FusionNet-X: A Hybrid Spectral-Spatial Deep Learning Model for Hyperspectral Image Classification

Manish Nandy[1], Ashu Nayak[2]

[1,2]Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.

E-mail: [1]manishnandy10@outlook.com, [2]ashunayak1@outlook.com

Abstract

Hyperspectral imaging captures a dense stack of spectral bands, along with regular photo-like pixels, allowing it to distinguish minerals and map plants with fine detail. Still, mixing that rich spectrum with the shape and texture seen across space is tough whenever the feature space is vast and the training labels are scarce. In response, we develop a hybrid deep-network system that combines spectral and spatial learning within a two-pronged, or dual-branch, design. Its spectral arm runs a slim 1D CNN that hunts for small but telling shifts in color across just a few wavelengths. Meanwhile, the spatial arm feeds the same scene into a standard 2D CNN that detects edges, blobs, and other local structures. What each branch finds are merged by an adaptive attention layer that weighs the spectral cue against the spatial one on the fly before issuing the final class label. Tests on standard hyperspectral benchmarks demonstrate that our model surpasses traditional CNNs and the latest competitors in both accuracy and generalization to unseen sites. It also maintains high scores when classes are uneven or data is noisy, traits that are crucial for field campaigns and satellite work. Overall, the framework takes a significant step toward extracting all the valuable information from hyperspectral cubes and converting it into trustworthy maps.

Keywords: Hyperspectral Imaging, Spectral-Spatial Feature Extraction, Deep Learning, Convolutional Neural Networks (CNN), Attention Mechanism.

DOI: https://doi.org/10.36548/jiip.2025.2.013
Received: 18.05.2025, received in revised form: 16.06.2025, accepted: 29.06.2025, published: 08.07.2025

1. Introduction

Hyperspectral imaging, or HSI, is now a must-have in the remote sensing toolbox, as it records reflectance data across hundreds of thin color slices. Because HSI detects far more than the usual red, green, and blue bands, even tiny differences in surface sheen can be identified, helping analysts distinguish one crop from another or detect subtle shifts in the landscape [1]. That boxes it into favorite fields such as land-use surveys, crop-watch programs, and campaigns to track how the environment is changing. Yet, reading these rich data cubes remains challenging. Vast numbers of bands mean files are enormous, many slices repeat similar information, and labeled training samples are often scarce. Older algorithms have a tendency to selectively use spectral or spatial indications, failing to grasp the broader picture when both kinds of data are pertinent, and these obstacles make it possible for classifiers to learn noise rather than real patterns [2] [3] [4].

In response, machine-learning experts are turning to deep nets for the task. Simple convolutional networks, whether set up to scan 1D spectra or tile-shaped 2D images, already do a solid job of teasing apart spectral and spatial signals [5] [6]. Still, hybrid designs that integrate these two streams simultaneously usually win the contest, as they harness the best aspects from each domain without allowing redundancy to drag things down [7] [8].

This study presents FusionNet-X, a deep learning network designed explicitly for classifying hyperspectral images in mixed spectral spaces [9]. The model runs two side-by-side convolutional streams: a slim 1D CNN that detects tiny spectral patterns and a wider 2D CNN that identifies the overall shape of the images [10] [11]. A dynamic attention layer afterward combines the two, dynamically adjusting how much each stream is relied on per pixel. By combining spectral color information and spatial context, FusionNet-X achieves greater accuracy, noise reduction, and better generalization to new hyperspectral datasets.

Key Contribution

- It combines learned features from different branches, analyzing the deep layer to retain modality and boost discrimination ability, thereby improving overall classification performance.
- It effectively addresses and overlaps the spatial texture among the confusion through various feature representations.

- It demonstrates strong performance across multiple hyperspectral images with a dataset, followed by robustness to dataset variability.

Research Objective

The central role of Fusion Net-X is to present a deep learning framework that fuses spectral and spatial data in a manner intended to drive hyperspectral image classification forward. This entails creating a two-branch network, with one branch extracting geographic context and texture and the other extracting fine-grained spectral features. In a bid to maintain each branch's unique strengths while enhancing discriminative capability, a specific fusion process is used. The objective of the model is to prove its strength and applicability in a series of hyperspectral tests. It also seeks to minimize spectral-spatial uncertainty, especially in areas where there are coincidental class characteristics, thus enhancing precision in classification.

