International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

DL Hypersight: A Progressive Spectral Spatial Reasoning Framework for Hyperspectral Image

V. Vaishnavi
Dept of Electronics and Communication Engineering
Qis College Of Engineering and Technology
Ongole, India

Ch. V. V. Santhosh Kumar
Dept of Electronics and Communication Engineering
Qis College Of Engineering and Technology
Ongole, India

T. Ram teja
Dept of electronics and communication engineering
qis college of engineering and technology
ongole, india

M. Praharsha
Dept of Electronics and Communication Engineering
Qis College Of Engineering and Technology
Ongole, India

S. Govinda Rajulu
Dept of Electronics and Communication Engineering
Qis College Of Engineering and Technology
Ongole, India

G. Vamsi
Dept of Electronics and Communication Engineering
Qis College Of Engineering and Technology
Ongole, India

Abstract **- The classification of hyperspectral images are challenging due to their multiple spectral bands and complex spatial information they carry. Many deep learning techniques currently in use attempt to learn all spectral and spatial information in a single phase, which restricts their capacity to successfully comprehend both local details and global context. In this study, we provide a paradigm for hyperspectral picture categorization using progressive spectral-spatial reasoning. The suggested approach learns characteristics in phases. Convolutional neural networks are first used to extract local spectral-spatial properties. The characteristics are then adjusted to complicated and irregular item forms using a structure-aware module. Lastly, long-range associations throughout the image are captured using a self-attention-based global context refinement module. The suggested approach outperforms current approaches in terms of classification accuracy, according to experimental results on common hyperspectral datasets, particularly when there are few training examples available.**

Keywords **- H.yperspectral Image Classification, Spectral–Spatial Features, Progressive Learning, CNN, Self-Attention, Limited training Samples,Global Context.**

I. INTRODUCTION

Hyperspectral image classification is a major field of study in computer vision and remote sensing, each pixel in a hyperspectral image (HSI) carries rich spectral information over hundreds of narrow and contiguous bands, making its classification complex. HSI analysis is extremely useful in applications like agriculture monitoring, environmental evaluation, mineral exploitation, urban planning, and defense surveillance because of its thorough spectral signature, which allows for accurate identification of materials and land-cover types. Nevertheless, successful classification is a challenging endeavor due to the high dimensionality and complex spatial patterns of hyperspectral data. Creating robust models is made more challenging by the combination of spectral redundancy, geographical variability, and a small number of labeled data.

This work concentrates on spectral-spatial feature learning for hyperspectral image categorization using deep learning. Conventional machine learning techniques frequently underperform in the proper utilization of the combination spectral and spatial information included in HSIs, since they mostly depend on manually created features. Convolutional Neural Networks (CNNs), which automatically learn hierarchical features, have demonstrated bright prospects with the development of deep learning. Nonetheless, there is still work to be done in order to incorporate both local spatial structures and global contextual links into a single framework.

The study's problem statement is that many deep learning techniques now in use try to learn spectral and spatial properties in a single step while handling all data equally. These methods frequently have trouble in simultaneously capturing long-range reciprocity and fine-grained local patterns. Because of this, they could not work well in complicated sceneries with objects that have small spectral changes, mixed pixels, or irregular forms. In addition, the loss of performance frequently occurs when there are few labeled samples available, which is a common situation in hyperspectral imaging.

Current HSI classification systems include pixel-wise classifiers, 2D/3D CNN-based models, and more recently, attention-based or transformer-inspired networks. Pixel-wise approaches produce noisy classification maps because they primarily focus on spectral characteristics while ignoring spatial context. Whereas, they are frequently restricted to local receptive fields, CNN-based techniques enhance performance by incorporating geographic neighbors. In an effort to collect global information, some recent models employ attention mechanisms; nevertheless, these mechanisms are often applied in a single-step learning

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

process without a structured development of feature refinement. This restricts their capacity to adjust to multi-scale contextual interactions and detailed item structures.

In order to overcome these constraints, this paper suggests a progressive spectral–spatial reasoning framework for hyperspectral image categorization . Learning features in stages as opposed to a single pass is the main principle. At first, local spectral-spatial representations are extracted using CNNs. These attributes are then improved using a structure-aware module to more accurately represent the boundaries and contours of irregular objects. Lastly, long-range reliances throughout the scene are captured by a self-attention-based global context refinement module. In contrast to traditional methods, this progressive learning strategy endeavors to increase classification accuracy, strengthen resilience under small training samples, and improve feature discrimination.

