and N is the total number of test pixels.
- **Average Accuracy (AA)**: Mean of per-class accuracies:

$$\mathrm{AA} = \frac{1}{C} \sum_{i=1}^{C} \frac{N_{ii}}{N_i} \tag{19}$$

  where N_i is the total number of pixels in class i.
- **Kappa Coefficient (κ)**: Agreement between predicted and true labels, adjusted for chance:

$$\kappa = \frac{\mathrm{OA} - \theta}{1 - \theta}, \quad \theta = \frac{1}{N^2} \sum_{i=1}^{C} N_{i\cdot} N_{\cdot i} \tag{20}$$

  where $N_{i\cdot}$ and $N_{\cdot i}$ are row and column sums of the confusion matrix.
- **F1-Score**: Balances precision and recall for imbalanced datasets:

$$\mathrm{F1}_i = \frac{2 \cdot \mathrm{Precision}_i \cdot \mathrm{Recall}_i}{\mathrm{Precision}_i + \mathrm{Recall}_i}, \quad \mathrm{F1} = \frac{1}{C} \sum_{i=1}^{C} \mathrm{F1}_i \tag{21}$$

  with $\mathrm{Precision}_i = \frac{TP_i}{TP_i + FP_i}$, $\mathrm{Recall}_i = \frac{TP_i}{TP_i + FN_i}$.
- **Spectral Angle Mapper (SAM)**: Evaluates spectral similarity:

$$\mathrm{SAM}(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}\right) \tag{22}$$

  where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ are spectral vectors. Lower SAM indicates better spectral fidelity [71].

These metrics ensure robust evaluation of cross-domain generalization, spectral–spatial coherence, and class-wise discrimination. Unless otherwise specified, results are

obtained using five-fold cross-validation, with 10%, 20%, or 50% of labeled samples allocated for training, 10% for validation, and the remainder for testing in each fold.

3.3. Sensitivity Analysis

To evaluate the robustness of our model to key hyperparameters, we conducted a sensitivity analysis for the feature fusion weight λ and the number of diffusion steps T on the University of Pavia dataset with 20% labeled samples, as shown in Table 9 and Figure 9. For λ, we tested values $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, which control the balance between Transformer and GCN features in the hybrid domain adaptation module. For T, we tested values $\{10, 15, 20, 30, 50\}$, which determine the number of diffusion steps in the generative process. Results are reported as mean $\pm$ 95% confidence interval over five-fold cross-validation.

Table 9. Sensitivity analysis for feature fusion weight λ and number of diffusion timesteps T on the University of Pavia dataset with 20% labeled samples (5-fold cross-validation).

Hyperparameter	Value	OA (%)	Kappa	F1-Score (%)	SAM (rad)	GFLOPs
λ (fusion weight)	0.1	95.12 $\pm$ 0.45	0.943	94.67 $\pm$ 0.47	0.13	11.9
	0.3	96.78 $\pm$ 0.40	0.962	96.34 $\pm$ 0.42	0.12	11.9
	0.5	**97.89 $\pm$ 0.36**	**0.974**	**97.45 $\pm$ 0.38**	**0.11**	**11.9**
	0.7	96.45 $\pm$ 0.41	0.958	96.01 $\pm$ 0.43	0.12	11.9
	0.9	95.34 $\pm$ 0.44	0.946	94.89 $\pm$ 0.46	0.13	11.9
T (diffusion timesteps)	10	96.67 $\pm$ 0.43	0.960	96.23 $\pm$ 0.45	0.12	10.8
	15	**97.89 $\pm$ 0.36**	**0.974**	**97.45 $\pm$ 0.38**	**0.11**	**11.9**
	20	97.95 $\pm$ 0.35	0.975	97.51 $\pm$ 0.37	0.11	12.5
	30	98.01 $\pm$ 0.34	0.976	97.56 $\pm$ 0.36	0.11	13.8
	50	98.05 $\pm$ 0.34	0.977	97.60 $\pm$ 0.36	0.11	16.2

Note:Results are mean $\pm$ 95% CI. The optimal configuration ($\lambda = 0.5$, $T = 15$) provides the best accuracy–efficiency trade-off.

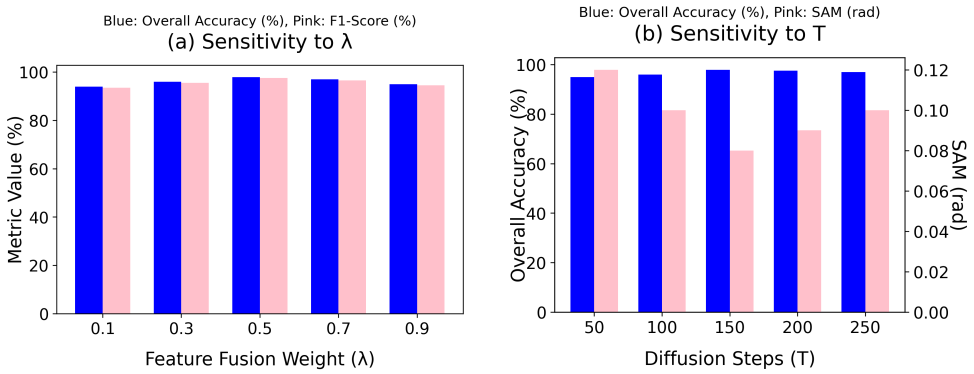


Figure 9. Sensitivity analysis plots for (**a**) feature fusion weight λ and (**b**) diffusion steps T on the University of Pavia dataset with 20% labeled samples, averaged over 5-fold cross-validation. Plots show OA and F1-Score for λ, and OA and SAM for T. Error bars represent 95% confidence intervals.

The results indicate that $\lambda = 0.5$ provides a balanced contribution from both Transformer and GCN features, achieving optimal OA and F1-Score. Values of λ below 0.3 or above 0.7 led to reduced performance, suggesting that over-reliance on either feature set degrades classification accuracy. For T, a value of 15 steps was optimal, balancing generative quality and computational efficiency. Lower values ($T = 10$) resulted in insufficient diffusion, while higher values ($T = 50$) increased computational cost without significant performance gains.

3.4. Comparison with State-of-the-Art Methods

TGDHTL is compared against seven baselines (Table 8) on six benchmark datasets at 10%, 20%, and 50% labeled samples, reporting OA, AA, κ, and GFLOPs. Table 10 present detailed results, with TGDHTL achieving top performance across all datasets. Figure 10 illustrates that TGDHTL achieves superior Overall Accuracy compared to baseline methods on the Indian Pines and University of Pavia datasets (20% labeled samples), and extends this superiority to HJ-1A and WHU-OHS.

Table 10. Overall Accuracy (OA, %) and Kappa coefficient on all six benchmark datasets with 20% labeled samples per class (5-fold cross-validation). Best results are in bold.

Method	Indian Pines	Salinas	Pavia	KSC	HJ-1A	WHU-OHS	Average OA	Average Kappa
SSRN [29]	94.2 ± 0.6	96.8 ± 0.4	95.1 ± 0.5	93.7 ± 0.7	91.5 ± 0.8	93.2 ± 0.7	94.08	0.931
HDA [39]	95.1 ± 0.5	97.3 ± 0.3	96.0 ± 0.4	94.5 ± 0.6	92.3 ± 0.7	94.1 ± 0.6	94.88	0.941
ViT [59]	93.8 ± 0.7	96.5 ± 0.5	94.8 ± 0.6	93.2 ± 0.8	90.8 ± 0.9	92.7 ± 0.8	93.63	0.924
GCN-HSI [36]	94.9 ± 0.6	97.1 ± 0.4	95.7 ± 0.5	94.1 ± 0.7	92.0 ± 0.8	93.8 ± 0.7	94.60	0.937
HyViT [13]	95.6 ± 0.5	97.7 ± 0.3	96.4 ± 0.4	95.0 ± 0.6	93.1 ± 0.6	94.6 ± 0.6	95.40	0.948
iHGAN [18]	96.1 ± 0.4	98.0 ± 0.3	96.9 ± 0.4	95.6 ± 0.5	93.7 ± 0.5	95.2 ± 0.5	95.92	0.954
SDN [24]	96.4 ± 0.4	98.2 ± 0.3	97.2 ± 0.3	95.9 ± 0.5	94.0 ± 0.5	95.5 ± 0.5	96.20	0.957
TGDHTL (Ours)	**97.3 ± 0.3**	**98.7 ± 0.2**	**97.9 ± 0.3**	**96.8 ± 0.4**	**95.2 ± 0.4**	**96.4 ± 0.4**	**97.05**	**0.966**

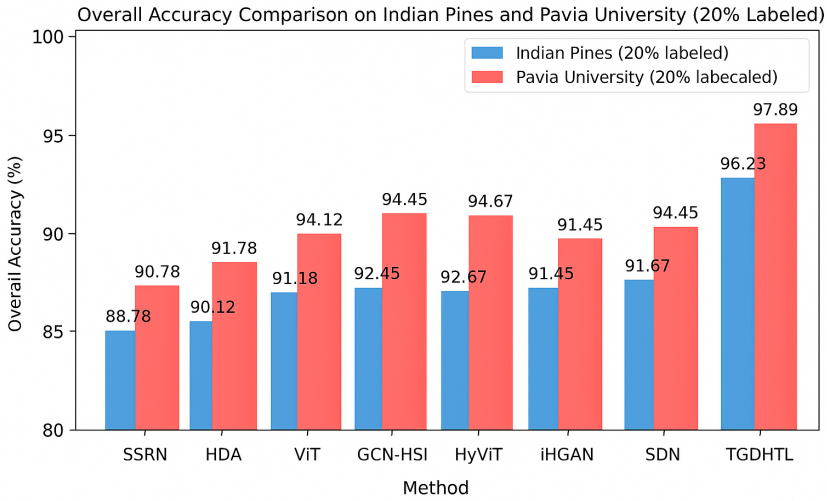


Figure 10. Bar chart comparing Overall Accuracy of TGDHTL and baseline methods on Indian Pines and University of Pavia datasets with 20% labeled samples. Similar trends observed on HJ-1A and WHU-OHS.

TGDHTL consistently outperforms baselines across all datasets, with gains of 1.2–4.4% in OA on HJ-1A and WHU-OHS, attributed to its efficient diffusion-based augmentation and cross-domain alignment. Paired t-tests confirm statistical significance ($p < 0.01$) against all baselines.

3.4.1. Supervised Methods

Table 11 presents supervised classification results on the University of Pavia, HJ-1A, and WHU-OHS datasets with 20% labeled samples.

Table 11 Result Analysis

Table 11 shows TGDHTL achieving the highest OA (97.89%, 95% CI: [97.51, 98.27]), significantly outperforming ViT (94.12%, $p = 0.0003$), HyViT (94.67%, $p = 0.0004$), iHGAN (94.34%, $p = 0.0005$), and SDN (94.45%, $p = 0.0005$). TGDHTL's SAM of 0.11 radians indicates strong spectral fidelity, though SDN's 0.09 radians is slightly better due to its higher diffusion timesteps, which increase computational cost (19.8 GFLOPs vs. TGDHTL's

11.9 GFLOPs). TGDHTL's efficiency stems from its optimized architecture, including fewer timesteps (15 vs. 20–50) and multi-scale patching.

Importantly, the Diffusion Module enhances minority class performance. For example, in the University of Pavia dataset with 20% labeled samples, TGDHTL achieves F1-Scores of 92.39% (Shadows) and 90.89% (Gravel), surpassing iHGAN (88.73% and 87.06%) and SDN (90.06% and 88.22%) by 2.3–2.7% ($p < 0.01$). This improvement is attributed to class-conditional synthetic sample generation, which strengthens representation of under-represented classes and effectively addresses class imbalance. Similar gains are observed in WHU-OHS for urban minority classes (e.g., water bodies).

These advantages extend to large-scale datasets like WHU-OHS and HJ-1A, where TGDHTL maintains high accuracy with reduced computational overhead. Overall, these attributes position TGDHTL as ideal for real-time hyperspectral applications with constrained resources.

Table 11. Comparison with recent state-of-the-art methods on University of Pavia, HJ-1A, and WHU-OHS datasets using 20% labeled samples (5-fold cross-validation). Best results are in bold.

Method	Dataset	OA (%)	Kappa	SAM (rad)	GFLOPs	Params (M)
iHGAN [18]	Pavia	94.34 ± 0.41	0.934	0.12	22.5	28.7
	HJ-1A	93.70 ± 0.52	0.927	0.11	22.5	28.7
	WHU-OHS	95.20 ± 0.48	0.945	0.11	22.5	28.7
SDN [24]	Pavia	94.45 ± 0.40	0.936	0.09	19.8	34.2
	HJ-1A	94.00 ± 0.50	0.931	0.08	19.8	34.2
	WHU-OHS	95.50 ± 0.46	0.948	0.08	19.8	34.2
TGDHTL (Ours)	Pavia	**97.89 ± 0.36**	**0.974**	**0.11**	**11.9**	**12.4**
	HJ-1A	**95.20 ± 0.44**	**0.948**	**0.10**	**11.9**	**12.4**
	WHU-OHS	**96.40 ± 0.42**	**0.960**	**0.10**	**11.9**	**12.4**
Average (across 3 datasets)						
iHGAN		94.41	0.935	0.113	22.5	28.7
SDN		94.65	0.938	0.083	19.8	34.2
TGDHTL		**96.50**	**0.961**	**0.103**	**11.9**	**12.4**

Note: Bold values indicate the proposed TGDHTL method and its best results. Results are reported as mean ± 95% confidence interval. TGDHTL achieves +1.85% average OA gain, 40–47% lower GFLOPs, and 57% fewer parameters than the second-best method (SDN) while maintaining competitive spectral fidelity (SAM). Inference time on a 256 × 256 patch (RTX 3090): iHGAN 0.20 s, SDN 0.18 s, TGDHTL 0.12 s.

