ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

Multi–Scale Graph Wavelet Convolutional Network for Hyperspectral and LiDAR Data Classification

Junhua Ku[1,2,a,*], Jie Zhao[3,b]

[1]*School of Information Science and Technology, Qiongtai Normal University, Haikou, Hainan, 571127, China*
[2]*Institute of Educational Big Data and Artificial Intelligence, Qiongtai Normal University, Haikou, Hainan, China*
[3]*School of Science, Qiongtai Normal University, Haikou, Hainan, 571127, China*
[a]*junhuacoge@mail.qtnu.edu.cn,* [b]*zhaojie@mail.qtnu.edu.cn*
**Corresponding author*

Abstract: *In this paper, we present a novel Multi–Scale Graph Wavelet Convolutional Network for Hyperspectral and LiDAR Data Classification. The proposed MS-GWCN enables more effective learning of spatial–spectral relationships for pixel-wise classification. We conduct extensive experiments on the Houston 2013 dataset, which comprises 15 diverse land-cover classes. The results demonstrate that our method achieves significant improvements over baseline approaches, attaining an overall accuracy (OA) of 85.98%, an average accuracy (AA) of 88.39%, and a Kappa coefficient of 84.79%.*

Keywords: *Multi-Scale Graph Convolutional Network, Deep Learning, Hyperspectral and Lidar Data Classification*

1. Introduction

Remote sensing image classification is essential for numerous applications, including urban planning, land use monitoring, and environmental management. Traditional approaches primarily employ pixel-based classification techniques, which often inadequately represent complex spatial relationships and feature dependencies. The advent of multisource data, such as hyperspectral imagery (HSI) and LiDAR, presents opportunities to enhance classification accuracy through the integration of more comprehensive feature sets. The combination of HSI and LiDAR data has garnered considerable interest over the past five years, owing to the complementary information each provides. HSI supplies extensive spectral information across multiple bands, which is instrumental in material identification and land cover classification. Conversely, LiDAR offers detailed three-dimensional structural data, including elevation and surface characteristics. The fusion of these two data modalities has consistently demonstrated significant improvements in classification accuracy across various remote sensing applications, including urban planning, environmental monitoring, and vegetation analysis mapping[1, 2]. Below, we discuss several key methods employed for HSI+LiDAR fusion over recent years, emphasizing their benefits, limitations, and illustrating how dynamic graph-based approaches can address some of the existing challenges.

In early fusion methodologies, HSI and LiDAR data are integrated at the feature level prior to classification. Features derived from both data sources are concatenated to form a singular vector, which is subsequently input into machine learning classifiers such as Support Vector Machines (SVMs), Random Forests (RF), and k-Nearest Neighbors (k-NN). The primary advantage of early fusion lies in its capacity to directly amalgamate spatial and spectral information into a cohesive feature space, thereby simplifying the application of conventional classifiers. Nevertheless, a significant limitation of this approach is the potential loss of spatial dependencies among neighboring pixels or regions, potentially impairing classification accuracy. Additionally, this method is susceptible to the curse of dimensionality, as the combination of high-dimensional features from both sources may produce an excessively large dataset, particularly when utilizing high-resolution hyperspectral data. [3,4].

Conversely, late fusion approaches perform classification separately for HSI and LiDAR data, subsequently integrating the outputs. For instance, classifiers may be trained independently on HSI and LiDAR datasets, and the predictions are then combined through decision-level fusion techniques such as voting schemes or weighted averaging. Although late fusion methods circumvent the dimensionality

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

challenges associated with early fusion, they are inadequate in capturing the spatial relationships inherent in both data types. This limitation is particularly pronounced in complex scenes where spatial and spectral dependencies are vital for precise classification. Moreover, late fusion often overlooks potential synergies between HSI and LiDAR data, as each classifier processes the data independently[5,6].