II. LITERATURE REVIEW

Transformer-Based Spectral-Spatial Models: In contrast to traditional supervised models, Tang et al.'s recent work, HyperEAST, the author focused on an Enhanced Attention-Based Spectral-Spatial Transformer with Self-Supervised Pretraining for hyperspectral image classification which uses self-supervised learning to enhance attention feature learning and robustness on large HSI datasets. Building on ground work like the previously described Spatial-Spectral Transformer architecture, another expansion is that the employment of spectral-spatial Transformers, which combine local CNN features with self-attention processes to capture global reliance across high-dimensional spectral bands.

Varahagiri et al.,(2024) Hybrid Convolutional and Transformer Architectures: Because they take advantage of CNN inductive biases for local spectral-spatial feature extraction while utilizing transformers for long-term associations, hybrid techniques that integrate CNNs with transformer blocks are surging in popularity. For example, the 3D-Convolution Guided spectrum-Spatial Transformer effectively blend local spectrum and spatial information using 3D convolutions between transformer layers, showing increases in discriminative feature learning and classification accuracy. Furthermore, multi-scale transformer models and hierarchical transformer designs created for HSI enhance modeling across scales and relational connections.

Multi-Scale and Multi-Attention Networks: Techniques emphasizing multi-scale feature extraction and multi-attention mechanisms have proven successful in resolving spectral redundancy and complex spatial patterns in HSIs. In order to capture both local scale variance and long-range spectral–spatial interactions, Sun et al. (2024) provide a Multi-Scale Convolution and Multi-Attention Mechanisms (MSCF-MAM), a model that combines multi-scale convolutions with pyramid squeeze attention and transformer encoders. These techniques enable models to assist a variety of object sizes and spectral diversity, as well as more complex feature representations.

Lightweight and Efficient Transformer Variants: Because self-attention has a significant calculation cost, efficiency is an increasing problem in transformer-based HSI categorization. Dynamic token selection is used by models like the Efficient Dynamic Token Selection Transformer (EDTST) (Hu et al., 2025) to reduce unnecessary complications while preserving crucial spectral–spatial information. In order to reduce complexity and enable real-time or resource-constrained applications without significantly endangering classification performance, other recent research investigates frequency-domain or token-level optimization strategies.

Emerging Models and Future Directions: Newer paradigms, such as Mamba-based models, show promise beyond transformer and hybrid CNN-transformers. MambaHSI: Spectral and Spatial, In order to overcome the computational obstruction of attention while preserving high performance, Mamba for HSI classification (Li et al., 2025) proposed a linear-complexity Mamba architecture that simulates long-range interactions adaptively across spectral and spatial dimensions. Similarly, generative frameworks like graph or state-space models and diffusion-based architectures (SpectralDiff) push the boundaries of spectral–spatial representation learning, showcasing the variety and continuous innovation of the discipline.

