Chapter 7

Conclusions and Future Research Directions

In this thesis, we present a systematic investigation into HSI classification. The investigation is done by leveraging DL architectures to integrate both spectral and spatial cues. Our research proposed a series of efficient and accurate hybrid DL architectures to address the intrinsic challenges posed by high dimensionality, spectral redundancy, class imbalance, and limited annotated data. Each framework has been designed with a dual objective: maximizing classification performance while minimizing computational overhead—key for deployment on real-time, resource-constrained platforms such as UAVs and edge devices. The final chapter summarizes key contributions and results, and suggests future research directions based on current limitations.

7.1 Contributions

Our motivation is derived from the growing importance of HSI data in precision agriculture, land-use monitoring, and environmental analysis, as discussed in Chapter 1. The rich spectral bands enable fine land-cover distinction, but high dimensionality and redundancy remain major challenges. Moreover, limited annotated samples, varying spatial resolution, and class imbalance further hinder accurate classification. Section 1.7 outlines these challenges and emphasizes the need for models that efficiently extract spectral and spatial features. The thesis contributions are discussed in Section 1.11. Chapter 2 reviews existing models based on DL feature types: spectral-only, spatial-only, and spectral-spatial. These features are detailed in Section 2.5. Spectral-only models capture rich spectral cues but miss the spatial context. Spatial-only models, like 2D CNNs, often ignore inter-band spectral dependencies. Hybrid models address these gaps by combining both domains, but many face high computational costs or poor generalization. Our review of CNNs, RNNs, Transformers, and graph-based models reveals a need for accurate, generalizable, and resource-efficient solutions. Section 2.6 further categorizes models by feature context: local (e.g., CNNs), global (e.g., Transformers), and hybrid (e.g., CNN-Transformer). Our analysis of architectures like SpectralFormer, IFormer, and LSGA-ViT shows that although they improve accuracy, their large parameter sizes and slow inference limit use on resource-constrained platforms like UAVs and edge devices.

Chapter 3 presents our MDCNN, a hybrid model that integrates 3D and 2D CNNs with morphological operations and dilated convolutions to capture multi-scale spectral-spatial features (Section 3.4.3). MDCNN uses PCA for dimensionality reduction and morphological filters (Section 3.4.2) to reduce spectral redundancy and enhance spatial representation. It is a hybrid architecture that balances accuracy with low computational cost. Experiments (Section 3.5) on IP, PU, and SA datasets show superior performance in OA, AA, and κ, outperforming CNNs and models like HybridSN, SSRN, and R-3D-CNN. For example, MDCNN achieved 99.80% OA on IP and 100% OA on SA (Section 3.5.4). Additionally, MDCNN reduces training and testing time (Figure 3.8, making it efficient for real-time applications. However, its focus on local features limits global context modeling. Future work will explore transformer integration and model compression for scalable HSI classification. Chapter 4 introduces LogGroupFormer, a lightweight DL framework combining CNN and ViT for efficient HSI classification (Section 4.4). Innovations like LGConv, SPCE, and $(SC)^2 PosEmbed$ enable local-global feature extraction while reducing the computational cost. Experiments (Section 4.5.5) show state-of-the-art accuracy: 98.12% on IP, 99.47% on PU, 99.23% on HU, and 99.85% on LK, outperforming SOTA models, especially with limited data. Theoretical analysis (Section 4.5.6, Eq. 4.2) shows LGConv3D reduces overhead by 65.63%, with no parameter increase ($\Delta P = 0$). LGConv2D also cuts parameters and FLOPs by 65.63% compared to 2D convolution (Table 4.1). Empirical results confirm efficiency, requiring only 192.34K parameters and 0.84 G-Flops, with the shortest training (13.60 s) and inference (0.93 s) times on IP. Ablation and sensitivity studies confirm the importance of each module. Future work will explore linear attention and adaptive patching to address scalability challenges in ViT.

Chapter 5 introduces CKGFLNet, a hybrid framework combining 3D-2D CNNs with a KGFLA transformer to improve HSI classification efficiency (Section 5.4). Replacing standard MHSA with a Flatten linear-complexity mechanism, CKGFLNet achieves SOTA accuracy while reducing training and inference times by 2.08× and 4.76×, respectively, compared to SSFTT and LGSA-ViT (Section 5.5.4). Experiments (Section 5.5.3) show superior performance: 98.60% OA on IP, 97.44% on PU, and 98.34% on SA, surpassing CNN- and transformer-based models. The complexity analysis (Section 5.5.4) shows CKGFLNet requires only 15.10 s training time and 0.39 s inference on IP, with 192.34K parameters, making it suitable for real-time applications. KGFLA module reduces computational overhead from $\mathcal{O}(N^2d)$ to $\mathcal{O}(Nd^2)$, preserving expressive power via Kaiming-Gaussian initialization. However, CKGFLNet has limitations, including a higher parameter count than lightweight CNNs and underperformance in minority classes due to class imbalance.