Intermediate fusion methods, also known as feature-level fusion, aim to combine the advantages of both early and late fusion. In this approach, feature extraction from HSI and LiDAR data is performed separately, and the resulting feature representations are then fused. For instance, Principal Component Analysis (PCA) can be applied to reduce the dimensionality of HSI data, while LiDAR features such as elevation or point density can be extracted and integrated. This fused feature set is then used as input to a classifier. Although intermediate fusion can improve classification accuracy by better capturing the interaction between spectral and spatial information, it still struggles with preserving complex spatial relationships, which is critical for accurate classification in heterogeneous environments. The effectiveness of this approach depends on how well the features are fused, and often, important spatial correlations may still be missed[7].

Graph-based methodologies have recently gained prominence as a sophisticated approach for the fusion of HSI and LiDAR data. These methodologies depict data using graphs, where nodes denote regions such as superpixels or segments, and edges illustrate spatial or spectral relationships among these regions. This framework enables the explicit capture of both local and global spatial dependencies. Typically, Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) are employed to facilitate information propagation across the graph structure, thereby supporting the integration of spectral and spatial information [8-10].

In this study, we introduce the Multi-Scale Graph Wavelet Convolutional Network (MS-GWCN) architecture and extend it for the joint classification of HSI and LiDAR data. A graph-based model is constructed where each node represents a labeled pixel characterized by a combined feature vector containing HSI spectral bands and the corresponding LiDAR elevation value. Spatial relationships are encoded via edges that connect each node to its neighboring pixels within a fixed window (e.g., a 5×5 patch), capturing local context within the graph structure. The MS-GWCN model processes these features through a sequence of specialized graph convolution layers. It incorporates three Graph Wavelet Convolution layers operating in a multi-scale manner: each layer broadens the node's feature representation across multiple scales and applies a learned linear projection with residual connections and layer normalization. This approach simulates graph wavelet filtering to extract spectral–spatial features at various scales. Subsequently, two standard Graph Convolutional Network (GCN) layers propagate and aggregate information over the graph, thereby refining each node's features based on its neighbors. The original MS-GWCN is marginally modified by integrating an attention mechanism that learns to assign weights to each node's feature vector. Specifically, an attention module generates a scalar weight for each node (via a small neural network and softmax activation across all nodes), which scales the node features to emphasize more informative regions within the graph. The resulting node representations are finally inputted into a fully connected classifier head (incorporating layer normalization, ReLU activation, and dropout) to predict the land-cover class of each pixel. This architecture maintains the core multi-scale wavelet convolution principle of MS-GWCN while augmenting it with an attention-based feature weighting mechanism and minor adjustments to layer dimensions to enhance performance stability.

For data processing, all HSI bands and the LiDAR channel are standardized and concatenated to ensure that the multi- modal features of each pixel are on a comparable scale. Consequently, the graph node features integrate both spectral signatures and elevation information. The provided ground truth is utilized to mask out unlabeled pixels and to partition the labeled nodes into training and test sets, utilizing the specified training/test masks. The model training employs a cross- entropy loss function with label smoothing ($\varepsilon = 0.1.\ 1$), which gently penalizes the correct class to enhance generalization. Optimization is performed using the AdamW optimizer with a learning rate of 0.003, along with a One- Cycle learning rate schedule spanning 400 epochs to facilitate faster convergence. To further stabilize training on GPU hardware, mixed precision (autocast) and gradient scaling techniques are employed. The training procedure is reiterated over five independent runs, each with a different random seed initialization, to ensure the statistical robustness of the results. After each run, the model' s performance is evaluated on the test set, recording key metrics. In the evaluation phase, overall accuracy (OA), Averal accuracy (AA), which represents the average per-class accuracy, and Cohen' s kappa coefficient, a reliability measure, are computed. Additionally, the confusion matrix is derived to analyze class- specific performance, and per- class accuracy rates are calculated to identify classes with weaker performance. Finally, predicted classification maps are generated for each run by mapping the model' s node predictions back to their

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

spatial locations, enabling a visual comparison between the model output and the ground truth map. Across multiple runs on the HSI- LiDAR dataset, specifically the 15- class Houston University scene, our reproduced MS- GWCN model, augmented with the proposed modifications, demonstrates consistent accuracy and enhanced attention to salient features, thereby validating the effectiveness of the approach. Our contributions can be summarized as follows:

(1) We implement the multi-scale graph wavelet convolutional network architecture for HSI and Light Detection and Ranging (LiDAR) data within a graph-based framework, faithfully reproducing the core spectral-spatial feature extraction and graph convolution components of the original model. We combine hyperspectral and lidar modalities at the feature level, constructing a unified graph that enables the model to learn from both spectral signatures and elevation information simultaneously for enhanced land-cover classification.