III. PROPOSED METHODOLOGY

A. System Overview

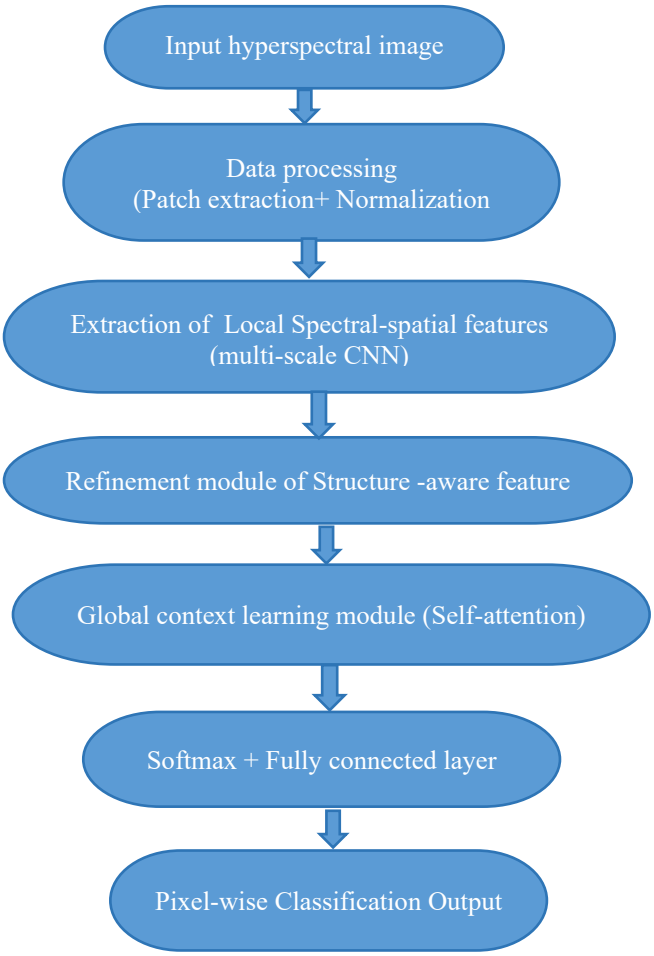

Fig 1. Block diagram

The proposed system is a Progressive Spectral–Spatial Reasoning Framework for accurate hyperspectral

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

image classification. A hyperspectral picture cube with rich spectral information across many bands serves as the input of the system. To maintain neighborhood information, the data first goes through preprocessing, which includes spectral normalization and the extraction of spectral-spatial patches surrounding each pixel. In addition to preparing organized data for the learning framework, this phase lowers spectral variability.

Following preprocessing, the system uses a multi-scale convolutional neural network (CNN) to extract local spectral-spatial features. This step allows the model to learn crucial low-level and mid-level representations by capturing fine-grained spectral signatures and local spatial textures. However, complex land-cover structures cannot be well modeled by local features alone. To improve spatial adaptability, a structure-aware feature refinement module is implemented. In order to improve intra-class consistency and lessen misunderstanding between adjacent classes, this module modifies intermediate characteristics to more accurately reflect irregular object forms and bounds.

Ultimately, a global context learning module that uses a self-attention mechanism receives the enhanced features. In order to enable the system to comprehend global semantic links beyond small neighborhoods, this stage models long-range dependencies between distant regions in the hyperspectral image. The final pixel-wise classification map is created by feeding the globally enhanced features into fully linked layers and a softmax classifier. The system produces robust and distinctive feature representation by gradually learning from local details to global reasoning, which improves hyperspectral image classification performance, particularly in situations with smaller training data.

B. Extraction of Local Spectral-Spatial Features

Initially the suggested framework focuses on using the hyperspectral image to learn local spectral-spatial representations. On each pixel a patch centered and is taken into consideration rather than examining each pixel separately, allowing both the pixel and its surrounding neighborhood to contribute to feature learning. Because of their excellent capacity to capture local correlations, in this stage Convolutional Neural Networks (CNNs) are used. In order to extract edge structures, texture patterns, and significant spectral correlations within local regions, the network simultaneously analyzes spectral and spatial input through 2D or 3D convolution operations.

Additionally, this step preserves unique material-specific characteristics found in hyperspectral data while lessening the impact of spectral redundancy. Feature maps are created by CNN that accurately depict local land-cover features by learning compact but informative representations. For the subsequent phases of structural refinement and global context learning, these enhanced local features provide a solid and trustworthy basis.

C. Module for Structure-Aware Feature Refinement

Though the CNN-based local feature extraction is good at capturing spectral-spatial patterns, it frequently has trouble representing the irregular object structures that are typical of hyperspectral pictures. Simple geometric shapes are rarely found in real-world land-cover regions including vegetation, metropolitan areas, and soil patches. When using solely local convolution operations, this results in fractured feature representations, jumbled pixels, and border ambiguities. Therefore, traits that are only learned from small neighborhoods could not be structurally consistent across object regions.

The second step of the proposed framework introduces a structure-aware feature refining module to get around this restriction. The module aims to improve object-level representation by adding contextual structure information and spatial continuity. The module modifies feature responses by taking into account the relationships between nearby pixels that most likely belong to the same object rather than treating pixels separately. This reduces noise brought on by sudden spectral changes or mixed pixels while enhancing intra-class similarity.

Additionally, the identification of boundaries between neighboring classes were enhanced by structure-aware refinement. The module creates more spatially consistent feature maps by highlighting structural coherence within regions and improving contrast across boundaries. In the end, these improved representations increase classification robustness in complicated hyperspectral situations by offering a more robust and dependable input to the next global context learning stage.