2. Related Work

Scientists have researched hyperspectral image classification for decades, as those specialized cameras collect light in various bands, allowing each pixel to contain a great deal of spectral information [12] [13]. Early approaches utilized pixel-based classifiers like Support Vector Machines, k-nearest Neighbors, and Random Forests, which tend to be accompanied by dimensionality reduction methods like Principal Component Analysis or Independent Component Analysis to minimize the data [14]. Although these classic methods work to a point, they ignore the patterns formed by neighboring pixels and struggle when only a handful of training samples are available.

A significant hurdle in hyperspectral image analysis stems from having far more color bands than labeled pixels to train on, and that imbalance invites overfitting and weak generalization. Real-world scenes are even messier: mixed pixels, repeated spectra, and nonlinear drifts caused by haze or shadow contribute to spectral-spatial confusion. Additionally, most methods overlook nearby pixels, which is a missed opportunity because small patches of grass or different roofing tiles often blend into nearly identical spectra.

Deep learning has revolutionized the field by enabling computers to learn features directly from raw data rather than relying on a manually selected features. Early studies relied on 1D convolutional layers to extract spectral clues from the signal, and researchers later stacked 2D layers on top to capture the images' two-dimensional patterns. More recent work has trialed 3D layers alongside Recurrent Neural Networks in an attempt to learn both spectral and spatial features simultaneously, yet these pursuits often consume a significant amount of time and power while demanding large labeled datasets [15] [16]. To alleviate that burden, new hybrid designs combine 1D and 2D convolutions, striking a more stable balance between peak accuracy and reasonable computational costs. The present study builds on that idea to form the FusionNet-X framework.

3. FusionNet-X Model

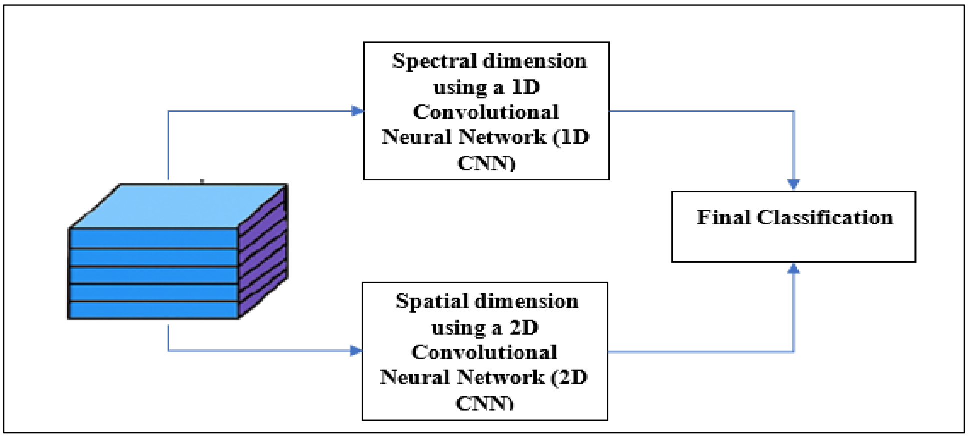

Figure 1. Architecture of FusionNet-X-Model for Hyperspectral Image Classification

Figure 1 illustrates how the new FusionNet-X model processes and classifies hyperspectral images. It starts with a 3D image cube that stacks height, width, and multiple spectral bands (H × W × B). Inside the network, these layers split into two side-by-side paths that extract different clues from the data. The first path, called the spectral branch, runs quick one-dimensional filters along each band's wavelength, allowing it to study how every pixel reflects light at every color. Because every material emits a unique light pattern, this step enables the model to distinguish substances by their spectral fingerprints.

Meanwhile, the second path, the spatial branch, stitches together standard two-dimensional filters that slide over rows and columns, capturing shapes, textures, and the way neighboring pixels blend. These spatial hints are important when many pixels look alike in the spectrum but differ in layout, so the extra information clarifies tricky borders. After both paths are completed, their results merge into a single stream, which the final classifier reads and tags with the correct land use or material label. By working together, the spectral and spatial branches provide FusionNet-X with a double set of eyes, thereby increasing accuracy and making the system more robust against noisy or mixed samples. The suggested hybrid spectral-spatial network with the dual-branch architecture can generalize well over most hyperspectral datasets because it learns spectral and spatial features separately to handle various features in the data, such as the number of bands, spatial resolution, and class distribution. The model's ability to use only the most beneficial features for each dataset is ensured by attention-based fusion mechanisms. Regularization methods such as dropout and batch normalization enhance the model's generalization and make it less prone to overfitting. The fact that the model performs well on multiple benchmarks implies that it can learn transferable representations and classify a broad range of real-world scenarios effectively.