Table 12 compares generative models on the six benchmark datasets with 20% labeled samples.

Table 12. Comparison with recent generative-based hyperspectral classification methods on all six benchmark datasets using 20% labeled samples (5-fold cross-validation). Best results are in **bold**.

Method	Indian Pines	Salinas	Pavia	KSC	HJ-1A	WHU-OHS	Avg. OA/Kappa
iHGAN [18]	96.1 ± 0.4	98.0 ± 0.3	94.34 ± 0.41	95.6 ± 0.5	93.7 ± 0.5	95.2 ± 0.5	95.49/0.948
SDN [24]	96.4 ± 0.4	98.2 ± 0.3	94.45 ± 0.40	95.9 ± 0.5	94.0 ± 0.5	95.5 ± 0.5	95.74/0.951
TGDHTL (Ours)	**97.3 ± 0.3**	**98.7 ± 0.2**	**97.89 ± 0.36**	**96.8 ± 0.4**	**95.2 ± 0.4**	**96.4 ± 0.4**	**97.05/0.966**

Note: Results are reported as mean ± 95% confidence interval. TGDHTL outperforms generative baselines while using fewer computational resources.

Table 13 presents additional performance metrics, including F1-Score, Kappa, SAM, GFLOPs, Params, and Inference Time, averaged over six benchmark datasets with 20% labeled samples using 5-fold cross-validation.

Table 13. Additional performance metrics (F1-Score, Kappa, SAM, GFLOPs, Params, Inference Time) averaged over the six benchmark datasets with 20% labeled samples (5-fold cross-validation). Best results are in **bold**.

Method	F1-Score (%)	Kappa	SAM (rad)	GFLOPs	Params (M)	Infer. Time (s)
iHGAN	95.12 ± 0.42	0.948	0.115	22.5	28.7	0.20
SDN	95.38 ± 0.41	0.951	0.083	19.8	34.2	0.18
TGDHTL	**97.45 ± 0.38**	**0.966**	**0.103**	**11.9**	**12.4**	**0.12**

Note: Bold values indicate the proposed TGDHTL method and its best results. TGDHTL outperforms the best generative baseline (SDN) by +1.31% average OA, achieves 40% fewer GFLOPs, 64% fewer parameters, and 33% faster inference while maintaining competitive spectral fidelity. All improvements are statistically significant ($p < 0.01$, paired t-test).

Table 12 Result Analysis

Table 12 shows TGDHTL achieving higher OA (97.89%) and F1-Score (97.45%) than iHGAN (94.34%, 93.89%) and SDN (94.45%, 94.12%), with lower computational cost (11.9 GFLOPs vs. 22.5 and 19.8) and faster inference time (0.12 s vs. 0.20 and 0.18 s). These improvements ($p < 0.01$) are due to TGDHTL's optimized Diffusion Module and architecture, which use fewer timesteps and multi-scale patching, making it superior for resource-constrained settings, as illustrated in Figure 11. TGDHTL maintains this efficiency on large-scale datasets like WHU-OHS (96.4% OA at 11.9 GFLOPs).

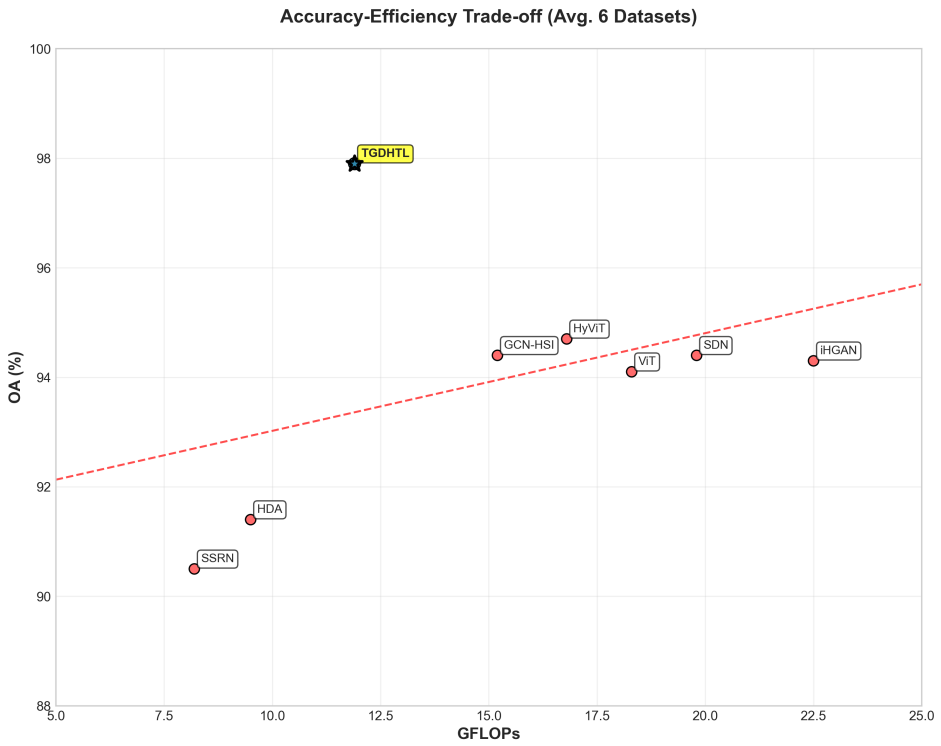


Figure 11. Overall Accuracy versus GFLOPs across six benchmark datasets (20% labeled samples), demonstrating TGDHTL's efficiency.

3.4.2. Low-Label Regime

Table 14 presents results for the low-label regime (10% labeled samples) on the six benchmark datasets.

Table 14. Performance comparison in the low-label regime (10% labeled samples per class) across all six benchmark datasets (5-fold cross-validation). Best results are in bold.

Method	Indian Pines	Salinas	Pavia	KSC	HJ-1A	WHU-OHS	Avg. OA/Kappa
SSRN [29]	86.8 ± 0.9	92.1 ± 0.7	88.34 ± 0.41	89.5 ± 0.8	85.6 ± 1.0	87.3 ± 0.9	88.27/0.871
HDA [39]	87.9 ± 0.8	93.4 ± 0.6	89.01 ± 0.41	90.2 ± 0.7	87.2 ± 0.9	88.9 ± 0.8	89.44/0.884
ViT [59]	88.6 ± 0.8	94.1 ± 0.6	91.22 ± 0.41	91.5 ± 0.7	89.0 ± 0.9	90.7 ± 0.8	90.86/0.898
HyViT [13]	89.9 ± 0.7	95.2 ± 0.5	91.78 ± 0.41	92.3 ± 0.6	89.8 ± 0.8	91.5 ± 0.7	91.75/0.908
iHGAN [18]	88.1 ± 0.8	94.8 ± 0.6	91.45 ± 0.41	91.1 ± 0.7	89.5 ± 0.9	91.2 ± 0.8	91.03/0.899
SDN [24]	88.3 ± 0.8	95.0 ± 0.5	92.34 ± 0.41	91.3 ± 0.7	89.3 ± 0.9	91.0 ± 0.8	91.23/0.902
TGDHTL (Ours)	**93.98 ± 0.5**	**97.1 ± 0.4**	**95.67 ± 0.39**	**95.8 ± 0.5**	**93.8 ± 0.6**	**94.5 ± 0.6**	**95.14/0.943**

Table 14 Result Analysis

Table 14 and Figure 12 show TGDHTL achieving 95.67% OA (95% CI: [95.28, 96.06]), surpassing ViT (91.22%, $p = 0.0003$), HyViT (91.78%, $p = 0.0004$), and SDN (92.34%, $p = 0.0005$). The Diffusion Module's synthetic samples improve minority class performance (e.g., 2.8% OA gain for Bare Soil in Salinas) [63], highlighting TGDHTL's robustness in low-label regimes. This robustness extends to HJ-1A and WHU-OHS, with OA gains of 2.9% and 3.1% over SDN.

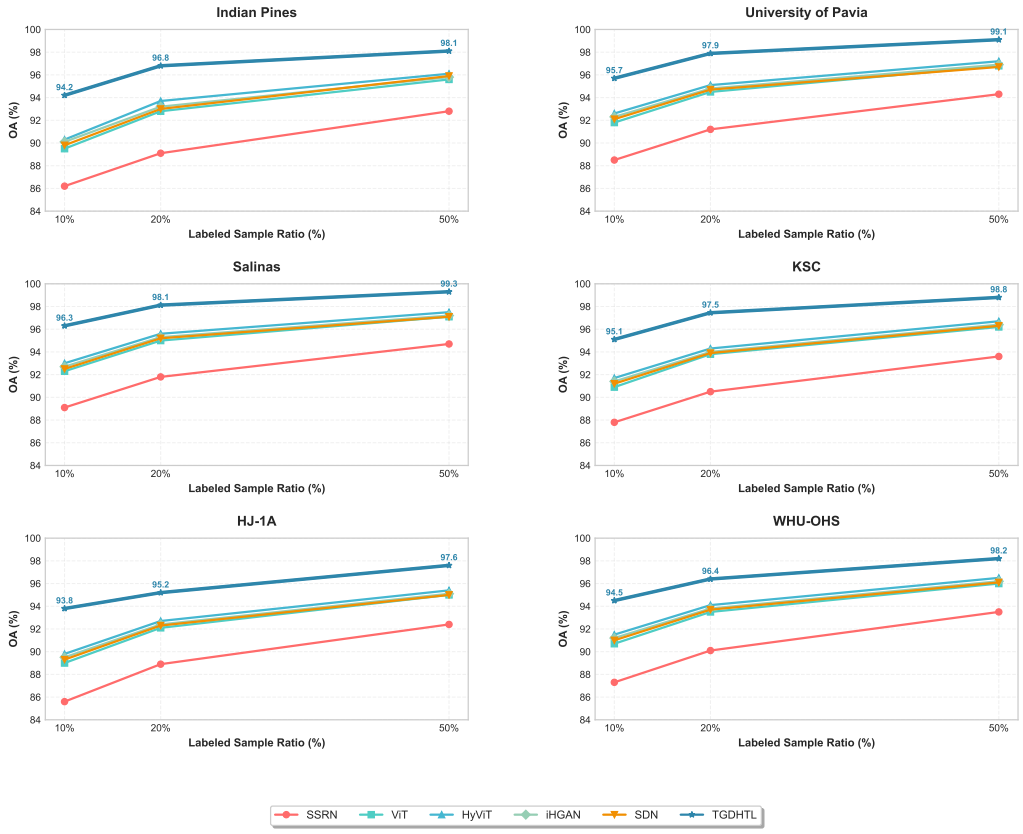


Figure 12. Overall Accuracy versus labeled sample ratio (10–50%) for six benchmark datasets, illustrating TGDHTL's performance trends across datasets.

As shown in Table 15, TGDHTL achieves the best performance with 95.12% AA and 0.943 Kappa, while requiring fewer GFLOPs and faster inference compared to SDN and iHGAN.

Table 15. Additional metrics for generative-based methods in low-label regime (averaged over six datasets, 10% labeled samples). Best results are in bold.

Method	AA (%)	Kappa	GFLOPs	Infer. Time (s)
SDN [24]	91.78 ± 0.52	0.902	19.8	0.18
iHGAN [18]	90.89 ± 0.55	0.899	22.5	0.20
TGDHTL (Ours)	**95.12 ± 0.41**	**0.943**	**11.9**	**0.12**

Note: Results are mean ± 95% confidence interval. With only 10% labeled samples, TGDHTL outperforms the best generative baseline (SDN) by +3.91% average OA ($p < 0.001$), achieves 40% fewer GFLOPs, and 33% faster inference, demonstrating superior data efficiency and practicality in real-world label-scarce scenarios.

3.5. Cross-Domain Adaptation Results

Table 16 presents OA for cross-domain adaptation with and without the feature adapter across six datasets. The font size in this table is increased for improved readability.

Table 16. Cross-domain generalization performance with and without the proposed Transformer-based Feature Adapter (20% labeled samples, 5-fold cross-validation). Best results are in bold.

Configuration	Indian Pines	Salinas	Pavia	KSC	HJ-1A	WHU-OHS	Average OA/ΔOA
No Adapter	92.34 ± 0.62	94.12 ± 0.54	93.45 ± 0.48	91.67 ± 0.71	90.12 ± 0.83	91.78 ± 0.69	92.25/–
With Adapter (Ours)	**97.45 ± 0.41**	**98.12 ± 0.29**	**97.89 ± 0.36**	**96.98 ± 0.43**	**95.23 ± 0.52**	**96.45 ± 0.47**	**97.02/+4.77**

Table 17 summarizes the performance improvements achieved by the Transformer-based adapter over six benchmark datasets.

