Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 10, pp. 3565–3574
DOI : 10.47974/JIOS-2484

Cross dimensional deep fusion learning for an efficient hyperspectral image classification

S. Tahseen Banu *
Department of Computer Science & Engineering
Jawaharlal Nehru Technological University Anantapur
Anantapur 515002
Andhra Pradesh
India

K. Subba Reddy †
Department of Computer Science & Engineering
Rajeev Gandhi Memorial College of Engineering & Technology
Nandyal 518501
Andhra Pradesh
India

Abstract

Hyperspectral image (HSI) classification is an emerging need in remote sensing applications. Extraction of spatial and spectral cues, and their relevance, are crucial for the classification of HSI. Notable research has progressed to address challenges in HSI classifications, including pixel boundary correspondences, spectral band similarities, and high-dimensional issues. Recent methods, specifically SclusterFormer, have primarily focused on these issues and have achieved significant improvements in HSI classification. It effectively enables spectral and spatial information using cluster-based attention and processes features independently and combines them through a simple additive fusion strategy, which limits its ability to model cross-dimensional dependencies. To address these limitations, a Cross-Dimensional Spectral–Spatial Fusion Framework (CDSSF) is proposed. The proposed method introduces an adaptive attention-based fusion mechanism that dynamically integrates spectral and spatial features, enabling effective interaction across dimensions. This enhances feature discrimination, improves robustness to spectral variability, and leads to superior classification performance compared to the baseline model.

Subject Classification: *Primary 68T45, Secondary 68T07.*

Keywords: *Hyper spectral image, Deep learning, Spectral features, Spatial features, Image classification.*

* *E-mail:* **tahseen511@gmail.com (Corresponding Author)**
† *E-mail:* **mrsubbareddy@yahoo.com**

© Taru Publications

1. Introduction

Hyperspectral image (HSI) classification has become a critical task in remote sensing, as reported by Zhong et al. [14] and Yu et al. [16]. As it provides high spectral resolution with hundreds of continuous bands. This leads to accurate material and land cover class identification, while HSI processing is required for numerous applications such as environmental, agricultural, urban and military. However, the curse of dimensionality, the scarcity of labelled samples, and spatial heterogeneity make feature extraction and classification in hyperspectral imagery quite difficult. With the advent of deep learning, HSI classification performance has greatly improved through the use of both spectral and spatial feature extraction. In particular, the study [17] proposed SClusterFormer, a transformer-based model that effectively captures hierarchical spectral–spatial features for hyperspectral image classification, excellent long-range modelling capabilities by leveraging cluster attention, discussed in Vasti and Dev [18]. Such methods can effectively extract spectral and spatial features, yielding better results than classic convolution-based models. Nonetheless, most modern approaches focus on extracting spectral and spatial features separately and then combining them using relatively simple techniques, namely summation and/or concatenation. However, such approaches do not account for the intrinsic cross-dimensional relationships between spectral features and spatial patterns. Most of the current methods, such as cross-scale spectral calibration, are image reconstruction approaches rather than classification approaches. The classification performance is also restricted by not taking into account the interactions between the spectral and spatial dimensions, which is the case for spatiotemporal fusion methods. To tackle this, CDSSF adaptively combines the spectral and spatial features with an attention mechanism. This yields a set of more discriminative characteristics and enhance the classification accuracy, spectral invariance and generalization between data sets.

2. Literature Study

Recent HSI classification methods use deep learning, graph networks, and transformers to exploit spectral–spatial information. The idea of 3D-2D-GAT was proposed by Zhang et al. [1] and the frequency-interaction transformer was proposed by Arshad et al. [2]. To enhance feature learning and dependency modelling, Peng et al. [3] proposed CASST, Chen, Liu, and Zhou [4] presented SSIT and Ding et al. [5] proposed GLT-Net. Later, Zhang et al. [6] proposed MATNet and Cheng et al. [7] proposed S3FAN for further improving the feature extraction. However, these methods often suffer from high computational complexity, inefficient fusion strategies, redundant information, and limited ability to capture cross-dimensional spectral–spatial relationships, reducing their adaptability and generalisation in complex scenarios. Furthermore, multimodal and cross-modal fusion techniques have been used to harness complementary information. Zhao et al. [8] introduced a self- and cross-attention-enhanced transformer (SCAET) for visible and thermal hyperspectral data fusion, and also