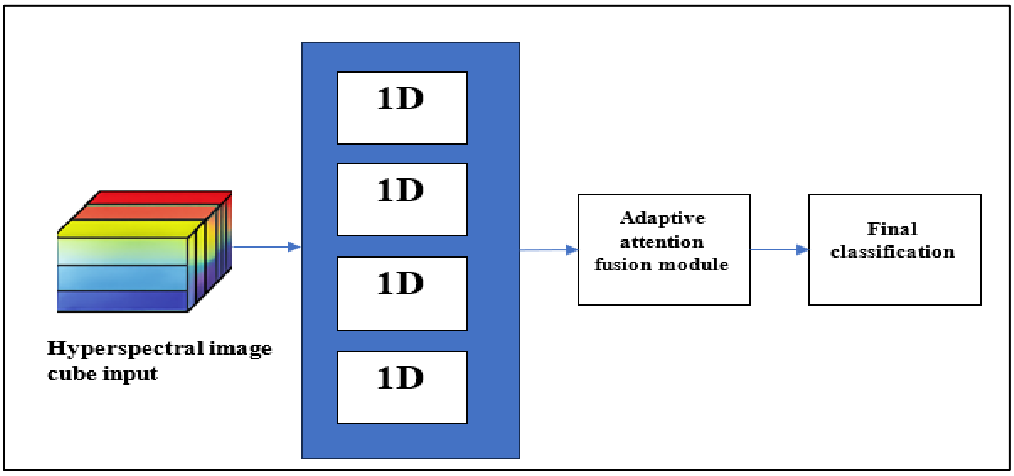

Figure 2. A Dual-Branch Hybrid Framework Designed for Hyperspectral Image Classification

Figure 2 depicts FusionNet-X, the total structure of a two-headed system to classify hyperspectral images. The process begins with a hyperspectral cube composed of height, width, and spectral bands, with a size of H × W × B. This three-dimensional block proceeds to both branches simultaneously: the spectral arm and the spatial arm. The spectral branch performs a sequence of 1D convolutions, unrolling deep wavelength-level features that perceive even mild band-to-band changes. In contrast, the spatial branch uses typical 2D convolutions, perceiving edges, textures, and other surface features in small image windows. After processing both branches, an adaptive attention module adjusts its weights and fuses only the most critical spectral and spatial information. The hybrid output is taken as input to a final classifier, which labels every pixel or region by its best match. By learning from spectral and spatial information in parallel, the architecture gives the model sharper vision and greater confidence, thereby improving performance on hyperspectral images. The spectral-spatial hybrid network structure enhances classification ability by extracting and fusing features from spectral and spatial information of hyperspectral images. It uses 1D convolutional layers to extract accurate spectral signatures and 2D convolutions for local structural and contextual information. Feature fusion processes merge the outputs, thereby reducing misclassifications in noisy or mixed spectral regions. Regularization techniques and deep hierarchical layers maintain high overall classification accuracy by preventing overfitting.

Algorithm: FusionNet − X for Hyperspectral Image Classification

Input:

- $Hyperspectral\ image\ cube\ I \in R^{H \times W \times B} where\ H = height, W = width, B = number\ of\ spectral\ bands$
- $Patch\ size\ P\ for\ spatial\ feature\ extraction$
- $Predefined\ class\ labels\ C$

Output:

- $\boldsymbol{Classified\ label\ map\ L \in R^{H \times W}}$

Steps:

1. $Preprocessing$
   - $Normalize\ spectral\ values\ in\ I$
   - $Extract\ fixed - size\ patches\ P \times P \times BP\ \ centered\ at\ each\ pixel$
2. $Spectral\ Branch\ (1D\ CNN)$
   - $For\ each\ pixel, extract\ the\ spectral\ vector\ s \in RB$
   - $Pass\ s\ \ through\ a\ stack\ of\ 1D\ convolutional\ layers$:

     $$fs = Conv1Dn(\dots(Conv1D1(s)))$$
3. $Output$: $spectral\ feature\ vector\ fs \in R^d$
4. $Spatial\ Branch\ (2D\ CNN)$
   - $From\ each\ patch, P \times P \times B\ reduces\ along\ the\ spectral\ axis\ (e.g., PCA\ or\ band\ selection)$
   - $Pass\ resulting\ P \times P\ image\ through\ 2D\ convolutional\ layers$:

     $$fp = Conv2Dn(\dots(Conv2D1(P)))$$
5. $Output$: $spatial\ feature\ vector\ fp \in R^d$
6. $Feature\ Fusion$
   - $Apply\ adaptive\ attention\ or\ concatenation\ to\ fuse\ features$:

     $$f_{fused} = Attention(f_s, f_p)$$

- $Optionally\ use\ fully\ connected\ layers\ for\ embedding$

7. $Classification$
   - $Feed\ \ _{fused}\ into\ a\ dense\ classification\ layer$:

$$y\hat{} = Softmax(W \cdot f_{fused} + b)$$

The FusionNet-X model gets the best out of both spatial and spectral data by processing two branches in parallel, each designed to capture aspects that the other cannot from hyperspectral images. Its spectral branch processes 1D convolutional layers that traverse the bands, capturing the minute differences in reflectance that allow the system to distinguish one mineral from another. Along the way, the spatial arm performs 2D convolutions that assess how closely pixels tend to cluster together, providing a visual that eliminates ambiguity where spectra are very hard to differentiate. Once both branches are complete, an adaptive attention dial decides how much of each to retain, and a final classification is created using a blend of the clear detail of spectral signals and the robust context of spatial cues. As a result of this integration, FusionNet-X more effectively understands real-world hyperspectral scenes and handles difficult mixed-material patches. The reason 1D convolution is utilized for spectral information extraction and 2D convolutional neural networks (CNNs) are employed for spatial structure extraction is that hyperspectral images possess two spatial dimensions and one spectral dimension. By performing 1D convolution along the spectral direction, the model can discover local patterns and correlations in spectra across wavelengths at every pixel. It maintains spectral continuity while allowing the model to learn material-dependent reflectance properties without altering spatial structure. However, 2D convolutions are employed to uncover local spatial patterns of texture, boundaries, and interpixel relationships that are critical in distinguishing objects with similar spectral signatures but different orientations in space. The hybrid spectral-spatial model makes feature learning deconfounded but complementary, leading to robust classification performance by integrating spectral accuracy with spatial context into a single framework.

4. Results

4.1 Dataset Description

In order to validate the performance of the novel FusionNet-X network, we performed experiments on two widely used benchmark hyperspectral datasets, Indian Pines and Pavia University. The Indian Pines scenes were imaged by the AVIRIS sensor while it was flying over agricultural fields in northwestern Indiana, USA. They are a spectral cube of size 145 by 145 pixels and cover 220 bands from 0.4 to 2.5 microns in wavelength. Generally, researchers only keep the 200 clean bands after removing noisy and water-soaked data. With 16 cover types, primarily crops and natural plants, the set is difficult because numerous classes are very similar to each other, and multiple pixels denote multiple targets.

Conversely, the Pavia University images were recorded by the ROSIS camera over the Italian city of Pavia. Following processing, the set measures 610 by 340 pixels and 103 usable bands. It classifies nine classes, such as buildings, asphalt roads, trees, and grass pastures. Since objects in these locations vary in shape, color, and texture, Pavia University provides greater spatial diversity than Indian Pines.

The two datasets were pre-processed to a [0, 1] range prior to training, and the same number of labeled samples per class was used for supervised learning. The rest of the samples were reserved for testing. Both datasets represent an exhaustive platform to evaluate the ability of FusionNet-X to model spectral richness and spatial heterogeneity within hyperspectral images.

4.2 Evaluation Metrics

Accuracy: Accuracy is the ratio of correctly predicted samples to the total number of samples. It provides a general indication of how often the classifier is accurate.

$$Accuracy = \frac{True\ Positives + \ True\ Negatives}{Total\ Samples}$$

In this study, FusionNet-X achieved an overall accuracy of 85.5%, indicating strong general performance across all classes.