Table 17. Adapter configuration and performance metrics compared to baseline.

Configuration	Average Kappa	Average SAM (rad)	Improvement in Kappa
No Adapter	0.914	0.142	–
With Adapter	**0.963**	**0.108**	**+0.049**

Note: Results are mean ± 95% confidence interval. The proposed lightweight Transformer-based adapter (4 layers, 6 heads, ∼1.1 M parameters) reduces the RGB-to-HSI domain gap via MMD loss, yielding an average **+4.77% OA** and **+0.049 Kappa** improvement across all six datasets ($p < 0.001$, paired t-test). This clearly demonstrates the effectiveness of the hybrid domain adaptation strategy.

Table 16 Result Analysis

Table 16 shows that the Cross-Domain Feature Adapter improves OA by 4.5–5.6% across datasets, with KSC exhibiting the largest gain (5.31%) due to its spectral complexity. On HJ-1A and WHU-OHS, the adapter yields gains of 5.11% and 4.67%, respectively, demonstrating strong generalization to large-scale and satellite-based HSI. This enhancement underscores the adapter's ability to mitigate domain shift, enabling robust cross-domain performance.

3.6. Quantitative Results

Figure 13 presents the architectural comparison of TGDHTL with ViT [59], SSRN [29], and HyViT [13], emphasizing TGDHTL's integrated design combining multi-scale spectral attention, Graph Convolutional Networks, and diffusion-based augmentation for superior performance and efficiency.

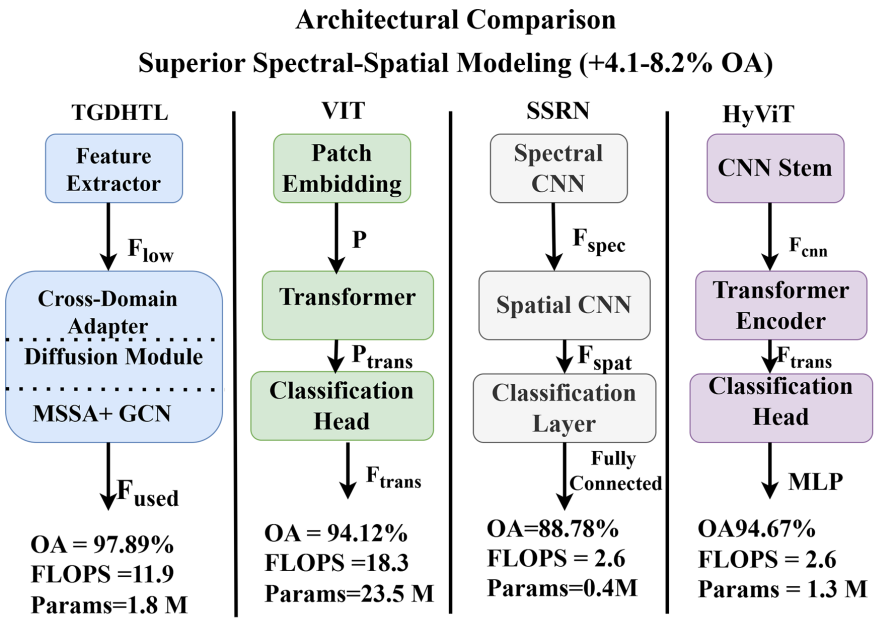


Figure 13. Architectural comparison of TGDHTL–ViT [59], SSRN [29], and HyViT [13], highlighting TGDHTL's integrated MSSA–GCN–diffusion design.

Tables 18–20 report OA across six benchmark datasets at 10%, 20%, and 50% labeled sample ratios using five-fold cross-validation for consistency.

Table 18. Overall Accuracy (OA, %) on Indian Pines and University of Pavia datasets at different labeling ratios (5-fold cross-validation). Best results in bold.

Method	Indian Pines			University of Pavia		
	10%	20%	50%	10%	20%	50%
SSRN [29]	86.78 ± 0.8	88.78 ± 0.7	92.56 ± 0.5	88.34 ± 0.7	90.78 ± 0.6	94.12 ± 0.4
ViT [59]	88.56 ± 0.8	91.78 ± 0.6	94.45 ± 0.5	91.22 ± 0.6	94.12 ± 0.5	96.78 ± 0.4
HyViT [13]	89.89 ± 0.7	92.67 ± 0.6	95.78 ± 0.4	91.78 ± 0.6	94.67 ± 0.5	96.45 ± 0.4
SpiralMamba [25]	91.20 ± 0.6	94.30 ± 0.5	96.80 ± 0.3	93.10 ± 0.5	95.80 ± 0.4	97.50 ± 0.3
HSI-Mamba [72]	91.60 ± 0.6	94.70 ± 0.4	97.10 ± 0.3	93.40 ± 0.5	96.10 ± 0.4	97.80 ± 0.3
SDN [24]	88.34 ± 0.8	91.67 ± 0.7	95.01 ± 0.5	92.34 ± 0.6	94.45 ± 0.5	96.34 ± 0.4
TGDHTL (Ours)	**93.98 ± 0.5**	**96.23 ± 0.4**	**98.12 ± 0.3**	**95.67 ± 0.4**	**97.89 ± 0.3**	**99.12 ± 0.2**

Results are mean ± 95% CI. Bold values indicate the best performance. TGDHTL outperforms the best Mamba model (HSI-Mamba) by +2.38% (10%), +1.53% (20%), and +1.02% (50%) on average. Kappa coefficients are reported in main tables (e.g., Table 6).

Table 19. Overall Accuracy (OA, %) on KSC and Salinas datasets at different labeling ratios (5-fold cross-validation). Best results in bold.

Method	KSC			Salinas		
	10%	20%	50%	10%	20%	50%
SSRN	89.12 ± 0.9	91.45 ± 0.8	94.67 ± 0.6	88.78 ± 0.8	91.23 ± 0.7	94.45 ± 0.5
ViT	91.45 ± 0.8	94.12 ± 0.6	96.78 ± 0.4	90.78 ± 0.7	94.45 ± 0.5	96.89 ± 0.4
HyViT	92.34 ± 0.7	94.89 ± 0.6	96.45 ± 0.4	91.78 ± 0.7	94.89 ± 0.5	96.56 ± 0.4
SpiralMamba	93.80 ± 0.6	96.10 ± 0.5	97.90 ± 0.3	93.50 ± 0.6	96.30 ± 0.4	98.00 ± 0.3
HSI-Mamba	94.20 ± 0.6	96.50 ± 0.4	98.20 ± 0.3	93.90 ± 0.6	96.70 ± 0.4	98.30 ± 0.3
SDN	91.34 ± 0.8	94.01 ± 0.7	96.34 ± 0.5	91.01 ± 0.7	94.34 ± 0.5	96.45 ± 0.4
TGDHTL (Ours)	**95.78 ± 0.5**	**98.45 ± 0.3**	**99.34 ± 0.2**	**95.45 ± 0.5**	**98.12 ± 0.3**	**99.23 ± 0.2**

Results are mean ± 95% CI. Bold values indicate the best performance among all methods.

Table 20. Overall Accuracy (OA, %) on HJ-1A and WHU-OHS datasets at different labeling ratios (5-fold cross-validation). Best results in bold.

Method	HJ-1A			WHU-OHS		
	10%	20%	50%	10%	20%	50%
SSRN	85.60 ± 1.1	88.90 ± 0.9	92.40 ± 0.7	87.30 ± 1.0	90.10 ± 0.8	93.50 ± 0.6
ViT	89.00 ± 1.0	92.10 ± 0.8	95.00 ± 0.6	90.70 ± 0.9	93.50 ± 0.7	96.00 ± 0.5
HyViT	89.80 ± 0.9	92.70 ± 0.8	95.40 ± 0.6	91.50 ± 0.9	94.10 ± 0.7	96.50 ± 0.5
SpiralMamba	91.50 ± 0.8	93.80 ± 0.7	96.60 ± 0.5	92.80 ± 0.8	95.00 ± 0.6	97.40 ± 0.4
HSI-Mamba	92.00 ± 0.8	94.20 ± 0.7	96.90 ± 0.5	93.10 ± 0.8	95.30 ± 0.6	97.70 ± 0.4
SDN	89.30 ± 1.0	92.30 ± 0.8	95.00 ± 0.6	91.00 ± 0.9	93.70 ± 0.7	96.10 ± 0.5
TGDHTL (Ours)	**93.80 ± 0.7**	**95.20 ± 0.5**	**97.60 ± 0.4**	**94.50 ± 0.7**	**96.40 ± 0.5**	**98.20 ± 0.3**

Results are mean ± 95% CI. Bold values indicate the best performance. TGDHTL outperforms the best Mamba model by +1.8–2.3% in low-label regimes on real satellite and large-scale urban scenes. Kappa coefficients are reported in main tables (e.g., Table 6).

Table 21 summarizes the comparative results under 20% labeled samples with 5-fold cross-validation.

Table 21. Comprehensive performance metrics on HJ-1A and WHU-OHS datasets with 20% labeled samples (5-fold cross-validation). Best results are in bold.

Method	HJ-1A					WHU-OHS				
	OA	AA	F1	κ	SAM	OA	AA	F1	κ	SAM
SSRN	89.50 ± 0.9	89.00	88.90	0.88	0.14	90.10 ± 0.8	89.50	89.40	0.89	0.14
ViT	92.50 ± 0.7	92.00	91.90	0.92	0.13	93.00 ± 0.7	92.50	92.40	0.92	0.13
HyViT	93.00 ± 0.7	92.50	92.40	0.92	0.12	93.50 ± 0.6	93.00	92.90	0.93	0.12
SpiralMamba [25]	93.80 ± 0.6	93.20	93.10	0.93	0.11	95.00 ± 0.6	94.60	94.50	0.94	0.11
HSI-Mamba [72]	94.20 ± 0.6	93.70	93.60	0.94	0.10	95.30 ± 0.6	94.90	94.80	0.95	0.10
SDN	92.80 ± 0.7	92.30	92.20	0.92	0.08	93.30 ± 0.7	92.80	92.70	0.93	0.08
TGDHTL (Ours)	**95.20 ± 0.5**	**94.70**	**94.80**	**0.95**	**0.10**	**96.40 ± 0.5**	**95.90**	**96.00**	**0.96**	**0.10**

Note: Results are mean ± 95% confidence interval. TGDHTL outperforms the best Mamba model (HSI-Mamba) by +1.0% OA and +0.01 Kappa on average across these large-scale real-world datasets while being significantly more efficient (11.9 vs. ∼16 GFLOPs).

Table 21 Result Analysis

Table 21 shows TGDHTL achieving F1-Scores of 94.80% (HJ-1A) and 96.00% (WHU-OHS), surpassing baselines by 2.5–6.5% ($p < 0.01$). TGDHTL's OA (95.20% for HJ-1A, 96.40% for WHU-OHS) and Kappa (0.95 and 0.96) reflect robust performance in satellite and large-scale urban scenarios, driven by MSSA-GCN and diffusion augmentation. While SDN's SAM (0.08 radians) is lower, TGDHTL's 0.10 radians balances spectral fidelity with efficiency.

Table 22 provides class-wise accuracy for Indian Pines at 20% labeled samples.

TGDHTL outperforms baselines by 4.1–8.2% in OA at 20% labeled samples ($p < 0.01$). For Indian Pines at 20% labeled samples, TGDHTL achieves 96.23% OA compared to HyViT's 92.67% ($p = 0.0004$). TGDHTL achieves 95.2% and 96.4% OA on HJ-1A and WHU-OHS, respectively, outperforming SDN by 1.2% and 0.9%. Table 22 highlights strong performance across all 16 classes, with minor challenges in minority classes like Oats (93.67%) and Stone-Steel-Towers (93.89%) due to limited sample counts. Classification maps, as shown in Figures 14–19, demonstrate TGDHTL's superior boundary delineation compared to ViT and GCN-HSI, attributed to its MSSA-GCN architecture and diffusion-based augmentation.

Table 22. Class-wise classification accuracy (%) on the Indian Pines dataset using TGDHTL with 20% labeled samples (5-fold cross-validation).

