# HYPERSPECTRAL IMAGE CLASSIFICATION USING MULTI-LAYER PERCEPTRON MIXER (MLP-MIXER)

**Ali Jamali [1], Masoud Mahdianpari [2,3*], and Alias Abdul Rahman [4]**

[1] Faculty of Engineering, University of Karabük, Karabük, Turkey
[2] Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John’s, NL A1B3X5, Canada
[3] C-CORE, 1 Morrissey Rd, St. John’s, NL A1B 3X5, Canada.
[4] Universiti Teknologi Malaysia (UTM), Faculty of Built Environment and Surveying.

**Commission IV, WG 7**

**KEYWORDS:** LULC Mapping, Big data, Hyperspectral, Image Classification, Machine Learning, Multi-layer Perceptron

**ABSTRACT**

The classifying of hyperspectral images (HSI) is a difficult task given the high dimensionality of the space, the huge number of spectral bands, and the small number of labeled data. As such, we offer a unique hyperspectral image classification methodology to address these issues based on sophisticated Multi-Layer Perceptron (MLP) algorithms. In this paper, we propose using MLP-Mixer to classify HSI data in three data benchmarks of Pavia, Salinas, and Indian Pines. Based on the results, the proposed MLP-Mixer achieved a high level of classification accuracy and produced noise-free and homogenous classification maps in all study areas. For the classification of HSI data in Salinas, Indian Pines, and Pavia, the proposed MLP-Mixer achieved an average accuracy of 99.82%, 99.81%, and 99.23%, respectively.

## 1. INTRODUCTION

Advances in image acquisition techniques have resulted in higher spatial-spectral image resolutions, more efficient image processing models, and the continuous creation of large amounts of high-quality data. As a result, low-cost, high-quality data from sensors, combined with the availability of advanced computing resources, including graphics processing units (GPUs), has resulted in superior computer models that allow researchers to better understand the morphological changes, ground surface, and human processes with greater precision and detail (Fu et al., 2017). Image classification and semantic segmentation are both well-studied sub-domains of computer vision that have been used in RS image analysis (A. Jamali et al., 2022; Jamali et al., 2021; Jamali and Mahdianpari, 2022), image segmentation and face recognition, among other applications. Image classification groups and categorizes all objects in an image into a single class, whereas semantic segmentation assigns each pixel in an image to a set of predetermined labels/classes, where the same labels share certain properties (Kemker et al., 2018).

Land cover mapping and change detection (Akar and Tunc Gormus, 2021; Q. Li et al., 2021), soil carbon prediction (Meng et al., 2022), vegetation classification (Gao et al., 2022), forest biomass understanding, and tree species identification/mapping (Vangi et al., 2021), urban monitoring and understanding (Wambugu et al., 2021), and other applications benefit from the rich spatial-spectral data gained from Hyperspectral (HIS) data. However, labeling HSI data is expensive and labor-intensive due to the various sensors utilized to gather the data and the domain expertise required. This explains why there are so few labeled HSI benchmark datasets, causing HSI classification to lag behind other vision-based and image-processing areas due to the scarcity of annotated labels and the intricate nature of HSI data (X. Cao et al., 2018). As such, the use of deep learning models such as deep and very deep Convolutional Neural Networks (CNNs) is a challenge in HSI data classification.

On the other hand, conventional classifiers, such as random forest and support vector machines, have had remarkable achievements in HSI classification (Belgiu and L. Drăguţ, 2016), but they only use spectral data, limiting their maximum efficiency. Furthermore, the high dimensionality of spatial-spectral data, combined with a small number of training examples, makes it difficult to increase classification performance (Z. Zhong et al., 2018). As such, to improve the classification accuracy of HIS data, in this paper, we will investigate the use of Multi-Layer Perceptron (MLP)-Mixer as one of the current advanced MLP algorithms for the classification of HSI data. It should be noted that the MLP-Mixer considers both spectral and spatial information in the classification process.

## 2. METHODS

### 2.1. MLP-Mixer

The architecture of the MLP-Mixer (Tolstikhin et al., 2021) is shown in Figure 1. CNNs are the computer vision model of choice at the moment. Moreover, attention-based networks, such as the Vision Transformer, have recently gained popularity. Although, the MLP-Mixer demonstrated that, while convolutions and attention are both effective for decent performance, neither is required. MLP-Mixer takes as input a succession of linearly projected image patches (also known as tokens) shaped as a "patches-channels" table and preserves this dimensionality. Mixer employs two sorts of MLP layers:

channel-mixing and token-mixing. MLPs that blend channels enable communication between them (Tolstikhin et al., 2021). They function independently on each token and take individual rows of the table as input. Token-mixing MLPs enable communication between distinct spatial locations (tokens); they work independently on each channel and accept specific table columns as inputs. Interleaving these two types of layers allows interaction of both input dimensions (Tolstikhin et al., 2021).