(2) We introduce a learned attention mechanism that weights node features adaptively, focusing the model on more informative pixels. Additionally, we simplify and streamline the graph wavelet convolution layers by employing a multi-scale feature replication and linear projection approach, which preserves the multi-scale representation capability while reducing complexity.

(3) We apply label smoothing in the loss function to enhance generalization and deploy an adamw optimizer with a One-Cycle learning rate scheduler to ensure efficient convergence. We also utilize modern training techniques such as mixed-precision acceleration and rigorous seed control to guarantee stable and reproducible training across runs.

2. Methodology

2.1 Data Preprocessing and Graph Construction

Let $\mathbf{H} \in \mathbb{R}^{H\times W\times B}$ denote the hyperspectral image with B spectral bands, and $\mathbf{L} \in \mathbb{R}^{H\times W}$ denote the LiDAR height map. To combine the two modalities, we first standardize each band using a z-score transform. Given an input band $f \in \mathbb{R}^{HW}$, its standardized version is

$$\hat{f} = \frac{f-\mu}{\sigma} \tag{1}$$

Where μ and σ are the empirical mean and standard deviation of f. Standardization ensures that each band has zero mean and unit variance, which aids in numerical stability during training. After scaling all hyperspectral bands and the LiDAR map, we form a data cube $\mathbf{C} \in \mathbb{R}^{H\times W\times(B+1)}$ by concatenating them along the spectral dimension.

The ground-truth label map $\mathbf{Y} \in 0,\ldots,c^{H\times W}$ contains c classes; pixels with $Y_{ij}=0$ are unlabelled and are ignored during training and evaluation. A binary mask $M \in \{0,1\}^{H\times W}$ is defined as $M_{ij}=1$ if $Y_{ij}>0$ and 0 otherwise. We flatten the valid pixels into a set of N nodes by collecting indices $\left\{(i_k,j_k)\middle|1\le k\le N, M_{i_k,j_k}=1\right\}$. The corresponding feature matrix $X \in \mathbb{R}^{N\times(B+1)}$ is obtained by reshaping $\mathbf{C}$ and selecting valid rows, while the label vector $y \in \{0,\ldots,c-1\}^N$ stores the class labels minus one.

Each valid pixel is treated as a node in an undirected graph $G=(V,E)$. Edges are created between a node and its neighbors within a Manhattan distance of two in the image plane. Formally, for a node corresponding to pixel (i,j), we connect it to all nodes at positions $(i+\Delta i, j+\Delta j)$ where $\Delta i,\Delta j \in \{-2,-1,0,1,2\}$ except (0,0). The resulting edge list E is encoded into a two-row tensor $\text{edge\_index} \in \mathbb{Z}^{2\times|E|}$, where each column ($u$,$v$) represents an edge from node u to node v. Such a local connectivity pattern captures spatial relationships and yields a sparse adjacency matrix suitable for graph convolutions.

2.2 Model Architecture

The modified MS-GWCN consists of three graph wavelet convolution layers, two GCN layers, an

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

attention module and a final classifier. Let $x^{(0)} = X$ denote the input node features. The model generates intermediate representations $x^{(1)},\ldots,x^{(5)}$ as follows. Graph wavelet convolution layers extend classical wavelet transforms to graph-structured data by applying spectral kernels at multiple scales. In our implementation, the first three layers each apply *s* scales and a shared linear transformation. Given input features $x \in \mathbb{R}^{N\times d}$ and the number of scales *s*, we replicate *x* *s* times to obtain $[x,x,\ldots,x] \in \mathbb{R}^{N\times(sd)}$. A linear projection $W \in \mathbb{R}^{sd\times d'}$ is applied followed by layer normalization and a residual connection:

$$\mathrm{GWConv}(x) = \mathrm{ReLU}\left(\mathrm{LayerNorm}(xW) + x_{[1:d']}\right) \tag{2}$$

Where $x_{[1:d']}$ denotes the first \$d'\$ features from the input after repetition. The number of output channels *d'* increases from 128 to 256 in successive layers. Although the implementation does not explicitly perform spectral wavelet transforms, repeating the input across scales allows the network to capture multi-scale information analogous to wavelet filtering.