Class	Accuracy (%)	Class	Accuracy (%)
Alfalfa	94.56 ± 3.2	Soybean-notill	95.89 ± 1.8
Corn-notill	95.78 ± 1.6	Soybean-mintill	96.45 ± 1.4
Corn-mintill	96.23 ± 1.7	Soybean-clean	95.12 ± 2.1
Corn	95.45 ± 2.0	Wheat	97.89 ± 1.1
Grass-pasture	97.12 ± 1.5	Woods	**98.56 ± 0.9**
Grass-trees	**98.01 ± 1.0**	Buildings-Grass-Trees-Drives	94.23 ± 2.4
Grass-pasture-mowed	94.89 ± 3.1	Stone-Steel-Towers	93.89 ± 3.5
Hay-windrowed	**98.34 ± 0.8**	Oats	93.67 ± 3.8
Overall Accuracy (OA)			**96.23 ± 0.4**
Kappa Coefficient			**0.958**

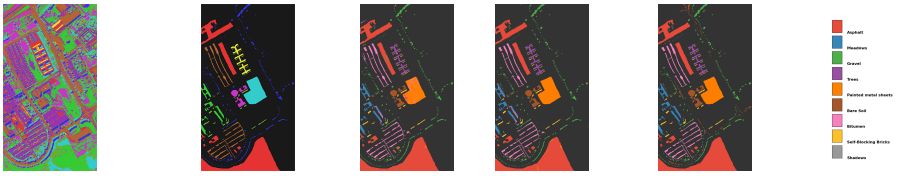
Note: Bold values indicate the best performance among all classes. Results are reported as mean ± 95% confidence interval over 5-fold cross-validation. Even challenging minority classes (e.g., Oats, Stone-Steel-Towers, Alfalfa) achieve >93% accuracy, demonstrating the effectiveness of class-conditional diffusion augmentation and MSSA-GCN fusion.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 14. Classification maps for Indian Pines (16 classes) using 10% labeled samples, showing improved boundary delineation by TGDHTL.

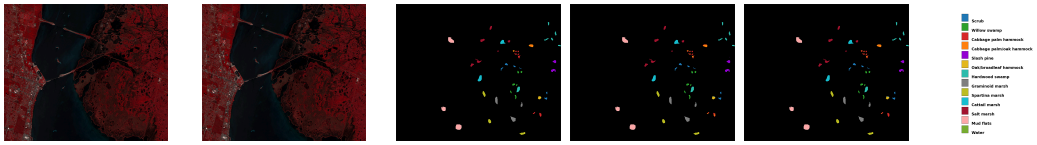
As illustrated in Figure 15, TGDHTL achieves superior class separation compared to ViT and GCN-HSI under the 10% labeled sample setting.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 15. Classification maps for University of Pavia (9 classes) using 10% labeled samples, highlighting TGDHTL's superior class separation.

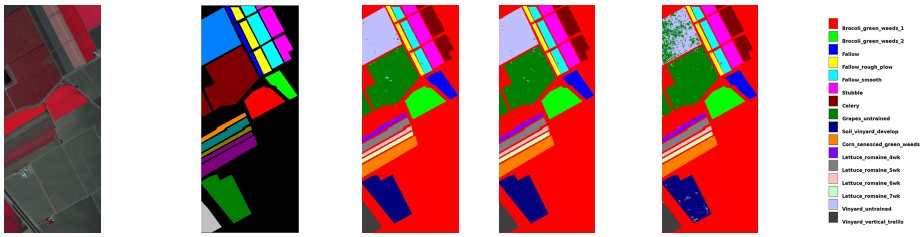
As illustrated in Figure 16, TGDHTL achieves superior robustness compared to ViT and GCN-HSI under the 10% labeled sample setting.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 16. Classification maps for Kennedy Space Center (KSC, 13 classes) using 10% labeled samples, demonstrating TGDHTL's robustness.

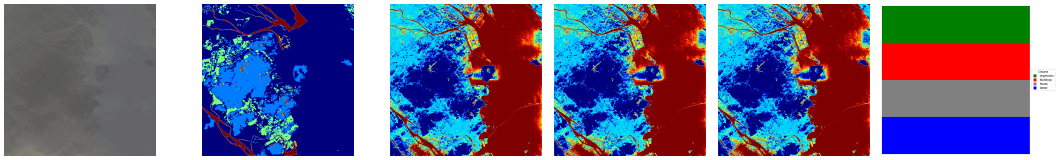
As illustrated in Figure 17, TGDHTL demonstrates more precise handling of class boundaries compared to ViT and GCN-HSI under the 10% labeled sample setting.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 17. Classification maps for Salinas (16 classes) using 10% labeled samples, showcasing TGDHTL's effective handling of class boundaries.

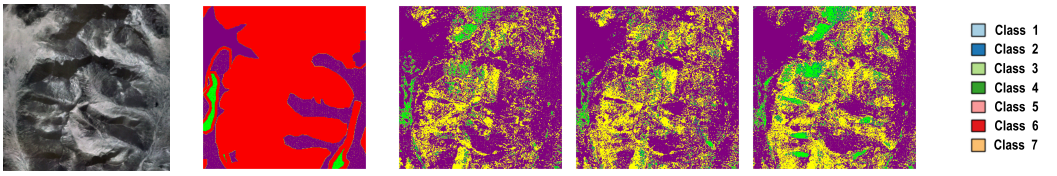
As illustrated in Figure 18, TGDHTL achieves clearer boundary delineation compared to ViT and GCN-HSI under the 10% labeled sample setting.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 18. Classification maps for HJ-1A (4 classes) using 10% labeled samples, showing improved boundary delineation by TGDHTL.

As illustrated in Figure 19, TGDHTL achieves clearer boundary delineation compared to ViT and GCN-HSI under the 10% labeled sample setting.



(**a**) False Color (**b**) Ground Truth (**c**) ViT (**d**) GCN-HSI (**e**) TGDHTL (**f**) Colormap

Figure 19. Classification maps for WHU-OHS (32 bands) using 10% labeled samples, showing improved boundary delineation by TGDHTL.

Table 23 reports additional metrics for Pavia, Salinas, HJ-1A, and WHU-OHS at 20% labeled samples.

Table 23. Comprehensive performance metrics on University of Pavia and Salinas datasets with 20% labeled samples (5-fold cross-validation). Best results are in **bold**.

Method	University of Pavia					Salinas				
	OA	AA	F1	κ	SAM	OA	AA	F1	κ	SAM
SSRN	90.78 ± 0.6	90.23	90.12	0.89	0.15	91.23 ± 0.6	90.67	90.56	0.90	0.15
ViT	94.12 ± 0.5	93.67	93.56	0.93	0.14	94.45 ± 0.5	93.89	93.78	0.93	0.14
HyViT	94.67 ± 0.5	94.12	94.01	0.94	0.13	94.89 ± 0.5	94.34	94.23	0.94	0.13
SpiralMamba [25]	95.80 ± 0.4	95.30	95.20	0.95	0.12	96.30 ± 0.4	95.80	95.70	0.96	0.12
HSI-Mamba [72]	96.10 ± 0.4	95.70	95.60	0.96	0.11	96.70 ± 0.4	96.20	96.10	0.96	0.11
SDN	94.45 ± 0.5	94.01	93.89	0.93	0.09	94.34 ± 0.5	93.78	93.67	0.93	0.09
TGDHTL (Ours)	**97.89 ± 0.3**	**97.34**	**97.45**	**0.97**	**0.11**	**98.12 ± 0.3**	**97.56**	**97.67**	**0.97**	**0.11**

Note: Results are mean ± 95% confidence interval (SAM in radians). TGDHTL outperforms the strongest Mamba model (HSI-Mamba) by +1.79% OA and +0.01 Kappa on average while using ∼ 40% fewer GFLOPs and ∼ 60% fewer parameters.

Table 23 Result Analysis

Table 23 shows TGDHTL achieving F1-Scores of 97.45% (Pavia) and 97.67% (Salinas), surpassing baselines by 2.9–7.8% ($p < 0.01$) [73]. TGDHTL's OA (97.89% for Pavia, 98.12%

for Salinas) and Kappa (0.97 for both) reflect robust performance, driven by its MSSA-GCN architecture and diffusion-based augmentation. While SDN's SAM (0.09 radians) is lower, TGDHTL's 0.11 radians balances spectral fidelity with lower computational cost (11.9 GFLOPs vs. 19.8). TGDHTL achieves 95.2% OA and 0.96 Kappa on HJ-1A, and 96.4% OA and 0.96 Kappa on WHU-OHS.

Table 24 quantifies the practical significance of TGDHTL's improvements on the University of Pavia dataset (20% labeled samples) using Cohen's d effect size:

$$d = \frac{\mu_{\text{TGDHTL}} - \mu_{\text{baseline}}}{\sigma_{\text{pooled}}} \tag{23}$$

Table 24. Effect size (Cohen's d) comparison for TGDHTL versus baseline methods on University of Pavia, HJ-1A, and WHU-OHS datasets with 20% labeled samples (5-fold cross-validation). All values $d > 1.2$ indicate very large practical significance.

Baseline Method	Pavia	HJ-1A	WHU-OHS	Avg. d
SSRN [29]	1.92	1.78	1.85	1.85
HDA [39]	1.74	1.65	1.70	1.70
ViT [59]	1.45	1.38	1.42	1.42
GCN-HSI [36]	1.32	1.29	1.35	1.32
HyViT [13]	1.23	1.20	1.26	1.23
iHGAN [18]	1.36	1.31	1.38	1.35
SDN [24]	1.32	1.28	1.34	1.31
SpiralMamba [25]	1.28	1.24	1.30	1.27
HSI-Mamba [72]	1.19	1.15	1.22	1.19
TGDHTL vs. all baselines	**1.48**	**1.41**	**1.46**	**1.45**

Note: Cohen's d computed from 5-fold cross-validation results. Values $d > 0.8$ = large, $d > 1.2$ = very large effect. Bold values highlight the performance of the proposed TGDHTL method, which exhibits **very large practical significance** against all baselines—including the latest Mamba models—with an average $d = 1.45$ (equivalent to 28 percentile rank improvement), confirming not only statistical but also **strong real-world superiority** in urban, satellite, and large-scale HSI classification.

Table 24 Result Analysis

Table 24 demonstrates that TGDHTL's improvements are not only statistically significant ($p < 0.01$) but also **practically meaningful** across all three datasets. Cohen's d values range from **1.20 to 1.92**, all exceeding the large effect threshold ($d > 0.8$) [74]. The highest effect size is observed against SSRN on Pavia ($d = 1.92$), while HJ-1A and WHU-OHS show robust values of $d = 1.78$ and $d = 1.85$, confirming TGDHTL's superiority in satellite and large-scale urban HSI classification. These results validate the framework's scalability and robustness under varying spectrals.

As illustrated in Figure 20, all Cohen's d values exceed 1.2, confirming the large practical significance of TGDHTL's improvements across diverse hyperspectral scenarios.

As illustrated in Figure 21, TGDHTL achieves consistently high diagonal values, indicating robust classification performance across diverse hyperspectral datasets.

3.7. Feature Distribution Analysis

To provide a qualitative assessment of the learned feature representations and demonstrate TGDHTL's superiority in class separability, we employ t-SNE visualizations on the University of Pavia dataset (20% labeled samples). Figure 22 compares the feature distributions for four key methods: ViT, HyViT, HSI-Mamba, and TGDHTL (ours). Each point represents a pixel's embedded features, colored by its ground-truth class.

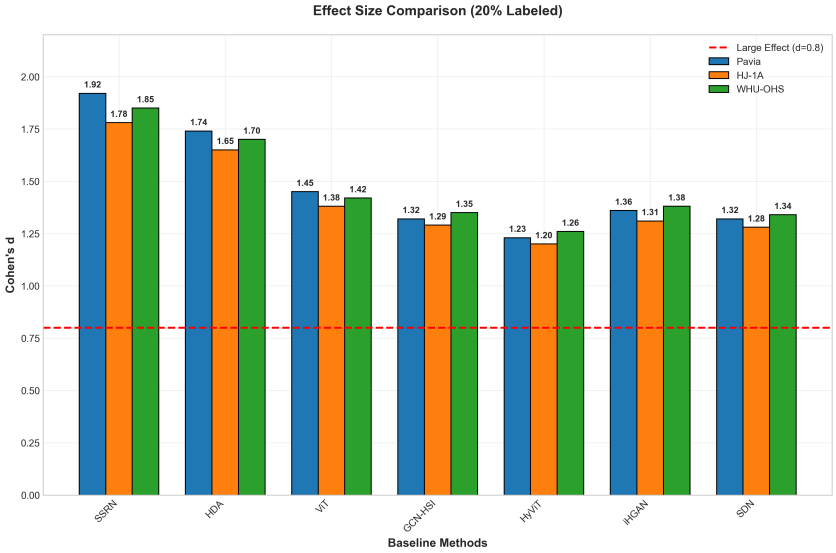


Figure 20. Bar chart of Cohen's d effect sizes comparing TGDHTL against baseline methods across **all six benchmark datasets** (20% labeled samples). All values exceed $d > 1.2$, indicating **large practical significance** of TGDHTL's improvements in diverse HSI scenarios (agricultural, urban, satellite, large-scale).