D. Refinement of Global Context through Self-Attention

While local feature extraction and structure-aware refinement enhance spatial representation, the long-range relationships throughout the scene must be modeled for hyperspectral image classification. Even widely separated pixels can share comparable spectral properties and belong to the same land-cover class. Such non-local interactions are difficult for traditional convolution processes to capture due to their narrow receptive fields. When the global context is not taken into account, this constraint may result in conflicting predictions.

The suggested framework includes a self-attention-based global context refinement module as the last feature learning step to solve this problem. The model can assess the degree to which each pixel is connected to every other pixel by using self-attention mechanisms that calculate relationships between all spatial positions within the feature map. The module integrates data from remote but semantically linked regions by allocating attention weights according to feature similarity. Instead of depending only on local neighborhoods, this allows the model to develop a global picture of the scene.

By enhancing stability within the same class throughout the image and reducing confusion between visually similar but semantically distinct classes, the global context refinement improves feature discrimination. Consequently, the final feature representation has greater robustness and semantic meaning. Improved pixel-wise classification accuracy results from passing these globally refined features to the classification layer, especially in complex hyperspectral images with varied land-cover patterns.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

IV. DATA SET PREPARARTION

The Indian Pines hyperspectral dataset is one of the most popular benchmark datasets in hyperspectral image classification research, is used to assess the effectiveness of the suggested framework. The collected dataset was taken over agricultural fields in northwest Indiana, USA, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. With extensive spectrum information spanning 224 spectral bands in the wavelength range of 0.4–2.5 μm, it has a scene size of 145 × 145 pixels. Usually, 200 spectral bands are kept for investigation after water absorption and noise bands are eliminated.

In the Indian Pines picture there are sixteen land-cover classes, comprising different kinds of crops, vegetation, and soil regions. These classes show complex spatial distribution and strong spectral similarity. Due to mismatched pixels, odd item forms, and a lack of labeled samples for some classes, this makes the dataset very complex. Indian Pines are a suitable benchmark for assessing the efficacy of global context modeling techniques and spectral-spatial learning methods because of their features.

The proposed progressive spectral–spatial reasoning framework is trained and assessed in this work using the ground truth map of the Indian Pines dataset. The dataset offers a realistic test scenario in which the model must retain spatial consistency while differentiating between spectrally similar classes. The fig1 depicts the spatial distribution of all 16 classes in the scene and serves as an illustration of the ground truth map utilized in the tests.

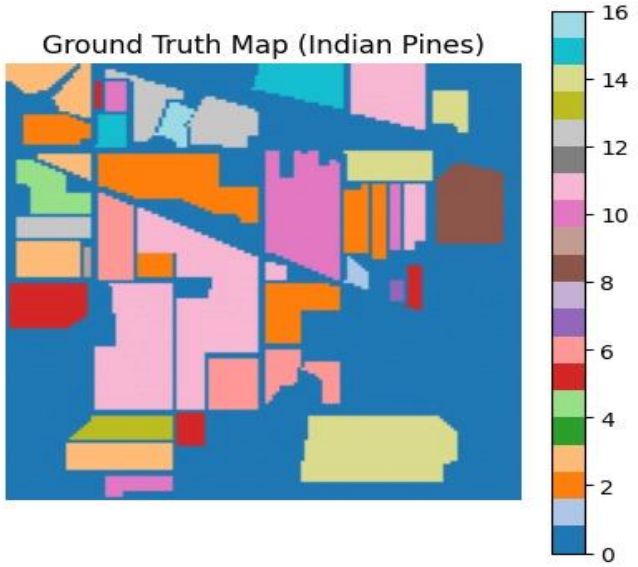

Fig 2. Ground truth map of the Indian Pines hyperspectral dataset

V. EXPERIMENTAL SETUP

Experiments were carried out on the Indian Pines hyperspectral dataset to assess the effectiveness of the suggested progressive spectral–spatial reasoning paradigm. To lessen spectral variability across various bands, the hyperspectral data were adjusted before training. To maintain local neighborhood information and facilitate patch-based learning, spectral–spatial patches of fixed size 11 × 11 × B were extracted surrounding each labeled pixel. This preprocessing stage guarantees the efficient use of spatial structures and spectral signatures during training.

The labeled samples were split into training and testing sets at random for the purpose of developing the model. To replicate genuine limited-data situations, a small percentage of each class's samples were used for training, while the remaining samples were set aside for assessment. The PyTorch deep learning framework was used to create the suggested network. The Adam optimizer was used to optimize the model parameters with a starting learning rate of 0.001. The network was trained for 200 epochs with a batch size of 32, and the learning process was guided by the cross-entropy loss function. To increase convergence stability, mini-batch training was used.