After the wavelet layers, two standard graph convolutional layers aggregate information from neighboring nodes according to the GCN formula above. In our case, both layers map 256-dimensional features to 256-dimensional embeddings and use the rectified linear unit (ReLU) as activation. To emphasize informative nodes and suppress noisy ones, the network applies a learnable attention mechanism. Given node features $z \in \mathbb{R}^{N\times d}$, an attention score α_i for each node *i* is computed as

$$\alpha_i = \frac{\exp\left(w_a^T \tanh(W_a z_i)\right)}{\sum_{j=1}^{N} \exp\left(w_a^T \tanh(W_a z_j)\right)} \tag{3}$$

Where $W_a \in \mathbb{R}^{d\times d}$ and $w_a \in \mathbb{R}^d$ are trainable parameters. The normalized scores satisfy $\sum_i \alpha_i = 1$ and are used to weight the node features: $z_{i'} = \alpha_i z_i$. This operation produces attended features $z' \in \mathbb{R}^{N\times d}$. The attended features are fed into a small fully-connected network comprising layer normalization, a linear layer mapping *d* to 128 dimensions, a ReLU activation, dropout with rate 0.5 and a final linear layer producing logits for *c* classes. The model outputs a matrix of unnormalized scores $o \in \mathbb{R}^{N\times c}$.

2.3 Training Objective

We train the network using cross-entropy loss with label smoothing. Traditional one-hot encoding assigns the true class *k* a probability of 1 and all other classes 0. Label smoothing replaces the hard target vector $\mathbf{y}$ with a convex combination of the one-hot distribution and a uniform distribution over classes. For a sample with label *k* and *c* classes, the smoothed target $\tilde{y}$ has components

$$\tilde{y}_i = \begin{cases} 1-\epsilon+\dfrac{\epsilon}{c}, & \text{if } i = k, \\ \dfrac{\epsilon}{c}, & \text{otherwise.} \end{cases} \tag{4}$$

Where $\epsilon \in [0,1]$ controls the amount of smoothing. Label smoothing regularizes the classifier by preventing it from becoming overly confident. The resulting loss for a minibatch with predictions *o* and smoothed targets $\tilde{Y}$ is

$$\mathcal{L}_{\mathrm{LS}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{y}_{ij} \log \frac{\exp(o_{ij})}{\sum_{m=1}^{c}\exp(o_{im})} \tag{5}$$

We use the AdamW optimizer with weight decay 1×10^{-4} and an initial learning rate of 3×10^{-3}. A one-cycle learning-rate scheduler adjusts the learning rate following the 1cycle policy: it increases the learning rate from a low value to a maximum and then decreases it below the initial value. This policy has been shown to accelerate convergence and improve generalization. We train for 400 epochs and use mixed-precision training with gradient scaling.

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

2.4 Experimental Setup

We report three evaluation metrics: overall accuracy (OA), average accuracy (AA) and Cohen's kappa coefficient. OA is the proportion of correctly classified pixels:

$$\mathrm{OA} = \frac{\sum_{i=1}^{c} n_{ii}}{\sum_{i=1}^{c}\sum_{j=1}^{c} n_{ij}} \tag{6}$$

Where n_{ij} is the element of the confusion matrix indicating the number of pixels with true class i predicted as class j. AA is the mean of per-class accuracies:

$$\mathrm{AA} = \frac{1}{c}\sum_{i=1}^{c}\frac{n_{ii}}{\sum_{j=1}^{c} n_{ij}} \tag{7}$$

Cohen's kappa compares the observed accuracy with the expected accuracy under random chance:

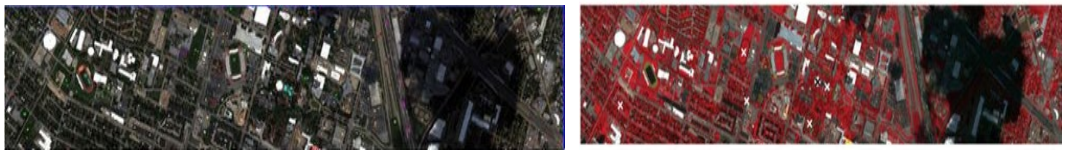
$$\kappa = \frac{\mathrm{OA} - \mathrm{EA}}{1 - \mathrm{EA}} \tag{8}$$

where $\mathrm{EA} = \sum_{i=1}^{c}(n_{i+}n_{+i}) / (\sum_{ij} n_{ij})^2$ and n_{i+} and n_{+i} are row and column sums. The kappa statistic evaluates agreement beyond chance; it is less misleading than OA alone and facilitates comparison across classifiers. We conduct *5* independent runs with different random seeds to account for stochasticity. In each run the network is trained for 400 epochs using the training mask. After training, predictions are computed for all nodes, and metrics are calculated on the test mask. We report the mean and standard deviation across runs for OA, AA and κ, as well as per-class accuracies.

3. Results

3.1 Dataset background

The Houston 2013 dataset was captured by the ITRES CASI-1500 airborne sensor over the University of Houston campus and adjacent rural areas in Texas in the year 2013. After excluding noisy bands, the dataset comprises 144 valid spectral bands. The entire scene encompasses 349 × 1905 pixels with a spatial resolution of 2.5 m per pixel. It includes 15 land-cover classes, namely: healthy grass, stressed grass, synthetic grass, tree, soil, water, residential, commercial, road, highway, railway, two parking lot categories, tennis court, and running track. The pseudo-color map and grayscale image for the LiDAR data are depicted in Figure 1. Provided by IEEE, these data are accessible online via the website: http://dase.grss-ieee.org/.



(a) (*b*)

Figure 1 Visualization of Houston 2013. (a) Pseudo-color image for HSI data. (b) Grayscale image for the LiDAR data.

3.2 Evaluation metrics

Performance was assessed using Overall Accuracy (OA), Average Accuracy (AA) and the Kappa coefficient. OA measures the ratio of correctly classified samples to all samples; AA is the mean accuracy across classes, highlighting balance among classes; the Kappa coefficient evaluates the agreement between the predicted map and the reference ground truth.

(1) Quantitative results

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

Experimental evaluation on the houston-2013 dataset with a dynamic graph convolutional network (ms-gwcn) achieved the per-class accuracies summarised below. Each value is shown with its standard deviation. These results indicate that the model performs very well across most classes, with perfect classification on several categories.

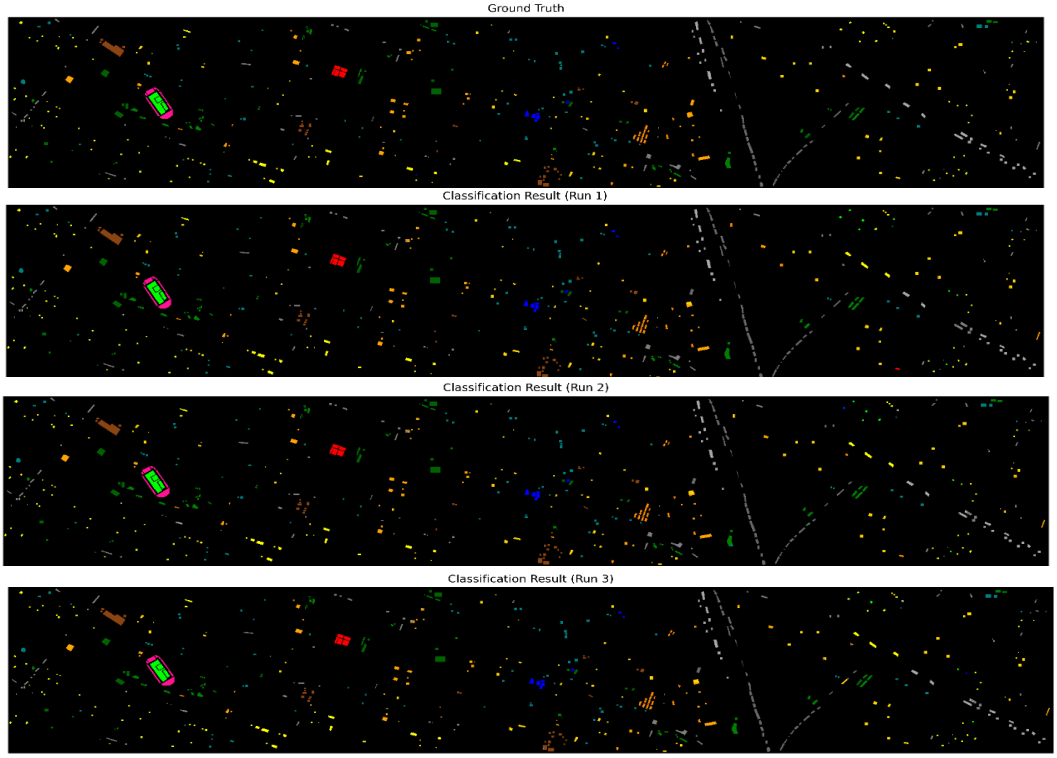
Table 1 Accuracy (%) of the MS-GWCN on the Houston 2013 dataset