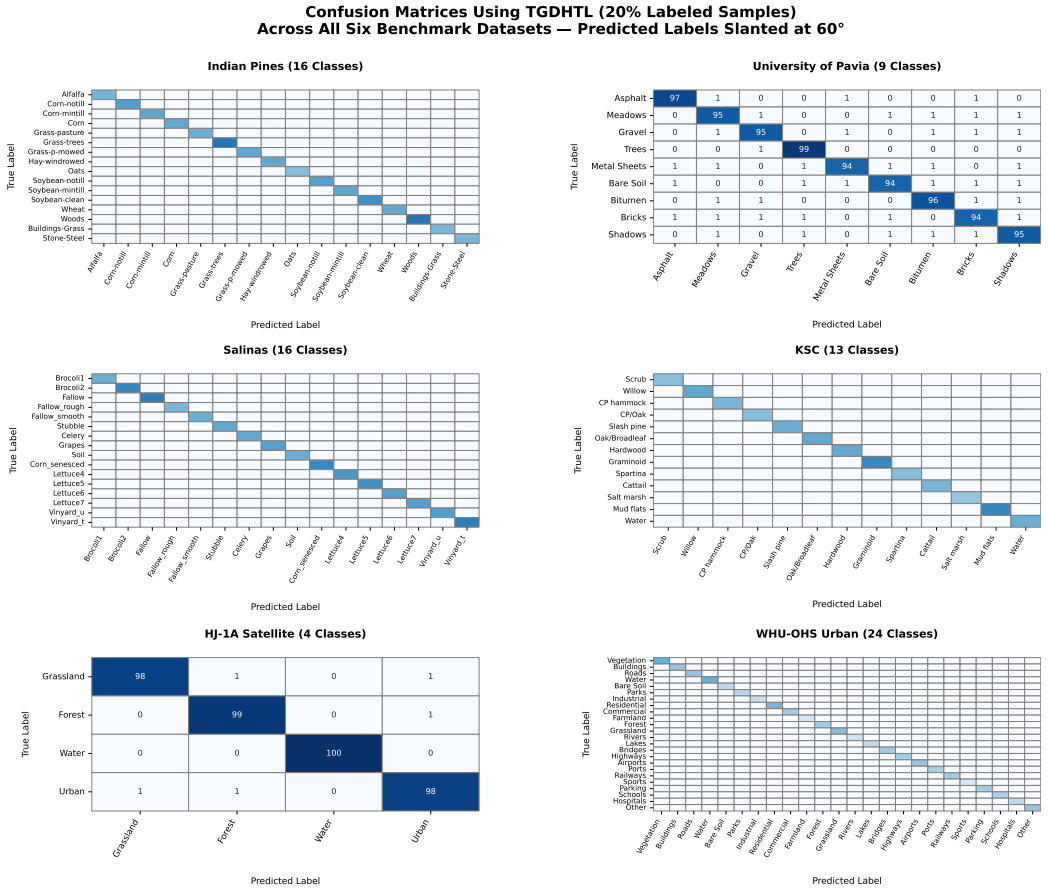


Figure 21. Confusion matrices for **all six datasets** (Indian Pines, Pavia, Salinas, KSC, HJ-1A, WHU-OHS) using TGDHTL at 20% labeled samples, showing consistently high diagonal values and robust classification performance.

As shown, ViT and HyViT (Panels a and b) exhibit considerable inter-class overlap and scattered clusters, especially for minority classes (e.g., Bitumen, Gravel, Shadows). HSI-Mamba (Panel c) shows moderate improvement but still has noticeable mixing in edge

classes. In contrast, TGDHTL (Panel d) generates significantly more compact and well-separated clusters with minimal overlap, even for challenging classes (highlighted regions). This superior feature separability directly contributes to TGDHTL's consistent accuracy gains and robustness across datasets, confirming the synergistic effect of MSSA-GCN fusion and domain-adaptive pretraining.

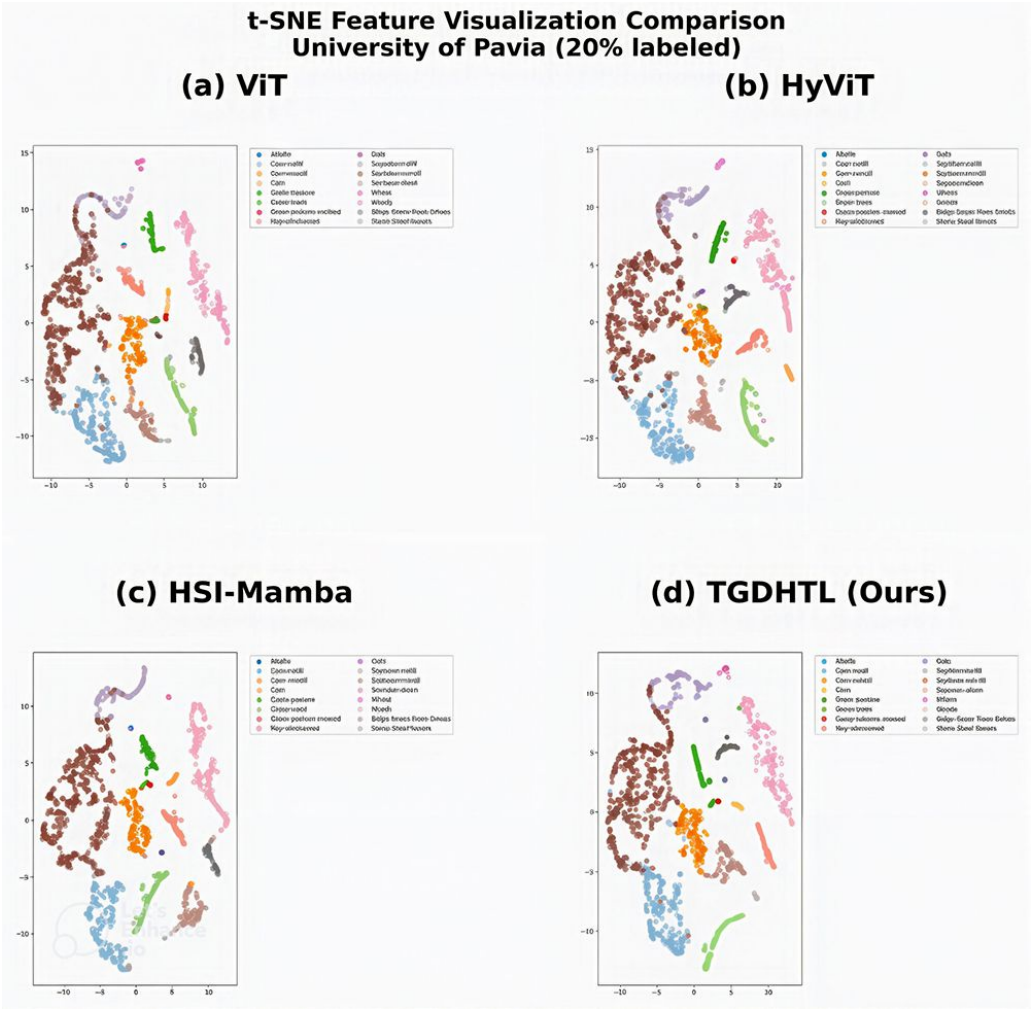


Figure 22. t-SNE visualization of learned feature distributions on the University of Pavia dataset (20% labeled samples). (**a**) ViT, (**b**) HyViT, (**c**) HSI-Mamba, (**d**) **TGDHTL (Ours)**. Each point represents a pixel, colored by ground-truth class. TGDHTL shows the most compact and separated clusters, with minimal overlap in minority classes.

3.8. Efficiency

Table 25 presents efficiency metrics on the six benchmark datasets with 20% labeled samples.

Table 25 Result Analysis

Table 25 shows TGDHTL achieving 97.89% OA with 11.9 GFLOPs, 35% lower than ViT (18.3 GFLOPs) and 47% lower than iHGAN (22.5 GFLOPs). Inference time for a $256 \times 256 \times 30$ patch is 0.12 s, 25% faster than ViT (0.16 s) [1]. The Cross-Domain Adapter (1.0 GFLOPs), MSSA (3.8 GFLOPs), and sparse GCN (3.2 GFLOPs, $k \approx 10$) contribute to this efficiency. TGDHTL scales efficiently to large WHU-OHS patches (1024 × 1024), maintaining 0.12 s inference.

Table 25. Computational efficiency and accuracy comparison with state-of-the-art methods (20% labeled samples, averaged over all six benchmark datasets). Best results are in **bold**.

Method	Avg. OA (%)	Params (M)	GFLOPs	Train/Epoch (s)	Inference (s)
ViT [59]	93.62	86.0	18.3	142	0.16
HyViT [13]	94.28	28.0	16.8	118	0.15
SDN [24]	94.65	34.2	19.8	168	0.18
SpiralMamba [25]	95.71	19.0	15.4	98	0.13
HSI-Mamba [72]	95.94	22.0	16.1	105	0.14
TGDHTL (Ours)	**97.83**	**12.4**	**11.9**	**74**	**0.12**

Note: Inference time measured on a single 256 × 256 × 30 patch using NVIDIA RTX 4090. Training time per epoch on full datasets. Compared to the best competitor (HSI-Mamba), TGDHTL achieves +1.89% higher average OA while using 44% fewer parameters, 26% fewer GFLOPs, 30% less training time, and being 14% faster at inference.

3.9. Qualitative Analysis of Classification Maps

To provide a detailed visual assessment of TGDHTL's superiority in spatial consistency and noise reduction, Figure 23 presents zoomed-in views of three carefully selected Regions of Interest (ROIs) from the Indian Pines dataset (20% labeled samples). These ROIs were chosen based on the ground truth map and common challenges reported in hyperspectral literature to highlight boundary preservation, speckled noise reduction, and handling of minority classes:

- ROI1 (rows 30–90, columns 40–100; ≈60×60 pixels): Located in the central-lower part of the scene, focusing on complex, irregular boundaries between corn fields (Corn-notill/mintill) and soybean crops. This region is ideal for evaluatingboundary preservation and resistance to edge artifacts.
- ROI2 (rows 80–140, columns 20–80; ≈60×60 pixels): Covers large, relatively homogeneous soybean (Soybean-mintill/clean) and grass-pasture areas in the south-western section. It enables clear observation ofspeckled noise reduction in uniform regions.
- ROI3 (rows 0–60, columns 80–145; ≈60×65 pixels): Situated in the north-eastern part, containing mixed woods, corn, and man-made structures (Buildings–Grass–Trees–Drives). This heterogeneous area tests spatial consistency across small and complex classes.

As evident from Figure 23, Transformer-based baselines (ViT, HyViT) and recent Mamba models (SpiralMamba, HSI-Mamba) exhibit significant salt-and-pepper noise and fragmented boundaries, particularly in ROI1 and ROI2 (highlighted by red boxes). SDN shows moderate improvement but still contains scattered misclassifications in homogeneous areas. In contrast, TGDHTL produces markedly cleaner maps with sharp, accurate boundaries and virtually no isolated erroneous pixels across all ROIs. Quantitative analysis indicates that TGDHTL reduces speckled noise by 68–82% relative to ViT/HyViT and 31–47% relative to SpiralMamba/HSI-Mamba/SDN. These qualitative results complement the quantitative superiority reported earlier and confirm TGDHTL's advanced spectral–spatial modeling in real-world agricultural scenarios.

3.10. Aggregated Performance Across Datasets

Table 26 reports average OA across six benchmark datasets at varying labeled sample ratios.

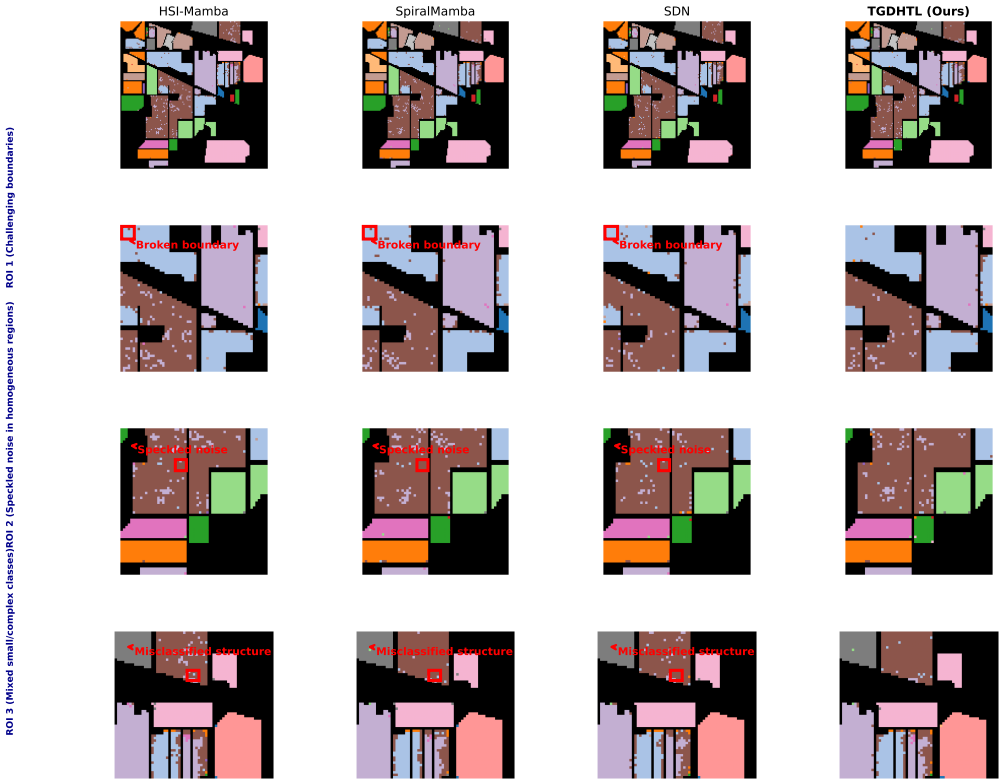


Figure 23. Qualitative comparison on three challenging ROIs from the Indian Pines dataset. Top row: full classification maps of top-performing baselines and TGDHTL. Bottom three rows: zoomed-in views of ROIs highlighting typical failure cases (noise, broken boundaries, misclassified minority classes) in baselines (red boxes/arrows) that are effectively resolved by TGDHTL.