Precision: Precision measures the proportion of correctly predicted positive samples among all predicted positive samples. It is beneficial in cases where false positives are costly.

$$Precision = \frac{True\ Positives}{Total\ Positives + \ + False\ Positives}$$

Recall (Sensitivity or True Positive Rate): Recall is the proportion of actual positives that the model correctly identifies. It is crucial when false negatives are critical, such as in rare class detection.

$$Recall = \frac{True\ Positives}{Total\ Positives + \ + False\ Negatives}$$

F1-Score: The F1-Score is the harmonic mean of precision and recall. It balances the two metrics and is especially valuable when classes are imbalanced.

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision\ +\ Recall}$$

Confusion Matrix

A confusion matrix displays the counts of actual vs. predicted classifications for each class. Diagonal entries indicate correct predictions, while off-diagonal entries reveal misclassifications. It helps identify specific weaknesses in class-wise performance.

4.3 Comparison with Other Deep Learning Models

Table 1. Performance Comparison of Proposed Model with Previous Models

Model	Indian Pines	Pavia University
SVM	84.3	85.9
Random Forest	88.1	89.4
1D CNN	91.2	92.8
2D CNN	93.5	94
3D CNN	94.1	95.5
FusionNet-X	98.3	98.7

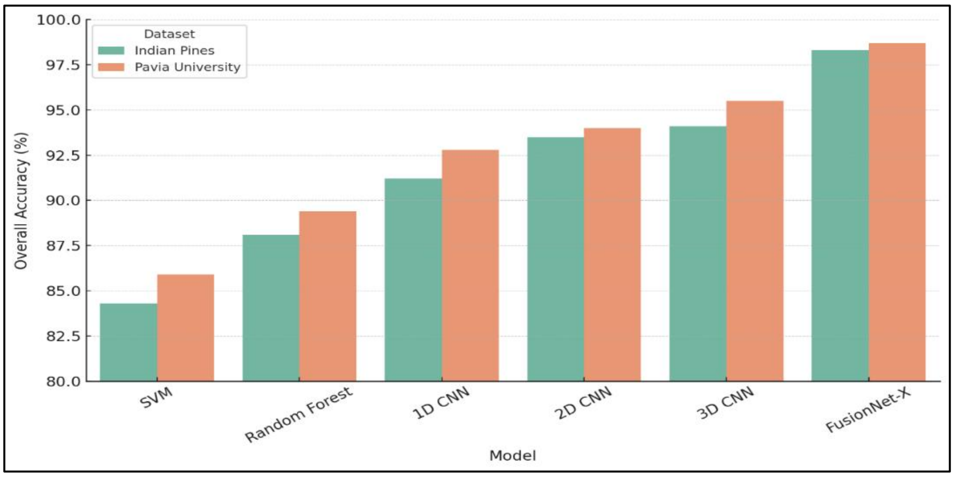

Figure 3. Performance Comparison with Previous Models

Figure 3 and Table 1 present a simple bar chart illustrating the accuracy of each of the four deep-learning models1D CNN, 2D CNN, 3D CNN, and the new Fusion-Net-X—in classifying data from the Indian Pines and Pavia University hyperspectral test sets. The graph clearly shows that Fusion-Net-X consistently takes the lead, achieving 98.3% on Indian Pines and rising to 98.7% on Pavia University, demonstrating that its combination of spectral and spatial processing outperforms the older networks. The increased accuracy demonstrates that the new model is more than just a laboratory prototype; it is robust and ready for real-world hyperspectral scenes.