proposed a MICF-Net for HSI-LiDAR fusion using cross-attention. Likewise, He et al. [9] proposed MADNet, a dynamic-scale attention-based model for multimodal feature fusion. While these models improve classification performance, they introduce significant complexity to the network architecture and rely on multimodal data, which is undesirable for single-modality HSI classification. Compared with this, Wang et al. [10] proposed an approach called NNCNet, which tackles the issue of limited labeled data by leveraging unlabeled data, thereby overcoming this limitation. However, these methods have difficulty with diverse data distributions and complex training strategies, and are ineffective in practical applications. Additionally, some approaches focus on spectral enhancement and reconstruction, Similarly, Ge et al. [11] explored transfer learning and channel-enhanced networks for hyperspectral image super-resolution. These techniques increase the representation of the spectrum, but they do not consider the spatial context and some are not developed for classification. Hybrid CNN–transformer and frequency-based models capture local and global features, but task specific models like that of Zhao, Zheng, and Hu [12] do not exist, and models without adaptive spectral–spatial fusion, such as the anomaly-detection model of Ge et al. [11] do exist. To fill this gap, CDSSF leverages attention-based fusion to fuse spectral and spatial information and efficiently models the interaction between spectral and spatial information to enhance the classification performance.

3. Proposed Methodology

The proposed Cross-Dimensional Spectral-Spatial Fusion (CDSSF) approach aims to capture the relationship between spectral and spatial features in hyperspectral images (HSI). CDSSF not only extracts these features separately but also proposes a dynamic attention-based fusion process that adaptively weights spectral and spatial features to achieve better representation.

Algorithm: Proposed CDSSF Methodology

Input:

- Hyperspectral Image X
- Ground Truth Labels Y

Output:

- Predicted Labels $\hat{Y}$
- Classification Metrics

Methodology:

Step 1: Data Preprocessing

1.1 Load hyperspectral dataset Xand labels Y. Normalize spectral values to remove scale variations.

1.2 Apply **Principal Component Analysis (PCA)** to reduce spectral dimensionality:

$$X = \mathrm{PCA}(X)$$

1.3 Extract labeled patches of size P × Pcentered around each pixel.
1.4 Split data into training and testing sets.

Step 2: Spectral Feature Learning

2.1 Pass input patch $X_i$through **Spectral Branch**:

$$F_{spec} = \text{ReLU}(\text{Conv}_{1\times1}(X_i))$$

Step 3: Spatial Feature Learning

3.1 Pass input patch $X_i$through **Spatial Branch**:

$$F_{spat} = \text{ReLU}\big(\text{Conv}_{3\times3}(X_i)\big)$$

Step 4: Cross-Dimensional Feature Interaction

4.1 Apply Global Average Pooling:

$$f_{spec} = \text{GAP}(F_{spec}), f_{spat} = \text{GAP}(F_{spat})$$

4.1 Combine pooled features:

$$f_{fusion} = f_{spec} + f_{spat}$$

Step 5: Adaptive Attention-Based Fusion (Core Step)

5.1 Learn adaptive weights:

$$\alpha = \sigma(W_2(\text{ReLU}(W_1(f_{fusion}))))$$

5.2 Perform cross-dimensional fusion:

$$F_{cdssf} = \alpha \cdot F_{spec} + (1-\alpha) \cdot F_{spat}$$

Step 6: Feature Refinement

6.1 Apply convolution for feature enhancement:

$$F_{refined} = \text{ReLU}(\text{Conv}_{3\times3}(F_{cdssf}))$$

6.2 Apply Global Average Pooling:

$$F_{global} = \text{GAP}(F_{refined})$$

Step 7: Classification

7.1 Feed features into Fully Connected layer:

$$\hat{Y} = \text{Softmax}(W \cdot F_{global})$$

Step 8: Model Training

8.1 Compute Cross-Entropy Loss:

$$\mathcal{L} = -\sum Y log(\hat{Y})$$

8.2 Update model parameters using Adam optimizer.

PCA is used to reduce noise and dimensionality while preserving valuable spectral data and fixed-size patches are taken around labelled pixels. The

CDSSF is a combination of two branches: a spectral branch (1×1 convolutions) and a spatial branch (3×3 convolutions), which learn complementary features. These features are adaptively fused with an attention module that concentrates on the most relevant information. The fused representation is fed to the fully connected and softmax layers, which are trained with the Adam optimiser and cross entropy loss to refine, pool and classify the representation.