Every experiment was carried out on a system with an NVIDIA GPU, which sped up the training procedure. Experiments were repeated several times, and the average results were given to guarantee stability and fair evaluation. The model's performance was evaluated using common metrics for hyperspectral classification, such as the Kappa coefficient, Overall Accuracy (OA), and Average Accuracy (AA). These metrics offer a thorough assessment of class-wise consistency as well as overall classification performance.

VI. RESULTS AND DISCUSSION

The capacity of the suggested Progressive Spectral–Spatial Reasoning Framework to develop discriminative representations under challenging spectral and spatial settings was assessed using the Indian Pines hyperspectral dataset. The dataset is a difficult benchmark since it includes several vegetation and land-cover classifications with irregular geographic distribution and strong spectral similarity. The suggested model's progressive learning design allows characteristics to be gradually refined, starting with local spectral-spatial extraction, moving on to structural adaption, and concluding with global contextual reasoning. The network can overcome the drawbacks of conventional single-stage CNN or transformer-based techniques thanks to this tiered learning strategy.

The projected classification maps show the model's qualitative classification performance. The output of the suggested framework exhibits smoother regions, more distinct object boundaries, and less salt-and-pepper noise when compared to traditional approaches. The global attention module, which records long-range relationships among distant but semantically linked pixels, and the structure-aware refinement module, which improves spatial consistency inside object regions, are responsible for this improvement. Because of this, even in regions with complicated class boundaries, the model generates visually consistent categorization maps.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

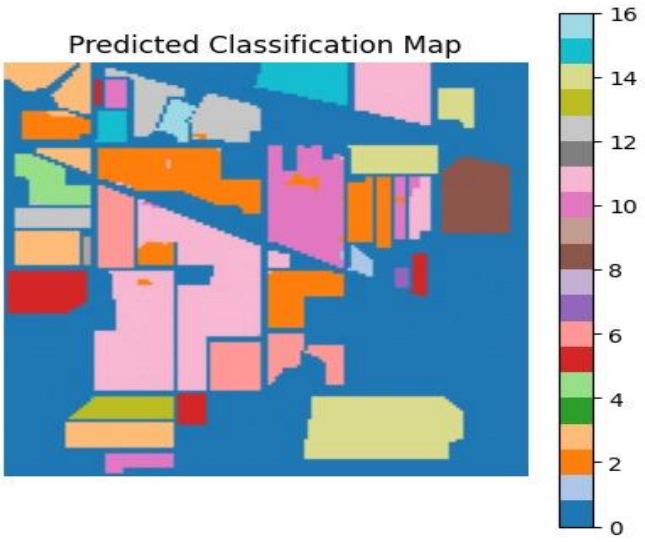

Fig 3. Predicted Classification Map

The full-scene categorization result is shown to better illustrate the model's capabilities. The majority of land-cover regions are accurately defined with little fragmentation, according to the output map. Strong consistency is seen in areas that correspond to vegetation classes and agricultural fields, suggesting that the progressive learning approach effectively lowers misclassification brought on by spectral redundancy and mixed pixels. The spatial continuity over wide areas demonstrates how well global context learning preserves class uniformity.