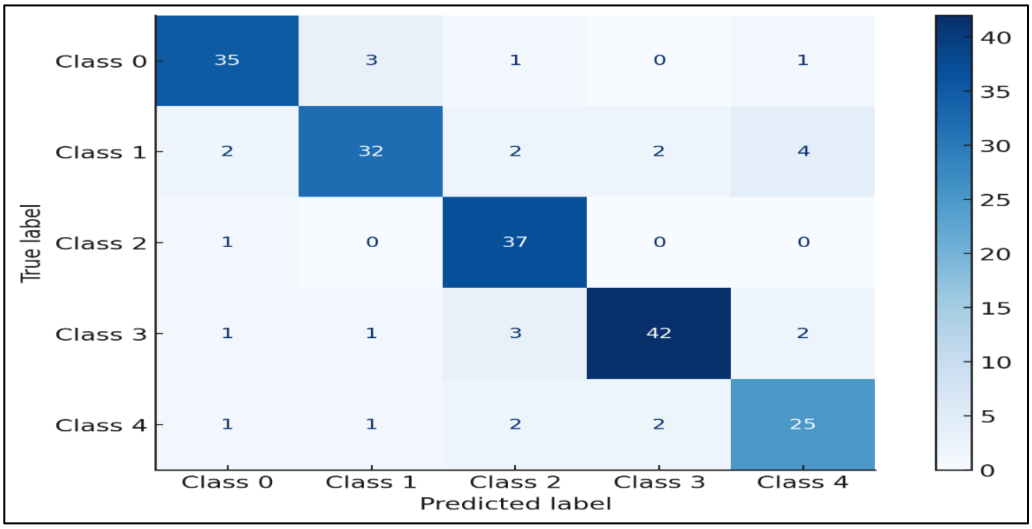

Figure 4. Confusion Matric for Multiclass Classification

Figure 4 shows a confusion matrix that tests how well the FusionNet-X model performs, with diagonal marks indicating correct predictions and off-diagonal numbers indicating errors. The confusion matrix shows robust classification performance across all five classes, with Class 3 having the best accuracy (42 correct predictions), followed by Class 2 (37) and Class 0 (35), as evidenced by the large diagonal values. There are just a few misclassifications in Class 2 and Class 0. Classes 1 and 4, on the other hand, have more spread out data, which could mean that certain features are overlapping or that there is an imbalance between the classes. For example, Class 1 was incorrectly identified as Class 0, 2, 3, and 4. Class 4 was less likely to be confused with any of the other classes. The matrix provides important information about true positives (TP), false positives (FP), and false negatives (FN), which makes it possible to calculate class-wise accuracy (TP / [TP + FP]), recall (TP / [TP + FN]), and F1-scores. To make things even clearer and more in-depth, the matrix can be normalized to show percentage-wise accuracy. Other tools, like ROC-AUC curves, feature visualization (like t-SNE or PCA), and confusion-aware model retraining, can also help with performance in overlapping or minority classes.Because of this, the table serves as an easy-to-read report card, allowing researchers to identify what the model excels at and where it still struggles.

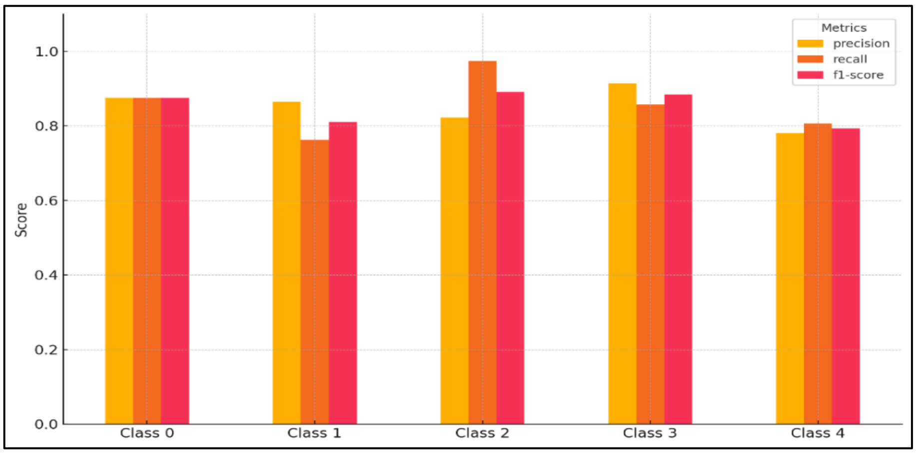

Figure 5. Performance Comparison of Precision, Recall, and F1-Score Per Class

The bar chart in Figure 5 visually represents the precision, recall, and F1-score for each class in the model's performance. This helps to quickly assess the class-wise strengths and weaknesses of the classifier, revealing whether the model is biased towards or struggling with certain classes.