Chapter 6 presents HieraKGTNet, a lightweight and robust framework that directly tackles the dual challenges of class imbalance and computational complexity in HSI classification. Previous models face specific limitations—MDCNN is computationally expensive, LogGroupFormer suffers from quadratic attention complexity, and CKGFLNet does not explicitly address class imbalance. HieraKGTNet overcomes these issues by integrating targeted modules and learning strategies to achieve balanced, efficient, and accurate classification across all classes. The architecture incorporates LGASS for efficient superpixel-based local-global feature extraction, HCBConv for hierarchical spectral-spatial learning, and FKGT for scalable global context modeling. Kaiming Semantic Tokenizer further strengthens the encoding of discriminative spectral features. A key innovation, MPF-Loss, adaptively emphasizes minority and hard-to-classify samples, enhancing convergence and fairness (Section 6.4.5). Unlike traditional cross-entropy or focal loss, MPF-Loss is specifically designed for multiclass HSI data. It ensures stable and unbiased learning across both majority and minority classes. Extensive experiments and computational analysis (Section 6.5) confirm the effectiveness and efficiency of **HieraKGTNet**. On the IP dataset, it achieves 34.04 s training time, 0.34 s inference time, 247.5K parameters, and 3.43 GFLOPs. For PU, it records 28.18 s training, 1.34 s inference, 247.1K parameters, and 3.10 GFLOPs. On HU, it reaches 36.38 s training, 0.48 s inference, 247.5K parameters, and 3.43 GFLOPs. For LK, the model maintains efficiency with 78.46 s training, 5.16 s inference, 243.3K parameters, and 1.91 GFLOPs. Across all datasets, **HieraKGTNet** consistently outperforms heavier models like DBDA, MCNN-CP, and IFormer, which exceed 2 million parameters, as well as lighter models like LogGroupFormer and CKGFLNet, which lack explicit handling of class imbalance.

In terms of classification accuracy (Section 6.4), **HieraKGTNet** achieves SOTA results: 98.60% OA on IP, 97.44% on PU, 99.23% on HU, and 99.85% on LK. Notably, it improves minority class accuracy by 5-12% over models like SSFTT and LGSA-ViT, especially for rare classes such as "Oats" in IP and "Lettuce-6wk" in SA. The **MPF-**

Loss function contributes to a 4.76% gain in average accuracy for underrepresented classes, reinforcing its effectiveness in achieving balanced and robust classification.

Compared to **MDCNN**, which is computationally heavy, and **LogGroupFormer**, which suffers from quadratic attention complexity, **HieraKGTNet** provides a lightweight, scalable, and fair solution. Unlike **CKGFLNet**, which focuses on linear attention but does not address class imbalance, **HieraKGTNet**'s **MPF-Loss** ensures balanced learning across all classes. This makes it particularly well-suited for real-world, imbalanced HSI classification scenarios.

7.2 Limitations

Despite the significant advancements of the proposed models, several limitations remain that provide scope for further improvements in HSI classification. These limitations highlight areas where the current approaches can be refined to achieve greater robustness, scalability, and applicability in real-world scenarios.

Scalability: While the proposed models perform effectively on widely used benchmark datasets, they face challenges when scaling to very large HSI datasets with higher spatial and spectral resolutions. The quadratic attention mechanisms and high-dimensional convolutional layers limit real-time analysis for large-scale or satellite-wide applications.

Class Imbalance: Class imbalance continues to pose a challenge, particularly for underrepresented land-cover classes. Although the HieraKGTNet framework partially addresses this issue using the proposed MPF-Loss, highly imbalanced datasets still degrade per-class performance, especially for minority categories.

Multi-modal Fusion: The models developed in this thesis focus exclusively on HSI data. However, real-world applications often require integrating HSI with other modalities such as LiDAR, SAR, or MSI data. The lack of multi-modal fusion currently limits robustness in complex environments.

Edge Deployment: Achieving efficient inference on edge devices such as UAVs, nanosatellites, or field-deployable mobile platforms remains a challenge. The current models, despite being optimized, still demand considerable computational resources, which restricts real-time deployment in resource-constrained devices.

Additionally, investigating domain adaptation for cross-sensor analysis and extending models to semi- and self-supervised frameworks are promising research avenues for improving generalization and reducing reliance on annotated data.

7.3 Future Research Directions

Building on the limitations discussed above, several promising research directions emerge that can enhance the scalability, generalization, and real-world applicability of HSI classification frameworks.

Linear-Time Attention and Scalable Architectures: Future work can explore Mamba-based or other linear-time attention mechanisms to handle large-scale hyperspectral cubes efficiently, enabling real-time analysis for high-resolution satellite data.

Advanced Solutions for Class Imbalance: Techniques such as generative adversarial networks (GAN)-based augmentation, adaptive focal loss, or curriculum learning strategies may improve recognition of underrepresented classes. Semi-supervised and self-supervised learning can also reduce reliance on large amounts of annotated data.

Multi-modal Data Fusion: Developing architectures that integrate HSI with complementary modalities such as LiDAR or SAR can provide richer contextual information and improve robustness in challenging environments like forests, coastal zones, and urban areas.

Edge and Low-power Deployment: Incorporating techniques such as pruning, quantization, and knowledge distillation can make HSI classification models more lightweight and adaptable for deployment on UAVs, nano-satellites, and portable field sensors.

Our thesis not only proposes innovative DL frameworks for HSI classification but also lays a strong foundation for future research to address current limitations, enhance scalability, and enable practical deployment in real-world applications.