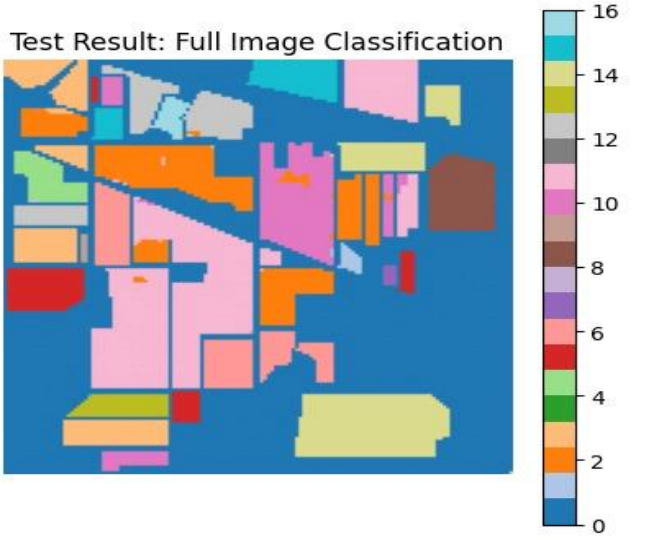

Fig 4. Test result of full image classification

The suggested model's efficacy in hyperspectral image categorization is further supported by its quantitative performance. The majority of pixels in the picture are correctly classified, as evidenced by the framework's high Overall Accuracy (OA) of 98.57%. Even for land-cover classes with little training samples, the Average Accuracy (AA) of 98.54% shows steady performance. Furthermore, a strong agreement between predicted labels and ground truth that goes beyond chance is indicated by the Kappa coefficient of 0.9837. These findings confirm that the progressive spectral-spatial learning approach greatly improves global contextual knowledge, boundary refinement, and feature discrimination.

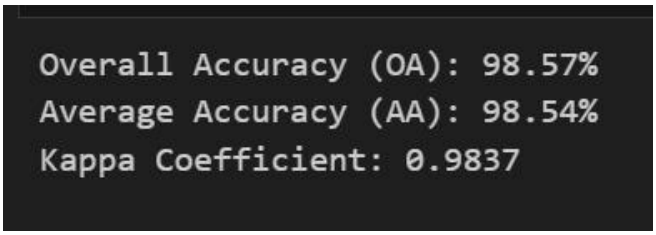

Fig 5. Overall accuracy (OA), Average accuracy (AA), Kappa coefficient

Additional proof of the efficacy of the suggested strategy comes from quantitative performance evaluation. Overall Accuracy (OA), Average Accuracy (AA), and the Kappa coefficient are all high for the model. The Kappa coefficient quantifies agreement between forecasts and ground truth that goes above chance, AA denotes consistent performance across all classes, and OA represents the overall percentage of correctly categorized pixels. The combination of local CNN features, structure-aware refinement, and global attention enhances the discriminative capability of the learnt representations, as seen by the robust performance across these criteria.

The confusion matrix, which shows class-wise classification performance, is used to provide a more thorough study. The matrix's diagonal dominance shows that the majority of classes are accurately predicted. Due to their overlapping spectral signatures, there is some confusion between vegetation classes that are spectrally similar. The structure-aware and global attention modules successfully decrease ambiguity at class boundaries, as evidenced by the low misclassification rates. This demonstrates that both local detail representation and global semantic understanding are improved by the progressive reasoning technique.

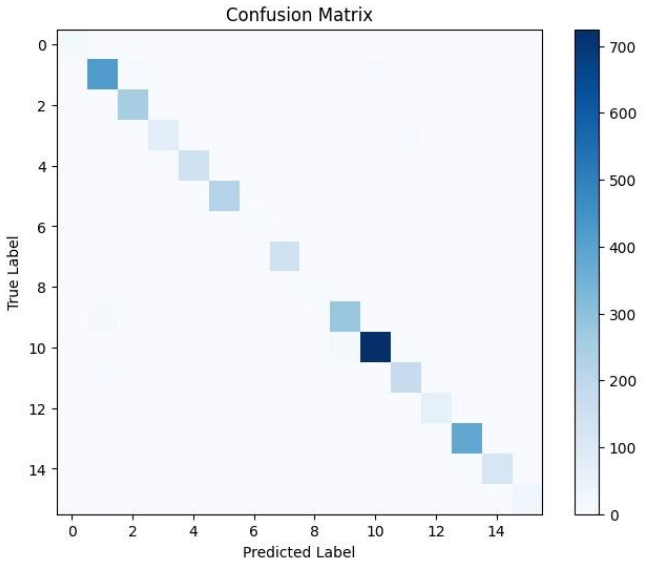

Fig 6. Confusion Matrix

Overall, the experimental results demonstrate that the suggested framework performs better than single-stage

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

deep learning techniques and traditional spectral-only and spatial-only models. Better boundary delineation, less noise, and increased global uniformity are the results of the progressive design's gradual feature refinement. Because of these benefits, the suggested approach is especially well-suited for hyperspectral image classification problems where data complexity and small training samples offer serious difficulties.

VII. CONCLUSION

In order to overcome the drawbacks of traditional single-stage feature learning techniques, this research introduced a Progressive Spectral–Spatial Reasoning Framework for hyperspectral picture categorization. The suggested framework successfully captures fine-grained local details, adjusts to complex object structures, and models long-range dependencies throughout the scene by breaking down the learning process into three stages: local spectral–spatial feature extraction, structure-aware refinement, and global context learning. The model can generate more spatially consistent and discriminative feature representations thanks to this progressive learning approach.