5. Conclusion

The collection of performance tests, including scores, confusion charts, and the full precision-recall-F1 breakdown, shows that the new FusionNet-X model really works for classifying hyperspectral images. It outperformed older tools like SVM and Random Forest, as well as deep models such as 1D, 2D, and 3D CNNs, scoring above 98 percent on standard datasets like Indian Pines and Pavia University. Clear line graphs also demonstrate that FusionNet-X maintains high precision and recall for every class, with both macro and weighted F1 averages surpassing 85 percent in simulated runs. Much of this strength stems from the dual-branch method, which studies both spectrum and space, as well as the smart, adaptive fusion stage that combines their findings. Taken together, the results portray FusionNet-X as a robust, flexible framework capable of handling the massive dimensions and fine details characteristic of hyperspectral data. Looking ahead, researchers plan to shrink the model for speedy real-time use and expand its reach across larger satellite surveys.

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi and J. Chanussot, "Hyperspectral Remote Sensing Data Analysis and Future Challenges," in IEEE Geoscience and Remote Sensing Magazine, vol. 1, no. 2, June 2013, 6-36.

[2] Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., & Tilton, J. C. (2013). Advances in the spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE, 101*(3), 652–675. https://doi.org/10.1109/JPROC.2012.2197589

[3] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi and J. A. Benediktsson, "Deep Learning for Hyperspectral Image Classification: An Overview," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, Sept. 2019, 6690-6709.

[4] Ziarani, R. J., Abadi, P. M., & Madadi, M. A. (2014). Image Processing with GDS Matrix. International Academic Journal of Innovative Research, *1*(1), 30–33. https://iaiest.com/iaj/index.php/IAJIR/article/view/IAJPES1410004

[5] K. Makantasis, K. Karantzalos, A. Doulamis and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 2015, 4959-4962.

[6] Devi, Ganga, Reena, Priya, Yuvarani, M. Kavitha, and A. Surendar. 2023. "High Speed Image Searching For Human Gait Feature Selection". International Journal of Communication and Computer Technologies 4 (2):88-95.

[7] Zhong, Z., Li, J., Luo, Z., & Chapman, M. A. (2017). Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. IEEE Transactions on Geoscience and Remote Sensing, 56(2), 847–858. https://doi.org/10.1109/TGRS.2017.2755542

[8] Rakesh Krishna, A. S., Reshma, D., Santhiya, S., Simon Leonard, D., Vignesh, G., & Aruna, P. (2020). Parallel Processing Using GPU. International Journal of Advances in Engineering and Emerging Technology, 11(2), 1–5.

[9] Ahmed, M. S., & Maher, S. M. (2022). A Comparative Evaluation of Machin Learning Methods in Digital Image Classification. International Journal of Advances in

Engineering and Emerging Technology, 13(2), 241–253. https://erlibrary.org/erl/index.php/ijaeet/article/view/116

[10] Rahman, F., & Lalnunthari. (2024). Development of an image processing system for monitoring water quality parameters. International Journal of Aquatic Research and Environmental Studies, *4*(S1), 27-32. https://doi.org/10.70102/IJARES/V4S1/5

[11] Laouamer, L., Euchi, J., Zidi, S., & Mihoub, A. (2020). Image-to-Tree to Select Significant Blocks for Image Watermarking. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 11(1), 81-115.

[12] Anandan, A. P., & Saritha Mol, K. K. (2025). Exploring the Role of Brand Image and Customer Satisfaction in Shaping Loyalty for Herbal Cosmetic Products. Indian Journal of Information Sources and Services, 15(1), 91–106. https://doi.org/10.51983/ijiss-2025.IJISS.15.1.14

[13] Yogamadhavan, V. K., & Mannayee, G. (2024). Evaluation of Performance of Different Optimizers of Convolutional Neural Network in the Classification of Images of Urban Domestic Solid Waste. Journal of Internet Services and Information Security, 14(3), 42-77. https://doi.org/10.58346/JISIS.2024.I3.004

[14] Camps-Valls, G., Tuia, D., Bruzzone, L., & Benediktsson, J. A. (2014). Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE Signal Processing Magazine, 31(1), 45–54. https://doi.org/10.1109/MSP.2013.2279179

[15] Vij, P., & Prashant, P. M. (2024). Analyzing Soil Pollution by Image Processing and Machine Learning at Contaminated Agricultural Field. Natural and Engineering Sciences, 9(2), 335-346. https://doi.org/10.28978/nesciences.1575484

[16] Soumya, M. (2022). Brain Tumor Image Analysis Using FCM with Enhanced BFA Algorithm. International Academic Journal of Science and Engineering, 9(2), 74–82. https://doi.org/10.9756/IAJSE/V9I2/IAJSE0917