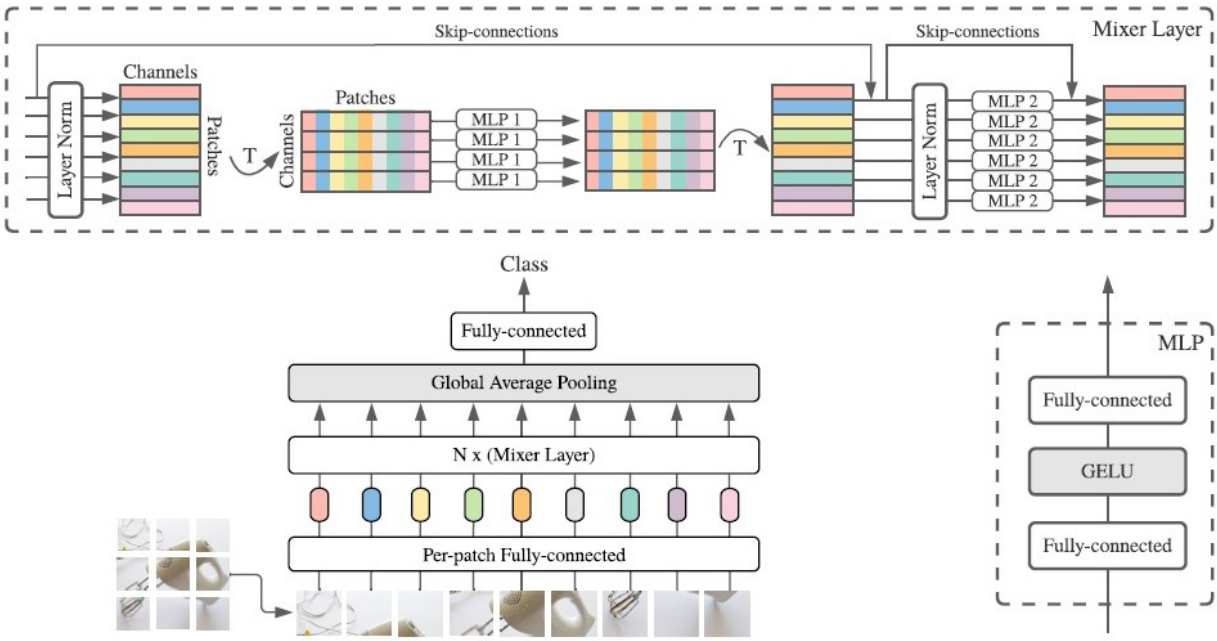


**Figure. 1**. The architecture of the MLP-Mixer. Per-patch linear embeddings, Mixer layers, and a classifier head comprise MLP-Mixer. One token-mixing MLP and one channel-mixing MLP, each with two fully connected layers and a GELU nonlinearity, are present in the mixer layers. Other features are skip-connections, dropout, and channel layer norm (Tolstikhin et al., 2021).

2.2 Study area and remote sensing data

We conduct experiments on three different HSI data benchmarks to assess the efficacy and performance of the proposed classifier:

- The Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS) sensor gathered the Indian Pines HSI, which represents northwestern Indiana. It has 145 by145 pixels with a spatial resolution of 20 m per pixel and 220 spectral bands. The ground truth data is divided into 16 classes.
- The AVIRIS sensor captured the Salinas image over Salinas, California, which consists of 512 by 217 pixels with a spatial resolution of 3.7 m per pixel and 224 spectral bands. The ground truth data is divided into 16 classes.
- The final HSI is Pavia University data taken using the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor, which has 610 by 610 pixels and 115 spectral bands, with a spatial resolution of 1.3 m. There are nine classes in the ground truth data.