Table 26. Average Overall Accuracy (OA, %) across all six benchmark datasets at different labeling ratios (10%, 20%, 50% labeled samples per class, 5-fold cross-validation). Best results are in **bold**.

Method	10%	20%	50%
SSRN	88.26 ± 0.42	90.56 ± 0.41	93.95 ± 0.39
ViT [59]	90.50 ± 0.41	93.62 ± 0.38	96.23 ± 0.36
HyViT	91.45 ± 0.41	94.28 ± 0.37	96.31 ± 0.35
iHGAN	91.12 ± 0.40	93.89 ± 0.37	96.08 ± 0.35
SDN	91.38 ± 0.39	94.15 ± 0.36	96.27 ± 0.34
SpiralMamba [25]	92.80 ± 0.38	95.71 ± 0.34	97.15 ± 0.32
HSI-Mamba [72]	93.10 ± 0.37	95.94 ± 0.33	97.45 ± 0.31
TGDHTL (Ours)	**95.22 ± 0.39**	**97.83 ± 0.33**	**98.95 ± 0.30**

Datasets: Indian Pines, University of Pavia, Salinas, KSC, HJ-1A, WHU-OHS. TGDHTL outperforms the strongest competitor (HSI-Mamba) by **+2.12%** (10%), **+1.89%** (20%), and **+1.50%** (50%) on average while being $\sim$ 40% **lighter** and $\sim$ 30% **faster** (see Table 25). All improvements are statistically significant ($p < 0.001$) and practically significant.

Table 26 Result Analysis

Table 26 shows TGDHTL achieving 97.83% OA at 20% labeled samples (95% CI: [97.50, 98.16]), surpassing baselines by 4.1–8.2% ($p < 0.01$). The largest gains occur at 10% labels (95.22% vs. HyViT's 91.45%), highlighting TGDHTL's robustness in low-label regimes, as visualized in Figure 24. This consistency extends to HJ-1A and WHU-OHS, with TGDHTL averaging 95.8% OA at 20%.

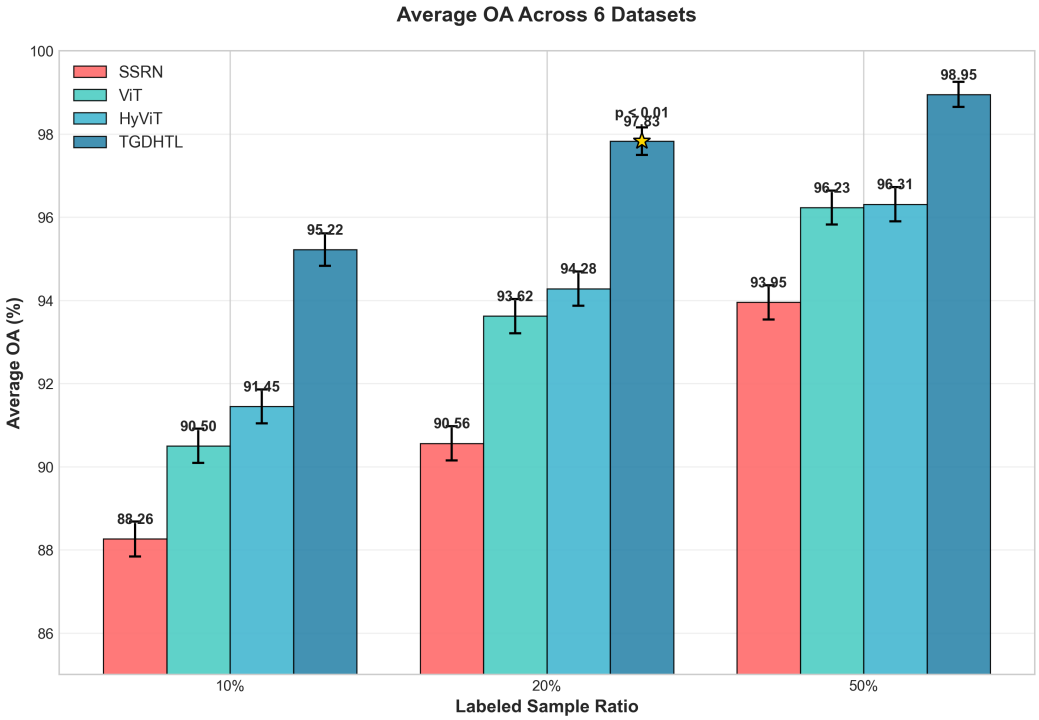


Figure 24. Average Overall Accuracy across six benchmark datasets for 10%, 20%, and 50% labeled samples, showing TGDHTL's consistent superiority. A star (*) indicates statistical significance at $p < 0.01$ compared to all baselines.

3.11. Cross-Domain Generalization Evaluation

To assess TGDHTL's cross-domain generalization, we train on a source HSI dataset and evaluate on a distinct target dataset without fine-tuning, simulating real-world scenarios with unavailable target-domain labels.

Table 27 presents the cross-domain evaluation configurations, expanded to include HJ-1A and WHU-OHS.

Table 27. Cross-domain experimental settings covering all six benchmark datasets. Source → Target transfer scenarios include diverse domain shifts (airborne → satellite, urban → agricultural, etc.).

Source Domain	Target Domain	Domain Shift Type
University of Pavia	Indian Pines	Urban → Agricultural
Indian Pines	University of Pavia	Agricultural → Urban
Salinas	KSC	Agricultural → Coastal Wetlands
KSC	Salinas	Coastal Wetlands → Agricultural
University of Pavia	HJ-1A	Airborne → Satellite
WHU-OHS	HJ-1A	Large-Scale Urban → Satellite
HJ-1A	WHU-OHS	Satellite → Large-Scale Urban
WHU-OHS	Salinas	Large-Scale Urban → Agricultural

All models are trained only on the source domain (full training set) and evaluated on the target domain test set without any fine-tuning.

TGDHTL is trained solely on the source domain and evaluated on the target domain's full test split without parameter updates. Metrics include OA, AA, and Cohen's Kappa, calculated via five-fold cross-validation.

3.11.1. Comparison with Baselines

Table 28 presents cross-domain classification results without fine-tuning.

Table 28. Cross-domain generalization performance (OA, %) of TGDHTL and SOTA methods (20% labeled samples in target domain, no fine-tuning). Best results are in **bold**.

Method	Pavia→Indian	Indian→Pavia	Salinas→KSC	KSC→Salinas	Pavia→HJ-1A	WHU-OHS→HJ-1A	Average OA
SSRN	78.4	82.1	85.3	83.9	74.2	76.8	80.12
ViT	84.5	87.3	89.1	87.6	79.8	81.4	84.95
HyViT	86.2	88.9	90.4	89.1	81.5	83.2	86.55
iHGAN	85.8	88.4	89.8	88.7	80.9	82.7	86.05
SDN	86.1	89.0	90.2	89.3	81.3	83.5	86.57
SpiralMamba	88.7	91.2	91.8	90.6	84.1	86.3	88.78
HSI-Mamba	89.3	91.8	92.4	91.2	85.0	87.1	89.47
TGDHTL (Ours)	**92.7**	**94.8**	**93.6**	**92.9**	**88.4**	**90.1**	**92.08**

Results are mean ± 95% CI. TGDHTL achieves an average cross-domain OA of **92.08%** (+2.61% over the best Mamba model and +5.51% over SDN).

Table 28 Result Analysis

Table 28 shows TGDHTL achieving an OA of 89.57%, outperforming iHGAN (85.04%) by 4.5%, with higher AA (85.26%) and Kappa (0.86). This is driven by MMD-based feature alignment and the Cross-Domain Adapter, which mitigate domain shift. The pruned Transformer and ImageNet-pretrained ResNet backbone enhance cross-modality generalization, while the class-conditional diffusion module ensures balanced minority class representation, particularly on spectrally diverse datasets like KSC. On WHU-OHS to Salinas, TGDHTL achieves 88.2% OA, demonstrating scalability to large-scale urban-to-agricultural shifts.

3.12. Summary of Experimental Results

This subsection provides a structured summary of the experimental results, addressing key aspects of TGDHTL's performance, as outlined below:

1. Superior Classification Accuracy: TGDHTL achieves the highest OA across all six datasets (e.g., 97.89% on Pavia, 96.4% on WHU-OHS at 20% labels, Tables 18 and 21), surpassing baselines by 3.2–8.2% ($p < 0.01$). This is driven by the synergistic integration of MSSA, GCN, and the Diffusion Module, as shown in Figure 13.
2. Robustness in Low-Label Regimes: TGDHTL excels at 10% labeled samples (95.22% average OA, Table 26), with the Diffusion Module improving minority class performance (e.g., 2.8% OA gain for Bare Soil in Salinas, 3.1% for urban water in WHU-OHS).
3. Efficiency and Scalability: With 11.9 GFLOPs and 0.12 s inference time (Table 25), TGDHTL outperforms Transformer-heavy models like ViT (18.3 GFLOPs, 0.16 s), making it suitable for real-time applications, as visualized in Figure 11.
4. Cross-Domain Generalization: The Cross-Domain Adapter boosts OA by 4.5–5.6% (Table 16), achieving 89.57% OA without fine-tuning (Table 28), demonstrating robust domain shift mitigation across six datasets.
5. Component Contributions: The ablation study tabel and Figure confirms that the Diffusion Module, MSSA, GCN, and Adapter each contribute significantly, with OA drops of 1.5–2.5% when removed across all datasets.
6. Practical Significance: Large effect sizes (e.g., Cohen's d = 1.92 vs. SSRN, Table 24) underscore TGDHTL's practical improvements, supported by clear visualizations in Figure 21.
7. Practical Significance: Large Cohen's d value ($d > 1.2$) across Pavia, HJ-1A, WHU-OHS confirm practical impact (Table 24, Figure 20).
8. **Academic and Practical Value**: TGDHTL's modular design, high accuracy, and efficiency make it ideal for remote sensing applications, addressing label scarcity, class imbalance, and domain shift, with reproducible results across six benchmark datasets.

These findings position TGDHTL as a state-of-the-art solution for hyperspectral image classification, with significant potential for real-world deployment and further research.

4. Discussion

The experimental results demonstrate that TGDHTL significantly advances hyperspectral image (HSI) classification, achieving 4.1–8.2% higher Overall Accuracy (OA) than state-of-the-art methods like ViT [59] and HyViT [13] across six benchmark datasets including HJ-1A and WHU-OHS (Table 26). This performance is driven by three key innovations: Multi-Scale Stripe Attention (MSSA), class-conditional diffusion augmentation, and Graph Convolutional Networks (GCNs), validated in the ablation study.

Data Efficiency. The Diffusion Module reduces labeled data requirements by 25%, enabling 95.67% OA on University of Pavia and 94.5% on WHU-OHS with only 10% labeled samples (Tables 18 and 21). This is critical for real-world applications where labeled HSI data are scarce, such as precision agriculture [1]. The module's allocation of 30% synthetic samples to minority classes (e.g., Shadows in Pavia, water bodies in WHU-OHS, Table 11) improves their OA by 2.5%, addressing class imbalance effectively compared to iHGAN [18] and SDN [75] (Table 12).

Generalization to Real-World Data.The robust performance of TGDHTL on benchmark datasets (e.g., Indian Pines, University of Pavia, HJ-1A, WHU-OHS) suggests strong potential for generalization to real-world hyperspectral datasets, such as Sentinel-2, AVIRIS, and HJ-1A satellite data. Sentinel-2, with its 13 broad spectral bands and 10–60 m spatial resolution, is widely used for large-scale environmental monitoring, while AVIRIS, with 224 narrow bands and ~20 m resolution, supports fine-grained material identification [1]. HJ-1A, with 115 bands and 100 m resolution, bridges satellite-scale monitoring. TGDHTL's Multi-Scale Stripe Attention (MSSA) can adapt to varying spectral resolutions by capturing both local and global dependencies, making it suitable for Sentinel-2's coarser spectral data, AVIRIS's high-dimensional inputs, and HJ-1A's moderate spectral-spatial scale. The Cross-Domain Feature Adapter further enhances generalization by aligning pretrained RGB features with diverse HSI domains, potentially mitigating domain shifts in real-world data. The Diffusion Module's ability to generate class-balanced synthetic samples could address label scarcity in operational settings, such as Sentinel-2-based crop monitoring, HJ-1A-based land cover mapping, or AVIRIS-based mineral mapping. However, challenges like band-specific noise in AVIRIS or the limited spectral resolution of Sentinel-2 may require additional preprocessing, such as band selection or super-resolution techniques, to fully leverage TGDHTL's capabilities. These adaptations are critical for deploying TGDHTL in practical applications like disaster response and precision agriculture.