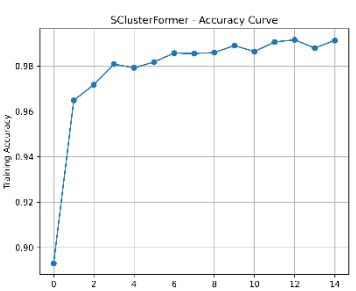
4. Experimental Results and Analysis

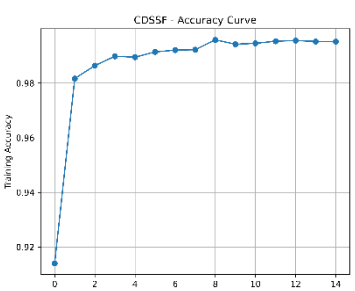
The proposed CDSSF framework was tested with three benchmark HSI datasets namely Pavia University (PaviaU), HongHu and HanChuan discussed in Zhong's et al. [14]. The assessment of performance was done according to overall accuracy (OA), the kappa coefficient, normalized mutual information (NMI), precision, recall, F1-score and AUC as done in Basha, Mouleeswaran, and Prasad [13] and Samantsinghar et al. [15]. The results show in Table 1, that CDSSF consistently outperforms the baseline across all datasets. On PaviaU, CDSSF achieved 99.57% OA compared to 98.90% for SClusterFormer, along with higher Kappa, NMI, and F1-score values, indicating better classification accuracy, class agreement, and feature discrimination.

Table 1

Performance Analysis between Recent SClusterFormer and Proposed CDSSF Method

Dataset	Model	OA	Kappa	NMI	Precision	Recall	F1	AUC
PaviaU	SClusterFormer	0.9890	0.9855	0.9684	0.9808	0.9755	0.9777	0.9999
PaviaU	CDSSF	0.9957	0.9943	0.9874	0.9913	0.9864	0.9885	1.0000
HanChuan	SClusterFormer	0.9607	0.9541	0.9274	0.9410	0.9327	0.9328	0.9992
HanChuan	CDSSF	0.9907	0.9891	0.9794	0.9856	0.9805	0.9828	0.9999
HongHu	SClusterFormer	0.9700	0.9621	0.9443	0.9508	0.9272	0.9313	0.9995
HongHu	CDSSF	0.9927	0.9908	0.9828	0.9834	0.9870	0.9848	1.0000





Contd...

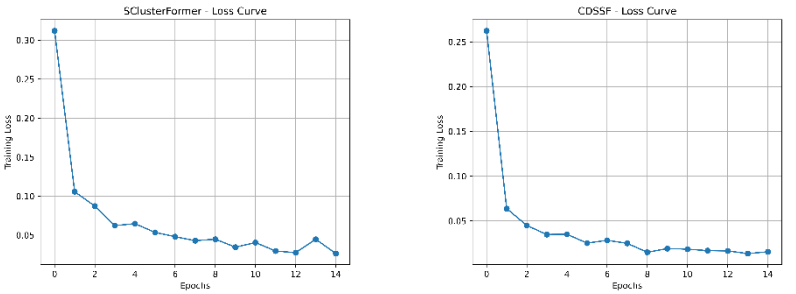


Figure 1
Accuracy and Loss Comparisons per Each Epoch between SClusterFormer and Proposed CDSSF for PaviaU dataset