2.3 Accuracy assessment

The classification results of the MLP-Mixer are evaluated in terms of overall accuracy, kappa index, and average accuracy (see Equations 1-3).

$$Overall\ Accuracy = \frac{(TP+TN)}{Total\ number\ of\ pixels} \times 100 \quad (1)$$

$$Kappa = \frac{p_0 - p_e}{1 - p_e}, p_0 = \frac{\sum x_{ii}}{N}, p_e = \frac{\sum x_{i+} x_{i+}}{N^2} \quad (2)$$

$$Average\ Accuracy = \frac{\sum_{i=1}^{n} Recall_i}{n},\ Recall = \frac{\text{True positive}}{(True\ positive + False\ negative)} \quad (3)$$

## 3. RESULTS AND DISCUSSION

To better understand the performance of the proposed MLP-mixer for HSI classification, its predicted classification maps are presented in Figure. 2. Moreover, statistical results of the MLP-Mixer in terms of average accuracy, kappa index, and overall accuracy are shown in Table 1. As seen in Figure 2 and Table 1, the proposed classifier illustrated great capability in the classification of HSI data in all three data benchmarks of Indian Pines, Salinas, and Pavia. For the Indian Pines dataset, the MLP-Mixer achieved kappa index, overall accuracy, and average accuracy of 99.78%, 99.8%, and 99.81%,

respectively. Moreover, average accuracy, kappa index, and overall accuracy of 99.23%, 99.49%, and 99.62% were obtained by the MLP-Mixer, respectively, for the Pavia dataset. In addition, the MLP-Mixer obtained kappa index, overall accuracy, and average accuracy of 99.7%, 99.73%, and 99.82%, respectively, for the Salinas dataset.

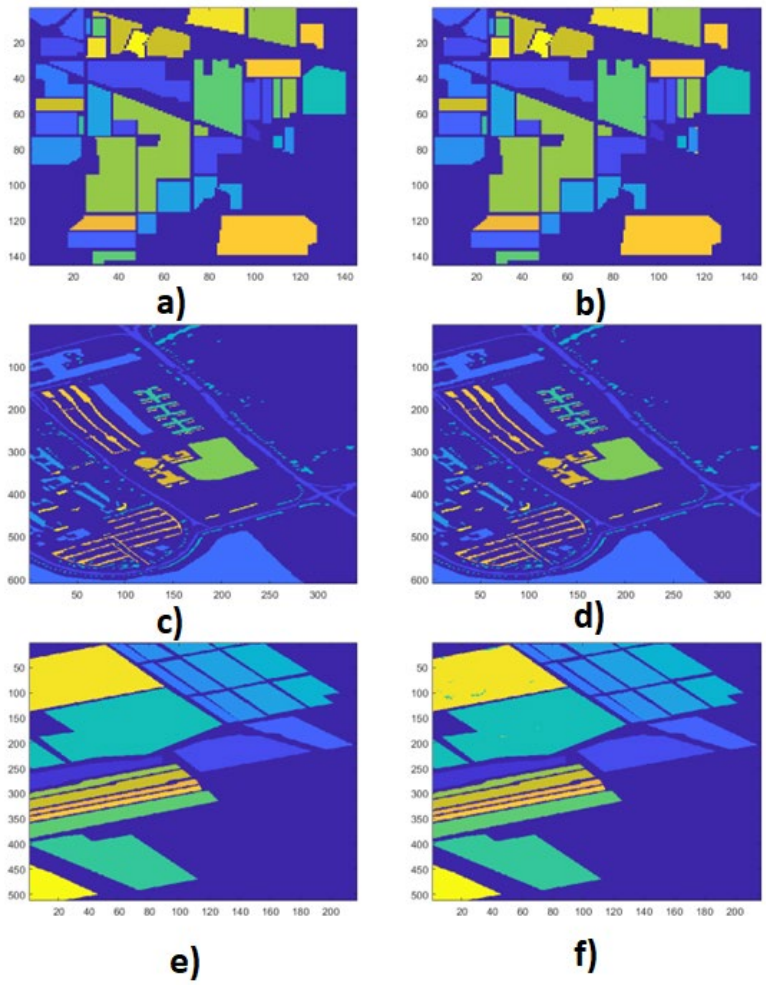


**Figure. 2**. Ground truth data and results of the MLP-Mixer; a) Indian Pines reference data, b) Classified Indian Pines HIS data c) Pavia reference data d) Classified Pavia HIS data e) Salinas reference data f) Classified Salinas HIS data.