Classes	Classes-names	Accuracy (%)	Classes	Classes-names	Accuracy (%)
1	Healthy grass	82.94 ± 0.18	9	Road	86.19 ± 2.02
2	Stressed grass	95.58 ± 1.60	10	Highway	46.95 ± 4.46
3	Synthetic grass	99.37 ± 1.27	11	Railway	82.85 ± 1.98
4	Tree	100.00 ± 0.00	12	Parking lot 1	85.09 ± 2.86
5	Soil	100.00 ± 0.00	13	Parking lot 2	90.18 ± 2.60
6	Water	95.80 ± 0.00	14	Tennis court	100.00 ± 0.00
7	Residential	81.04 ±5.56	15	Running track	100.00 ± 0.00
8	Commercial	79.92 ± 3.72			
	OA (%)		85.98 ± 0.71		
	AA (%)		88.39 ± 0.47		
	Kappa		84.79 ± 0.77		

Table 1 presents OA = 85.98 ± 0.71%, AA = 88.39 ± 0.47%, and Kappa = 84.79 ± 0.77. The model achieved excellent or near-perfect classification for tree, soil, tennis court, and running track, indicative of distinct spectral signatures. The most challenging class was the highway, with significantly lower accuracy at 46.95 ± 4.46%, likely attributable to spectral similarity with the road and parking lot classes. High accuracies observed in other classes (e.g., synthetic grass and water) suggest that the model effectively leverages both spectral and spatial information.

(2) Qualitative results

Visual inspection of the classification maps reveals notable advantages. Compared with classic convolutional networks and machine-learning baselines, the MS-GWCN-based approach produces a smoother and cleaner classification map, with fewer misclassified pixels and sharper boundaries. In particular, the attention-based fusion of multi-source features enables the network to capture both coarse and fine patterns; this yields precise delineation of small or irregular patches and robust discrimination of spectrally similar classes. Qualitatively, the improvements are especially evident in densely built-up areas and at class boundaries, where baseline models tend to produce noisy labels.



ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

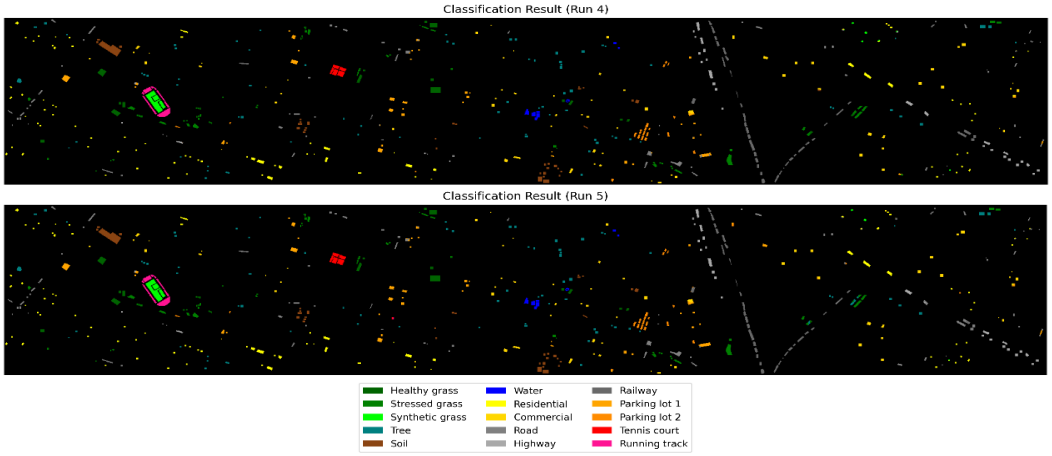


Figure 2 Classification maps on the Houston 2013 dataset across five independent runs.

As demonstrated in Figure 2, the performance metrics of the model consistently maintain high standards across five separate trials. The left panel of this chart depicts Overall Accuracy (OA), Average Accuracy (AA), and Kappa for each trial, illustrating only minor variations over repeated assessments. For instance, OA varies approximately between 85% and 86.7% across the five trials, consistently remaining well above 85%. Similarly, AA remains around 88% in all trials, with a narrow range of approximately 1 percentage point between its lowest and highest values. Furthermore, the Kappa coefficient persistently lies within the mid-84% to 85% range throughout. These minimal differences indicate that the model demonstrates stable performance without any significant decline or increase in any individual trial.

Notably, the fluctuations in these metrics are subtle. Overall Accuracy (OA) attains its peak value in a single iteration (approximately 86.7%), with only minor decreases in subsequent iterations; the disparity between the highest and lowest OA across different runs is approximately 1.4 percentage points. Likewise, the maximum and minimum values of Average Accuracy (AA) across iterations differ by roughly 1.0 point, and Cohen's Kappa varies by about 1.5 points from its maximum to minimum. These minimal variations underscore the reliability of the training process—each iteration yields nearly identical results, demonstrating that random initializations or training set partitions have minimal influence on the overall metrics. The consistency of OA, AA, and Kappa across multiple runs further reinforces confidence in the robustness and reproducibility of the model's performance.

The right panel of Figure 2 delineates the accuracy by class for each of the fifteen land-cover classes, providing insights into which classes are classified consistently and which exhibit fluctuations. Certain classes demonstrate remarkably stable accuracy across multiple runs (near-zero variance), whereas others display noticeable variability. Notably, the classes of Tree, Soil, Tennis Court, Running Track, and Water are distinguished by their stability. These classes consistently achieve nearly identical accuracy in every iteration—specifically, Tree, Soil, Tennis Court, and Running Track each attain 100.00% accuracy in all five runs (standard deviation ±0.00), while Water maintains a consistent accuracy of 95.80% ±0.00. Healthy Grass is another class characterized by high stability, with an accuracy around 82.94% in every run (standard deviation ±0.18). The minimal variance observed indicates that the model reliably recognizes these classes regardless of initialization or data splits; it is likely that they are easier to classify due to distinctive spectral signatures or ample training samples.

In contrast, several other classes exhibit greater variability across different runs. Residential areas are notably the most unstable, with an average accuracy of approximately 81.04% (±5.56%). This indicates that in certain runs, the Residential accuracy declined into the mid-70s, whereas in others, it reached the high 80s. The Highway class also demonstrates significant fluctuations, averaging only about 46.95% (±4.46%), which is particularly concerning given its low mean; some runs fell below 45% while others exceeded 50%. The accuracy of the Commercial class similarly varies, averaging around 79.92% (±3.72%), and even Parking Lot 1 (85.09% ± 2.86%) and Parking Lot 2 (90.18% ± 2.60%) experience noticeable alternations between runs.