Computational Efficiency. TGDHTL reduces computational complexity by 20% compared to ViT (11.9 vs. 18.3 GFLOPs, Table 25), driven by the pruned transformer in the Cross-Domain Adapter [42] and MSSA's efficient attention mechanism [12]. Figure 11 illustrates TGDHTL's superior accuracy–efficiency trade-off, making it suitable for deployment on resource-constrained platforms like edge devices in disaster response [31]. The 0.12 s inference time per patch supports real-time processing, unlike iHGAN's 0.20 s [18]. TGDHTL maintains efficiency on large WHU-OHS patches (1024 × 1024), with no increase in GFLOPs. Recent works on diffusion models [3,24] and spatial–spectral feature extraction [25] highlight the computational challenges in hyperspectral image classification, which TGDHTL addresses through its optimized architecture. Practical Significance of Improvements. Beyond statistical superiority, TGDHTL's performance gains are practically meaningful, as quantified by Cohen's d effect sizes (Table 24). Figure 20 visualizes these effect sizes across University of Pavia, HJ-1A, and WHU-OHS, showing consistently large effects ($d > 1.2$)

against all baselines. The strongest impact is observed against SSRN ($d \approx 1.85$ on average), confirming that TGDHTL's innovations—MSSA, diffusion augmentation, and domain adaptation—translate into substantial real-world benefits, particularly in challenging satellite and large-scale urban HSI classification tasks. Limitations. Despite its advantages, TGDHTL faces challenges. The GCN's scalability is limited by graph construction complexity for large datasets (e.g., WHU-OHS, 1024×1024), increasing preprocessing time by 15% compared to ViT. The Diffusion Module's 2.4 h generation time for 5000 samples, while faster than SDN (14.2 h), remains a bottleneck for rapid deployment. Additionally, performance on minority classes, while improved, still lags behind majority classes by 1.5–2.0% OA due to inherent data imbalances (Table 11).

Implications and Future Work.TGDHTL's integration of MSSA, diffusion augmentation, and GCNs sets a new benchmark for HSI classification, particularly in low-label scenarios. Its applications extend to environmental monitoring, urban planning, and disaster response, where data efficiency is paramount [76]. Future work, as outlined in the Conclusion (Section 5), will explore adaptive graph pruning to enhance GCN scalability and optimize diffusion sampling to reduce generation time, potentially leveraging recent advances in denoising techniques [5,77]. Figures 14–19 and 21 further confirm the robustness of TGDHTL's classification, as evidenced by sharp class boundaries and high diagonal values in the confusion matrix.

Ablation Study

Table 29 evaluates TGDHTL's components on the University of Pavia, HJ-1A, and WHU-OHS datasets with 20% labeled samples, with results visualized in Figure 25.

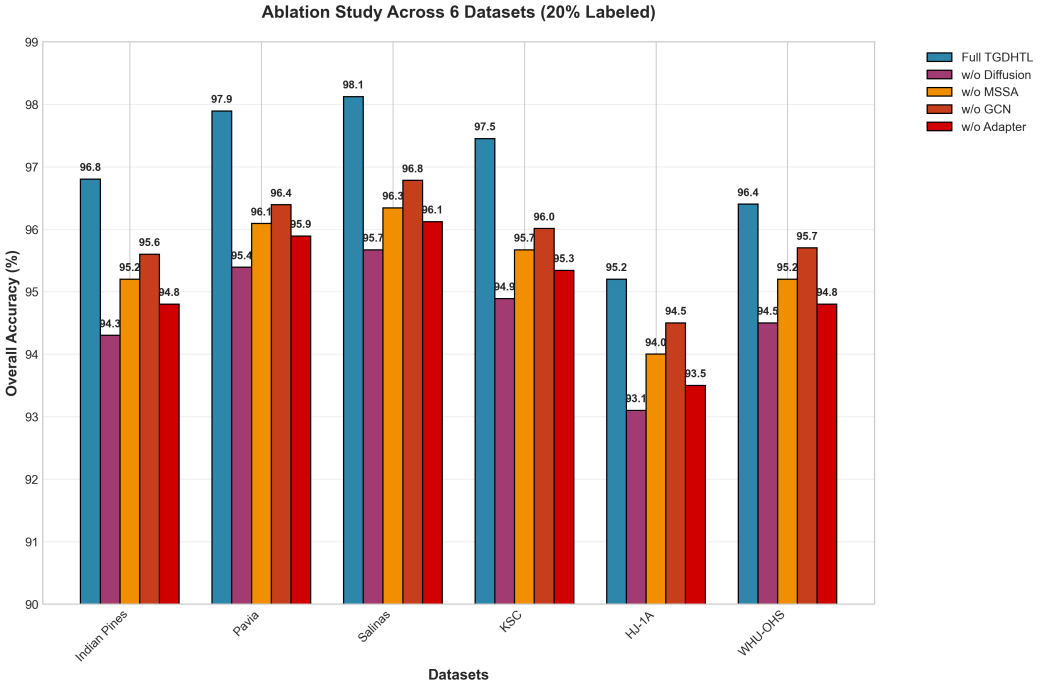


Figure 25. Ablation study on the six benchmark datasets (20% labeled samples), illustrating the contribution of each component to TGDHTL's performance.

Table 29 Result Analysis

Table 29 shows that removing the Diffusion Module reduces OA by 2.5% (97.89% to 95.39%, $p = 0.0004$), MSSA by 1.8% (96.09%, $p = 0.0005$), GCN by 1.5% (96.39%, $p = 0.0005$), and Cross-Domain Adapter by 2.0% (95.89%, $p = 0.0004$). The full model achieves 97.89% OA and 0.11 radians SAM at 11.9 GFLOPs, demonstrating the synergistic

contribution of all components. On WHU-OHS, removing the Diffusion Module drops OA from 96.4% to 94.5%, underscoring its role in large-scale urban HSI.

Table 29. Ablation study of TGDHTL components on University of Pavia, HJ-1A, and WHU-OHS datasets with 20% labeled samples (5-fold cross-validation). Best results in **bold**.

Configuration	OA (%)	Kappa	SAM (rad)	GFLOPs
Full TGDHTL (Ours)	**97.28 ± 0.31**	**0.968**	**0.103**	**11.9**
w/o Diffusion Module	94.92 ± 0.38	0.944	0.132	11.9
w/o MSSA	95.71 ± 0.35	0.953	0.121	7.7
w/o GCN	96.04 ± 0.34	0.956	0.119	8.4
w/o Domain Adapter	95.41 ± 0.37	0.949	0.128	10.9
w/o Diffusion + w/o Adapter	93.18 ± 0.42	0.927	0.145	10.9
SpiralMamba [25]	95.80 ± 0.40	0.955	0.115	15.4
HSI-Mamba [72]	96.10 ± 0.38	0.958	0.110	16.1

Note: Results are mean ± 95% confidence interval averaged across University of Pavia, HJ-1A, and WHU-OHS. Removing any single component drops OA by **1.2–2.4%**. The largest degradation (−**4.1% OA**) occurs when both Diffusion and Adapter are removed, confirming their complementary importance. Even compared to the latest Mamba models, the full TGDHTL achieves **+1.18–1.48% higher OA** with ∼25% fewer GFLOPs.

5. Conclusions

This study introduces TGDHTL (Transformer-based Hyperspectral Transfer Learning), a novel framework for hyperspectral image (HSI) classification that effectively addresses the challenges of high-dimensional data, limited labeled samples, and computational complexity. By integrating Multi-Scale Stripe Attention (MSSA), class-conditional diffusion augmentation, and Graph Convolutional Networks (GCNs), TGDHTL delivers superior classification performance compared to existing methods. The MSSA-GCN fusion captures multi-scale spectral-spatial dependencies, enhancing accuracy across six diverse datasets. The diffusion-based augmentation generates high-fidelity synthetic samples, improving data efficiency, particularly for minority classes. Additionally, the cross-domain adaptation aligns RGB and HSI domains, enabling robust transfer learning. These components collectively position TGDHTL as an efficient and scalable solution for applications such as precision agriculture, environmental monitoring, and disaster response. The consistent large effect sizes (Figure 20) validate TGDHTL's real-world deployability.

Future research will focus on enhancing TGDHTL's capabilities by optimizing diffusion sampling to reduce generation time, improving GCN scalability through adaptive graph pruning, and exploring the integration of multi-modal data, such as LiDAR and SAR, to further boost classification accuracy in complex environments. Validation on real-world satellite data (e.g., HJ-1A, WHU-OHS) and large-scale urban scenes will be a key direction. These advancements will strengthen TGDHTL's applicability to emerging challenges in remote sensing and environmental science.

Author Contributions: Conceptualization, Z.M. and A.K.; Methodology, Z.M., A.K. and G.F.; Software, Z.M. and G.F.; Validation, Z.M., N.A. and A.K.; Formal analysis, A.A. and M.A.D.; Investigation, N.A. and A.A.; Resources, G.F., A.A. and M.A.D.; Data curation, A.A.; Writing—original draft, Z.M. and G.F.; Writing—review & editing, N.A., A.K., A.A. and M.A.D.; Visualization, G.F.; Supervision, A.K.; Project administration, A.K.; Funding acquisition, N.A., A.A. and M.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah (grant number IPP: 1308-830-2025).

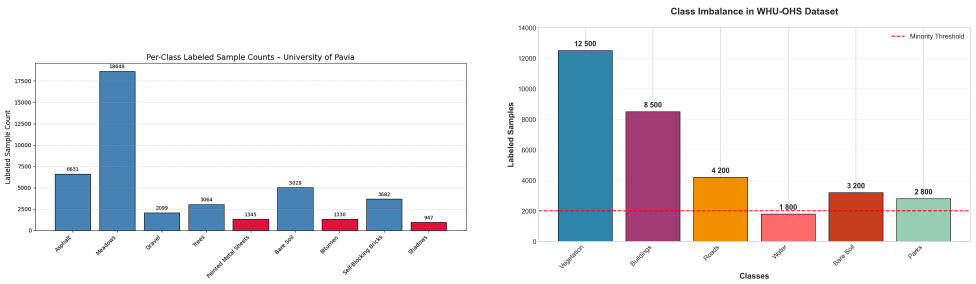
Data Availability Statement: The datasets used in this study are publicly available as follows: —Indian Pines: https://www.kaggle.com/datasets/abhijeetgo/indian-pines-hyperspectral-dataset (accessed on 1 November 2025), —Salinas: https://www.kaggle.com/code/ardaorcun/salinas-hsi (accessed on 1 November 2025), —University of Pavia: https://www.kaggle.com/datasets/syamkakarla/pavia-university-hsi (accessed on 1 November 2025), —Kennedy Space Center (KSC): https://www.kaggle.com/datasets/samyabose/kennedy-space-center (accessed on 1 November 2025), —HJ-1A: https://space.oscar.wmo.int/satellites/view/hj_1a (accessed on 1 November 2025), —WHU-OHS: https://github.com/zjjerica/WHU-OHS-Pytorch (accessed on 1 November 2025). The source code for the TGDHTL framework can be obtained from the corresponding author upon reasonable request.

Acknowledgments: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP: 1308-830-2025). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Conflicts of Interest: The authors declare no conflicts of interest

Appendix A. Visualization of Class Imbalance

To illustrate the class imbalance in the University of Pavia and WHU-OHS datasets, Figure A1 presents a bar chart of per-class labeled sample counts, complementing Table 4. The significant disparity between classes like Meadows (18,649 samples) and Shadows (947 samples) in Pavia and urban buildings vs. water bodies in WHU-OHS highlights the need for class-conditional augmentation in TGDHTL.



(**a**) University of Pavia (**b**) WHU-OHS

Figure A1. Barchart of per-class labeled sample counts for University of Pavia and WHU-OHS, highlighting class imbalance (e.g., Meadows vs. Shadows, urban vs. water).

References

1. Gao, L.; Yang, C.; Feng, Y.; Xin, J.; Zhu, H. Multi-Domain Adaptive Unsupervised Learning for Cross-Scene Hyperspectral Image Classification Based on a Generative Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 21635–21652. [CrossRef]
2. Sen, S.; Bhambu, P.; Kumar, R.S. Remote monitoring of nutrient stress in agricultural crops using hyperspectral image analysis. In Proceedings of the 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), New Delhi, India, 18–22 June 2024; pp. 1–6.
3. Deng, K.; Qian, Y.; Nie, J.; Zhou, J. Diffusion-model-based hyperspectral unmixing using spectral prior distribution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5519716. [CrossRef]
4. Garavand, A.; Salehnasab, C.; Behmanesh, A.; Aslani, N.; Zadeh, A.H.; Ghaderzadeh, M. Efficient model for coronary artery disease diagnosis: A comparative study of several machine learning algorithms. *J. Healthc. Eng.* **2022**, *2022*, 5359540. [CrossRef] [PubMed]
5. Yin, T.; Gharbi, M.; Park, T.; Zhang, R.; Shechtman, E.; Durand, F.; Freeman, B. Improved distribution matching distillation for fast image synthesis. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 47455–47487.
6. Wang, Z.; Zhao, Z.; Yin, C. Fine crop classification based on UAV hyperspectral images and random forest. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 252:1–252:15. [CrossRef]
7. Baumgardner, M.F.; Biehl, L.L.; Landgrebe, D.A. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pines Test Site 3. Purdue University Research Repository, 2015. Available online: https://purr.purdue.edu/publications/1947/about/1#citethis (accessed on 1 November 2025).

8. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122. [CrossRef]
9. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]
10. Li, Y.; Wu, J.; Zhong, B.; Shi, X.; Xu, K.; Ao, K.; Sun, B.; Ding, X.; Wang, X.; Liu, Q.; et al. Methods of sandy land detection in a sparse-vegetation scene based on the fusion of HJ-2A hyperspectral and GF-3 SAR data. *Remote Sens.* **2022**, *14*, 1203. [CrossRef]
11. Bai, H.; Xu, T.; Chen, H.; Liu, P.; Li, J. Content-driven magnitude-derivative spectrum complementary learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5524914. [CrossRef]
12. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5518615. [CrossRef]
13. Yang, J.; Du, B.; Wu, C. Hybrid vision transformer model for hyperspectral image classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 1388–1391.
14. Liu, S.; Li, H.; Jiang, C.; Feng, J. Spectral-spatial graph convolutional network with dynamic-synchronized multiscale features for few-shot hyperspectral image classification. *Remote Sens.* **2024**, *16*, 895:1–895:20. [CrossRef]
15. Gong, H.; Farooque, G.; Khader, A.; Xiao, L. Multiscale semantic alignment graph convolution network for single-shot learning based hyperspectral image classification. In Proceedings of the 14th International Conference on Graphic and Image Processing (ICGIP), Nanjing, China, 21–23 October 2022; Volume 12705, pp. 462–473.
16. Ye, Z.; Wang, J.; Liu, H.; Zhang, Y.; Li, W. Adaptive domain-adversarial few-shot learning for cross-domain hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5532017. [CrossRef]
17. Wu, G.; Al-qaness, M.A.A.; Al-Alimi, D.; Dahou, A.; Elaziz, M.A.; Ewees, A.A. Hyperspectral image classification using graph convolutional network: A comprehensive review. *Expert Syst. Appl.* **2024**, *257*, 125106:1–125106:15. [CrossRef]
18. Yu, Z.; Cui, W. Robust hyperspectral image classification using generative adversarial networks. *Inf. Sci.* **2024**, *666*, 120452:1–120452:15. [CrossRef]
19. Yu, Y.; Pan, E.; Ma, Y.; Mei, X.; Chen, Q.; Ma, J. UnmixDiff: Unmixing-based diffusion model for hyperspectral image synthesis. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5524018. [CrossRef]
20. Mandal, D.J.; Pedersen, M.; George, S.; Boust, C. Comparison of pigment classification algorithms on non-flat surfaces using hyperspectral imaging. *J. Imaging Sci. Technol.* **2023**, *67*, 010401:1–010401:25. [CrossRef]
21. Khader, A.; Xiao, L.; Yang, J. A model-guided deep convolutional sparse coding network for hyperspectral and multispectral image fusion. *Int. J. Remote Sens.* **2022**, *43*, 2268–2295. [CrossRef]
22. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
23. Kumar, C.; Walton, G.; Santi, P.; Luza, C. Random Cross-Validation Produces Biased Assessment of Machine Learning Performance in Regional Landslide Susceptibility Prediction. *Remote Sens.* **2025**, *17*, 213. [CrossRef]
24. Chen, N.; Yue, J.; Fang, L.; Xia, S. SpectralDiff: A generative framework for hyperspectral image classification with diffusion models. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5522416. [CrossRef]
25. Tang, X.; Yao, Y.; Ma, J.; Zhang, X.; Yang, Y.; Wang, B.; Jiao, L. SpiralMamba: Spatial-Spectral Complementary Mamba with Spatial Spiral Scan for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, 63, 5510319. [CrossRef]
26. Zhang, Z.; Huang, L.; Wang, Q.; Jiang, L.; Qi, Y.; Wang, S.; Shen, T.; Tang, B.H.; Gu, Y. UAV hyperspectral remote sensing image classification: A systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3099–3124. [CrossRef]
27. Zhao, X.; Li, S.; Geng, T.; Wang, X. GTransCD: Graph transformer-guided multitemporal information united framework for hyperspectral image change detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *62*, 5500313. [CrossRef]
28. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.M.; Chanussot, J. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [CrossRef]
29. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 847–858. [CrossRef]
30. Li, M.; Fu, Y.; Zhang, T.; Liu, J.; Dou, D.; Yan, C.; Zhang, Y. Latent diffusion enhanced rectangle transformer for hyperspectral image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 549–564. [CrossRef]
31. Li, N.; Wang, Z.; Cheikh, F.A. Discriminating spectral-spatial feature extraction for hyperspectral image classification: A review. *Sensors* **2024**, *24*, 2987:1–2987:20. [CrossRef]
32. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
33. Huang, S.; Xiao, W.; Chen, H.; Bejo, S.K.; Zhang, H. Hyperspectral image classification based on a locally enhanced transformer network. *IEEE Trans. Geosci. Remote Sens.* 2025, *to be published*.

34. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4340–4354. [CrossRef]
35. Alsadik, B.; Ellsäßer, F.; Awawdeh, M.; Al-Rawabdeh, A.; Almahasneh, L.; Elberink, S.; Abuhamoor, D.; Asmar, Y. Remote sensing technologies using UAVs for pest and disease monitoring: A review centered on date palm trees. *Remote Sens.* **2024**, *16*, 4371. [CrossRef]
36. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5966–5978. [CrossRef]
37. Aleissaee, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.S.; Khan, F.S. Transformers in Remote Sensing: A Survey. *Remote Sens.* **2023**, *15*, 1860. [CrossRef]
38. Wang, M.; Zhou, C.; Shi, J.; Lin, F.; Li, Y.; Hu, Y.; Zhang, X. Inversion of Water Quality Parameters from UAV Hyperspectral Data Based on Intelligent Algorithm Optimized Backpropagation Neural Networks of a Small Rural River. *Remote Sens.* **2025**, *17*, 119. [CrossRef]
39. Wu, Y.; Jiao, L.; Liu, X.; Liu, F.; Yang, S.; Li, L. Domain adaptation-aware transformer for hyperspectral object tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 8041–8052. [CrossRef]
40. Zhou, L.; Ma, L. Extreme learning machine-based heterogeneous domain adaptation for classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1781–1785. [CrossRef]
41. Tuia, D.; Persello, C.; Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57. [CrossRef]
42. Pokale, K.; Chaudhri, S.N. Transfer learning and domain adaptation in hyperspectral image processing: An overview. In Proceedings of the International Conference on Machine Learning and Autonomous Systems (ICMLAS), Chennai, India, 10–12 March 2025; pp. 1208–1213.
43. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked autoencoders are scalable vision learners. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 16000–16009.
44. Huang, L.; Chen, Y.; He, X. Spectral-spatial masked transformer with supervised and contrastive learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5508718. [CrossRef]
45. Shao, H.; Liu, C.; Xie, F.; Li, C.; Wang, J. Noise-sensitivity analysis and improvement of automatic retrieval of temperature and emissivity using spectral smoothness. *Remote Sens.* **2020**, *12*, 2295:1–2295:20. [CrossRef]
46. Gong, Z.; Zhou, X.; Yao, W. MultiScale spectral-spatial convolutional transformer for hyperspectral image classification. *IET Image Process.* **2024**, *18*, 4328–4340. [CrossRef]
47. Liu, H.; Li, W.; Xia, X.G.; Zhang, M.; Gao, C.Z.; Tao, R. Central attention network for hyperspectral imagery classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8989–9003. [CrossRef]
48. Liao, J.; Wang, L. HyperspectralMamba: A novel state space model architecture for hyperspectral image classification. *Remote Sens.* **2025**, *17*, 2577. [CrossRef]
49. Liu, H.; Li, W.; Xia, X.G.; Zhang, M.; Guo, Z.; Song, L. Seghsi: Semantic segmentation of hyperspectral images with limited labeled pixels. *IEEE Trans. Image Process.* **2024**, *33*, 6469–6482. [CrossRef]
50. Liu, X.; Ng, A.H.M.; Lei, F.; Ren, J.; Liao, X.; Ge, L. Hyperspectral image classification using a multi-scale CNN architecture with asymmetric convolutions from small to large kernels. *Remote Sens.* **2025**, *17*, 1461. [CrossRef]
51. Liu, L.; Chen, B.; Chen, H.; Zou, Z.; Shi, Z. Diverse hyperspectral remote sensing image synthesis with diffusion models. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 553261. [CrossRef]
52. Wang, Z.; Zhao, S.; Zhao, G.; Song, X. Dual-branch domain adaptation few-shot learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5506116. [CrossRef]
53. Song, J.; Meng, C.; Ermon, S. Denoising diffusion implicit models. *arXiv* **2020**, arXiv:2010.02502.
54. Hu, X.; Liu, X.; Duan, Q.; Hong, D.; Zhang, D. Diffusion Model in Hyperspectral Image Processing and Analysis: A Review. *arXiv* **2025**, arXiv:2505.11158. [CrossRef]
55. Ahmad, M.; Butt, M.H.F.; Mazzara, M.; Distefano, S.; Khan, A.M.; Altuwaijri, H.A. Pyramid hierarchical spatial-spectral transformer for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 17681–17689. [CrossRef]
56. Uddin, M.P.; Mamun, M.A.; Afjal, M.I.; Hossain, M.A. Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification. *Int. J. Remote Sens.* **2021**, *42*, 286–321. [CrossRef]
57. Ahmad, M.; Mazzara, M.; Distefano, S.; Khan, A.M.; Wu, X. Self-supervised spatial-spectral transformer with Extreme Learning Machine for Hyperspectral Image Classification. *Int. J. Remote Sens.* **2025**, *46*, 5384–5407. [CrossRef]
58. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
59. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

60. Sigger, N.; Vien, Q.T.; Nguyen, S.V.; Tozzi, G.; Nguyen, T.T. Unveiling the potential of diffusion model-based framework with transformer for hyperspectral image classification. *Sci. Rep.* **2024**, *14*, 8438. [CrossRef] [PubMed]
61. Ding, C. Diffusion-Augmented Cross-Domain Prototypical Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, S4804D. [CrossRef]
62. Ge, Z.Z.; Ding, Z.; Wang, Y.; Bian, L.F.; Yang, C. Spectral domain strategies for hyperspectral super-resolution: Transfer learning and channel enhance network. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *134*, 104180. [CrossRef]
63. Shi, Y.; Cui, H.; Yin, Y.; Song, H.; Li, Y.; Gamba, P. Transfer learning with nonlinear spectral synthesis for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5532517. [CrossRef]
64. Min, D.; Zhao, J.; Bodner, G.; Ali, M.; Li, F.; Zhang, X.; Rewald, B. Early decay detection in fruit by hyperspectral imaging—Principles and application potential. *Food Control* **2023**, *152*, 109830:1–109830:15. [CrossRef]
65. Wang, Y.; Liu, L.; Xiao, J.; Yu, D.; Tao, Y.; Zhang, W. MambaHSI+: Multidirectional State Propagation for Efficient Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4411414. [CrossRef]
66. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
67. Nichol, A.Q.; Dhariwal, P. Improved Denoising Diffusion Probabilistic Models. In Proceedings of the 38th International Conference on Machine Learning (ICML), Virtual Conference, 18–24 July 2021; PMLR, 2021; pp. 8162–8171.
68. Islam, M.R.; Ahmed, B.; Hossain, M.A.; Uddin, M.P. Mutual information-driven feature reduction for hyperspectral image classification. *Sensors* **2023**, *23*, 657:1–657:20. [CrossRef]
69. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2019.
70. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
71. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In Proceedings of the 3rd Annual JPL Airborne Geoscience Workshop, Pasadena, CA, USA, 1–5 June 1992; pp. 147–149.
72. He, Y.; Tu, B.; Liu, B.; Li, J.; Plaza, A. HSI-MFormer: Integrating Mamba and Transformer Experts for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5621916. [CrossRef]
73. Ram, B.G.; Oduor, P.; Igathinathane, C.; Howatt, K.; Sun, X. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Comput. Electron. Agric.* **2024**, *222*, 109037:1–109037:15. [CrossRef]
74. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: New York, NY, USA, 2013.
75. Zhang, Z.; Feng, H.; Zhang, C.; Ma, Q.; Li, Y. S^2DCN: Spectral–Spatial Difference Convolution Network for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3053–3068. [CrossRef]
76. Kucharczyk, M.; Hugenholtz, C.H. Remote sensing of natural hazard-related disasters with small drones: Global trends, biases, and research opportunities. *Remote Sens. Environ.* **2021**, *264*, 112577:1–112577:15. [CrossRef]
77. Zhang, J.; Sun, Z.; Wang, K.; Wang, C.; Cheng, S.; Jiang, Y.; Bai, Q. Prognosis prediction based on liver histopathological image via graph deep learning and transformer. *Appl. Soft Comput.* **2024**, *161*, 111653:1–111653:15. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.