**Table 1**. Classification results of HIS data by the MLP-Mixer (KI= Kappa Index, OA= Overall Accuracy, AA= Average Accuracy)

| Data | MLP-Mixer |
|---|---|
| **Indian Pines** | |
| KI (%) | 99.78 |
| OA (%) | 99.8 |
| AA (%) | 99.81 |
| **Pavia** | |
| KI (%) | 99.49 |
| OA (%) | 99.62 |
| AA (%) | 99.23 |
| **Salinas** | |
| KI (%) | 99.7 |
| OA (%) | 99.73 |
| AA (%) | 99.82 |

## 4. CONCLUSIONS

This paper proposes a novel approach for spatial-spectral classification of HSI data based on the current state-of-the-art MLP algorithms. The proposed classifier of MLP-Mixer demonstrated excellent capability in the classification of HSI data in three data benchmarks of Pavia, Indian Pines, and Salinas. The presented MLP-Mixer achieved an average accuracy of 99.82%, 99.81%, and 99.23% for the classification of HSI data in Salinas, Indian Pines, and Pavia, respectively. Our method can be compared to other MLP classifiers in future works.

## REFERENCES

A. Jamali, M. Mahdianpari, F. Mohammadimanesh, A. Bhattacharya, S. Homayouni, 2022. PolSAR Image Classification Based on Deep Convolutional Neural Networks Using Wavelet Transformation. IEEE Geoscience and Remote Sensing Letters 19, 1–5. https://doi.org/10.1109/LGRS.2022.3185118

Akar, O., Tunc Gormus, E., 2021. Land use/land cover mapping from airborne hyperspectral images with machine learning algorithms and contextual information. null 1–28. https://doi.org/10.1080/10106049.2021.1945149

Belgiu, M., L. Drăguţ, 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. Remote Sensing 9. https://doi.org/10.3390/rs9050498

Gao, Y., Song, X., Li, W., Wang, J., He, J., Jiang, X., Feng, Y., 2022. Fusion Classification of HSI and MSI Using a Spatial-Spectral Vision Transformer for Wetland Biodiversity Estimation. Remote Sensing 14. https://doi.org/10.3390/rs14040850

Jamali, A., Mahdianpari, M., 2022. Swin Transformer and Deep Convolutional Neural Networks for Coastal Wetland Classification Using Sentinel-1, Sentinel-2, and LiDAR Data. Remote Sensing 14. https://doi.org/10.3390/rs14020359

Jamali, A., Mahdianpari, M., Brisco, B., Granger, J., Mohammadimanesh, F., Salehi, B., 2021. Deep Forest classifier for wetland mapping using the combination of Sentinel-1 and Sentinel-2 data. GIScience & Remote Sensing 1–18. https://doi.org/10.1080/15481603.2021.1965399

Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 145, 60–77. https://doi.org/10.1016/j.isprsjprs.2018.04.014

Meng, X., Bao, Y., Wang, Y., Zhang, X., Liu, H., 2022. An advanced soil organic carbon content prediction model via fused temporal-spatial-spectral (TSS) information based on machine learning and deep learning algorithms. Remote Sensing of Environment 280, 113166. https://doi.org/10.1016/j.rse.2022.113166

Q. Li, H. Gong, H. Dai, C. Li, Z. He, W. Wang, Y. Feng, F. Han, A. Tuniyazi, H. Li, T. Mu, 2021. Unsupervised Hyperspectral Image Change Detection via Deep Learning Self-Generated Credible Labels. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 9012–9024. https://doi.org/10.1109/JSTARS.2021.3108777

Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A., 2021. MLP-Mixer: An all-MLP Architecture for Vision, in: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 24261–24272.

Vangi, E., D’Amico, G., Francini, S., Giannetti, F., Lasserre, B., Marchetti, M., Chirici, G., 2021. The New Hyperspectral Satellite PRISMA: Imagery for Forest Types Discrimination. Sensors 21. https://doi.org/10.3390/s21041182

Wambugu, N., Chen, Y., Xiao, Z., Tan, K., Wei, M., Liu, X., Li, J., 2021. Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review. International Journal of Applied Earth Observation and Geoinformation 105, 102603. https://doi.org/10.1016/j.jag.2021.102603

X. Cao, R. Li, L. Wen, J. Feng, L. Jiao, 2018. Deep Multiple Feature Fusion for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 3880–3891. https://doi.org/10.1109/JSTARS.2018.2866595

Z. Zhong, J. Li, Z. Luo, M. Chapman, 2018. Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. IEEE Transactions on Geoscience and Remote Sensing 56, 847–858. https://doi.org/10.1109/TGRS.2017.2755542