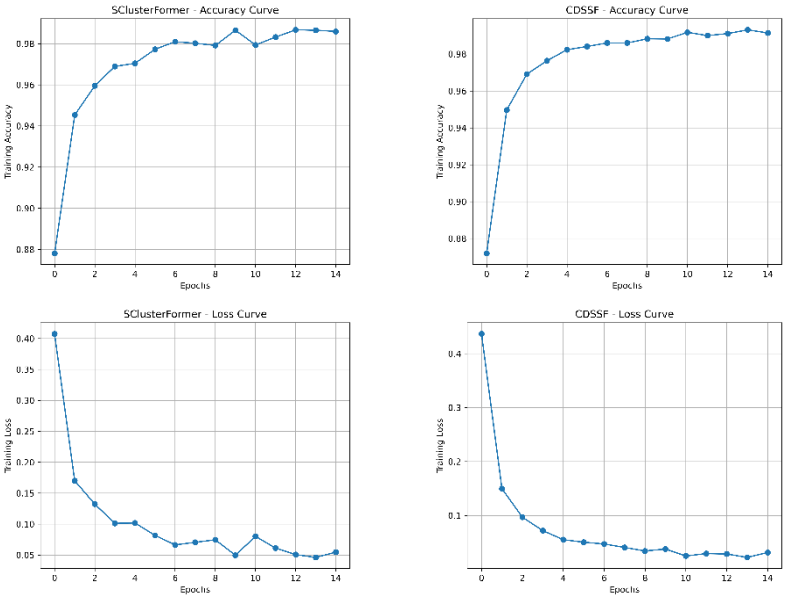


Figure 2
Accuracy and Loss Comparisons per Each Epoch between SClusterFormer and Proposed CDSSF for HongHUdataset

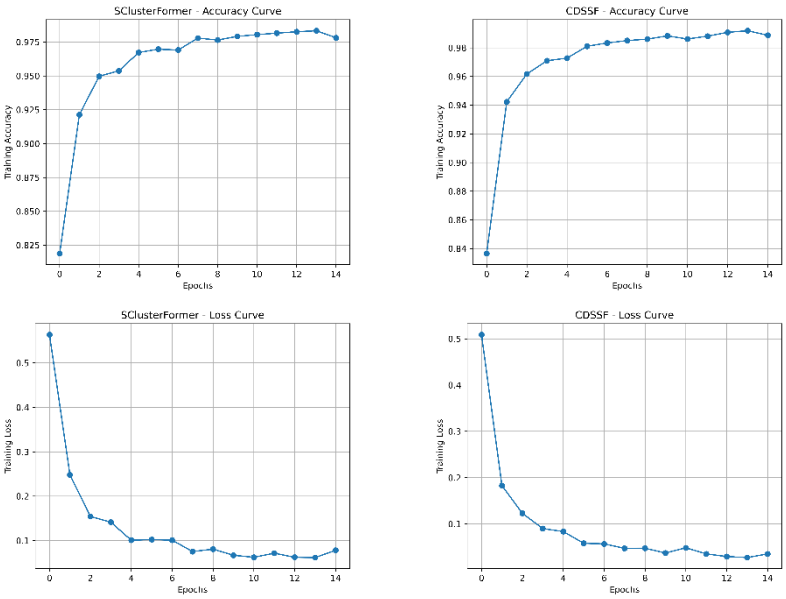


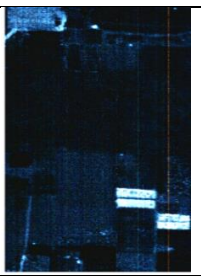
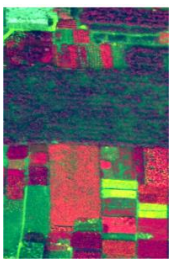
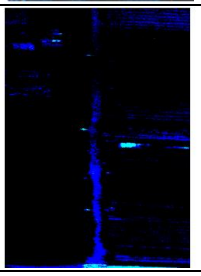
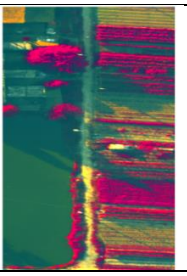
Figure 3
Accuracy and Loss Comparisons per Each Epoch between SClusterFormer and Proposed CDSSF for PaviaU dataset

CDSSF achieves high OA values on the HanChuan and HongHu datasets of 99.07% and 99.27%, respectively, which are superior to those of SClusterFormer, respectively. The complex spatial patterns and spectral variations do not reduce its performance as shown by the high Precision, Recall and F1 scores and AUC near 1.0 (Table 2), which illustrate its good class discrimination.

Table 2
HSI Results for Proposed CDSSF Method

Dataset	Input	HSI Result
PaviaU		

Contd...

HongHU		
HanChuan		

The accuracy and loss curves Figure 1, Figure 2 and Figure 3 shows that CDSSF learns faster and more stable than SClusterFormer, achieving higher accuracy with lower loss.

5. Conclusion and Scope of the Work

The proposed CDSSF framework addresses the limitations of existing methods, such as SClusterFormer, by effectively integrating spectral and spatial features via an adaptive attention mechanism. Experimental results show consistent improvements across all datasets, achieving higher OA, Kappa, NMI, and F1-scores. CDSSF also achieves near-perfect AUC values, demonstrating strong class discrimination. These results confirm that the cross-dimensional fusion strategy improves feature separability, spectral invariance, classification accuracy, and overall generalisation performance compared to existing approaches.