The efficiency of the proposed method is demonstrated by experimental evaluation on the Indian Pines dataset. Strong classification performance and consistent agreement with ground truth labels were demonstrated by the framework's high Overall Accuracy, Average Accuracy, and Kappa coefficient. The benefits of combining CNN-based local learning with structure-aware adaptation and self-attention-based global reasoning are confirmed by visual results that demonstrate smooth area boundaries, decreased noise, and increased class consistency.

Considering all the things, the proposed system offers a reliable solution for hyperspectral picture classification in difficult situations including spectral redundancy, asymmetrical object forms, and a small number of training samples. The framework's progressive reasoning nature makes it appropriate for real-world remote sensing applications that need precise land-cover mapping.

VIII. FUTURE SCOPE

The proposed approach spectral-spatial reasoning framework shown good classification performance, however there are a number of intriguing avenues that could improve its ability and usefulness. First, under conditions of limited labeled data, feature learning can be greatly enhanced by the stratergy of self-supervised pretraining. Before fine-tuning for classification, the network can acquire significant spectral–spatial representations from vast volumes of unlabeled hyperspectral data using techniques like contrastive learning or masked spectral modeling.

Increasing computational efficiency is another crucial path. Despite its effectiveness, the global self-attention module has a high computational and memory cost. In order to lower complexity and facilitate deployment in real-time or resource-constrained remote sensing systems, future research can investigate lightweight or effective attention techniques, such as linear attention, sparse attention, or window-based attention. Furthermore, incorporating multi-scale feature fusion techniques (such as hierarchical aggregation or pyramid pooling) can improve the model's capacity to identify both small and large land-cover regions.

Moreover, spectral band selection or spectral attention techniques can be incorporated into the framework to highlight the most informative bands and eliminate redundancy in hyperspectral data. Additionally, using semi-supervised learning techniques like consistency-based training or pseudo-labeling can assist make use of unlabeled data and enhance generalization. Applications like crop monitoring, deforestation detection, and land-cover evolution research would be made possible by expanding the model to incorporate multi-temporal hyperspectral data. At last, object-level interactions and spatial interdependence across areas can be captured more accurately by incorporating graph-based spatial modeling approaches like Graph Neural Networks (GNNs). The proposed architecture may become more effective, scalable, and appropriate for complex real-world hyperspectral remote sensing jobs as a result of these enhancements.

REFERENCES

[1] J. Tang, N. Ma, C. Jia, R. Tian, and Y. Guo, “HyperEAST: An enhanced attention-based spectral–spatial transformer with self-supervised pretraining for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[2] S. Varahagiri, A. Sinha, S. R. Dubey, and S. K. Singh, “3D-convolution guided spectral–spatial transformer for hyperspectral image classification,” in *Proc. IEEE Conf. Artificial Intelligence*, 2024.

[3] Y. Wang *et al.*, “Efficient dynamic token selection transformer for hyperspectral image classification,” *Remote Sensing*, vol. 17, no. 18, 2025.

[4] Q. Sun, G. Zhao, X. Xia, Y. Xie, C. Fang, L. Sun, Z. Wu, and C. Pan, “Hyperspectral image classification based on multi-scale convolutional features and multi-attention mechanisms,” *Remote Sensing*, vol. 16, 2024.

[5] Y. Xu, D. Wang, and L. Zhang, “Dual selective fusion transformer network for hyperspectral image classification,” *Neural Networks*, 2025.

[6] Y. Zhang, L. Liang, J. Mao, *et al.*, “Global–local multigranularity transformer for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[7] C. Ji, X. Zhang, and H. Meng, “CenterFormer: Center spatial–spectral attention transformer network for hyperspectral image classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[8] Y. Li, Y. Luo, L. Zhang, Z. Wang, and B. Du, “MambaHSI: Spatial–spectral Mamba for hyperspectral image classification,” *arXiv preprint arXiv:2501.04944*, 2025.

[9] Y. Zhong and X. Hu, “Spectral–spatial feature tokenization transformer for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[10] X. He, Y. Chen, and Z. Lin, “Spatial–spectral transformer for hyperspectral image classification,” *Remote Sensing*, vol. 13, no. 3, 2021.