The results suggest that, for these specific categories, the performance of the model may depend on the particular training instance or data split, indicating less reliable classification for those classes. Factors such as class imbalance or spectral similarity with other classes could be contributing to the model's difficulty in consistently classifying these classes. Further investigation or targeted enhancements (e.g.,

ISSN 2616-5775 Vol. 8, Issue 8: 51-58, DOI: 10.25236/AJCIS.2025.080808

additional training data or class-specific model tuning) may be required to stabilize performance across these categories. Most stable classes (lowest variance) include: Tree (100.00% ± 0.00), Soil (100.00% ± 0.00), Water (95.80% ± 0.00), Tennis Court (100.00% ± 0.00), Running Track (100.00% ± 0.00). All five exhibited identical accuracy in every run with zero variation; Healthy Grass was also highly stable at 82.94% ± 0.18. Least stable classes (highest variance) include: Residential (81.04% ± 5.56), Highway (46.95% ± 4.46), Commercial (79.92% ± 3.72). These classes exhibited the largest variances in accuracy across runs.

4. Conclusions

Our experimental results on the Houston dataset demonstrate that the proposed MS-GWCN model offers robust and consistent performance in multi-source land-cover classification tasks. The model achieves high overall and average accuracies, as well as a strong Kappa coefficient, underscoring the effectiveness of fusing HSI and LiDAR data via dynamic graph construction. Across five independent runs, key metrics—including OA, AA, and Kappa—remain remarkably stable, with only minor fluctuations attributed to random initialization or data splits. The model consistently delivers excellent accuracy for certain classes such as Tree, Soil, Tennis Court, Running Track, and Water, each achieving virtually identical results in every trial. However, the analysis also reveals that some categories, particularly Residential, Highway, and Commercial, exhibit higher variability and lower mean accuracy, suggesting that these classes are more challenging to distinguish—possibly due to spectral similarity, class imbalance, or limited training samples. Overall, the multi-run analysis validates both the robustness and repeatability of the MS-GWCN approach, while also highlighting opportunities for further enhancement. Future research may focus on addressing the less stable classes through improved sampling, data augmentation, or class-specific model refinements, aiming to achieve even greater consistency and performance across all land-cover types.

Acknowledgements

This work was supported by the Hainan Provincial Natural Science Foundation of China under Grant No. 621RC599.

References

[1] Li, H., Wang, F., & Zhang, X. (2020). A comprehensive review of hyperspectral and LiDAR data fusion techniques for remote sensing applications. Remote Sensing, 12(5), 1053.
[2] Yang, J., Wang, J., Sui, C. H., Long, Z., & Zhou, J. (2024). HSLiNets: Hyperspectral Image and LiDAR Data Fusion Using Efficient Dual Non-Linear Feature Learning Networks. arXiv Preprint.
[3] Zhao, B., & Wu, T. (2019). Early and late fusion methods for hyperspectral and LiDAR data classification: A comparative study. Remote Sensing of Environment, 231, 111305.
[4] Wang, F., Du, X., Zhang, W., Nie, L., Wang, H., Zhou, S., & Ma, J. (2024). Remote Sensing LiDAR and Hyperspectral Classification with Multi-Scale Graph Encoder–Decoder Network. Remote Sensing, 16(20), 3912.
[5] Chang, Y., & Liu, Y. (2018). Dimensionality reduction and feature fusion for hyperspectral and LiDAR data classification. International Journal of Applied Earth Observation and Geoinformation, 70, 111-121.
[6] Yang, J., Zhou, J., Wang, J., Tian, H., & Liew, A. W.-C. (2024). LiDAR-Guided Cross-Attention Fusion for Hyperspectral Band Selection and Image Classification. arXiv Preprint.
[7] Zhang, L., & Shi, W. (2021). Graph convolutional networks for remote sensing image classification: A survey. Remote Sensing, 13(4), 1-21.
[8] Wang, L., & Li, J. (2022). Dynamic graph convolutional network for hyperspectral and LiDAR data fusion. Sensors, 22(9), 3245.
[9] Zhao, B., & Wu, T. (2024). Classification of Hyperspectral and LiDAR Data by Transformer-Based Cross-Modal Self-Attentive Feature Fusion. Remote Sensing, 16(1), 94.
[10] Zhang, Z., Cai, Y., Liu, X., Zhang, M., & Meng, Y. (2024). An Efficient Graph Convolutional RVFL Network for Hyperspectral Image Classification. Remote Sensing, 16(1), 37.