References

[1] H. Zhang, K. Tu, H. Lv, and R. Wang, "Hyperspectral image classification based on 3D–2D hybrid convolution and graph attention mechanism," *Neural Process. Lett.*, vol. 56, Art. no. 117 (2024).

[2] T. Arshad, B. Peng, A. Rahman, R. Khan, S. U. Khan, S. Alnazi, and N. Alturki, "Spectral-spatial wave and frequency interactive transformer for hyperspectral image classification," *Sci. Rep.*, vol. 15, Art. no. 27259 (2025), doi: 10.1038/s41598-025-12489-3.

[3] Y. Peng, Y. Zhang, B. Tu, Q. Li, and W. Li, "Spatial–spectral transformer with cross-attention for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15 (2022).

[4] Z. Chen, C. Liu, and J. Zhou, "SSIT: A spatial–spectral interactive transformer for hyperspectral image denoising," *Sci. Remote Sens.*, vol. 12, Art. no. 100276 (2025), doi: 10.1016/j.srs.2025.100276.

[5] K. Ding, T. Lu, W. Fu, S. Li, and F. Ma, "Global–local transformer network for HSI and LiDAR data joint classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13 (2022), doi: 10.1109/TGRS.2022.3216319.

[6] B. Zhang, Y. Chen, Y. Rong, S. Xiong, and X. Lu, "MATNet: A combining multi-attention and transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–15 (2023).

[7] S. Cheng, Z. Xue, Z. Li, A. Xu, and H. Su, "Spectral–spatial score fusion attention network for hyperspectral image classification with limited samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14521–14542 (2024), doi: 10.1109/JSTARS.2024.3440254.

[8] E. Zhao, Y. Su, N. Qu, Y. Wang, and J. Zeng, "Self- and cross-attention enhanced transformer for visible and thermal infrared hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 13408–13422 (2025), doi: 10.1109/JSTARS.2025.3571226.

[9] Y. He, B. Xi, G. Li, T. Zheng, Y. Li, C. Xue, and J. Chanussot, "Multilevel attention dynamic-scale network for HSI and LiDAR data fusion classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Art. no. 5529916 (2024), doi: 10.1109/TGRS.2024.3456754.

[10] M. Wang, F. Gao, J. Dong, H.-C. Li, and Q. Du, "Nearest neighbor-based contrastive learning for hyperspectral and LiDAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Art. no. 5501816 (2023).

[11] Z.-Z. Ge, Z. Ding, Y. Wang, L.-F. Bian, and C. Yang, "Spectral domain strategies for hyperspectral super-resolution: Transfer learning and channel enhance network," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 134, Art. no. 104180 (2024), doi: 10.1016/j.jag.2024.104180.

[12] M. Zhao, W. Zheng, and J. Hu, "HTC-HAD: A hybrid transformer-CNN approach for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 10144–10156 (2025).

[13] M. Suleman Basha, S. K. Mouleeswaran, and K. Rajendra Prasad, "Hybrid visual computing models to discover the clusters assessment of high dimensional big data," *Soft Comput.,* vol. 27, pp. 4249–4262 (2023).

[14] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H^2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ*., vol. 250, Art. no. 112012 (2020), doi: 10.1016/j.rse.2020.112012.

[15] U. Samantsinghar, S. Sethi, R. C. Sahoo, and D. C. Panda, "Optimizing spectrum sensing performance and prediction of trustworthy users by implementing machine learning algorithms in cognitive radio network," *J. Inf. Optim. Sci.*, vol. 46, no. 4-B, pp. 1369–1383 (2025).

[16] H. Yu, H. Yang, L. Gao, J. Hu, A. Plaza, and B. Zhang, "Hyperspectral image change detection based on gated spectral–spatial–temporal attention network with spectral similarity filtering," *IEEE Trans. Geosci. Remote Sens*., vol. 62, pp. 1–13 (2024), doi: 10.1109/TGRS.2024.3373820.

[17] B. Gayathri and R. Gopi, "Some properties of k-even mean graph," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 4, pp. 1579–1583 (2025), doi: 10.47974/JDMSC-2244.

[18] M. Vasti and A. Dev, "Ensemble learning based predictive modelling on a highly imbalanced multiclass data," *J. Inf. Optim. Sci.,* vol. 45, no. 8, pp. 2141–2164 (2024), doi: 10.47974/JIOS-1778.

Received March, 2026