PERFORMANCE COMPARISON OF MULTISPECTRAL CHANNELS FOR LAND USE CLASSIFICATION

Tejasri Nampally[1], *Jiantao Wu*[2,3], *Soumyabrata Dev*[2,3]

[1]Department of Artificial Intelligence, Indian Institute of Technology Hyderabad, India
[2]The ADAPT SFI Research Centre, Dublin, Ireland
[3]School of Computer Science, University College Dublin, Ireland

ABSTRACT

Land cover classification using satellite imagery plays a crucial role in monitoring changes on the earth's surface. This paper presents an analysis of the EuroSAT dataset using state-of-the-art deep learning models to benchmark the impact of additional bands on classification accuracy. The dataset consists of 27,000 images across 10 classes captured by the Sentinel-2 satellite, including RGB and multispectral bands. Performance evaluation was conducted using popular convolutional neural network models based on Resnet variants and Vision Transformer (ViT). The results show that the combination of all bands achieved the highest accuracy, with ResNet-152 achieving a validation accuracy of 96.63% on the multispectral dataset. Precision, recall, and F1 scores were also utilized to assess the models' performance. The findings highlight the significance of incorporating additional bands for improved classification accuracy in satellite image analysis.

Index Terms— Land cover classification, Satellite imagery, EuroSAT dataset, Sentinel-2 satellite.

1. INTRODUCTION

Land cover classification through satellite imagery has become a crucial tool for monitoring changes on the earth's surface [1]. It involves categorizing different features such as vegetation, infrastructure, water bodies [2] and soil to analyze patterns and detect changes over time. Image classification techniques have played a vital role in remote sensing and image analysis, enabling researchers to study and monitor various aspects of our planet, including climate change [3], forest cover mapping, pollution [4], wetland mapping [5] and land cover analysis [6]. Remote sensing data availability has dramatically increased over the past decade, increasing the demand for effective deep learning-based image processing and analysis techniques. The frequency of launch of satellites has increased to 300 in 2017 and 2018. Many of them are image satellites used for either commercial or earth observation activities, such as the Copernicus program of the European Space Agency. As a result of this, large data sets such as the EuroSAT dataset [7] are publicly available.

In the past decade, significant progress has been made in parallel GPU computing, facilitating the shift from CPU-based to GPU-based training methods. This transition has played a crucial role in enabling the training of deep neural networks. The usage of deep learning (DL) techniques has led to the development of increasingly sophisticated and intelligent algorithms, surpassing human performance levels. [8]. Convolutional Neural Networks (CNNs) have played a significant role in the evolution of deep learning methods. These have shown remarkable success in image classification tasks due to their ability to capture spatial features in images [9]. CNNs have been widely deployed in various geoscience applications such as land cover classification, canopy classification etc, [10] [11]. Vision transformers (ViT) have shown great potential in image classification tasks and have outperformed Convolutional Neural Networks (CNN) on several benchmark datasets. ViT is effective due to its ability to capture long-range dependencies in images and incorporate global information [12]. In addition to RGB channels, multispectral data has also been of paramount importance with its additional bands, such as near-infrared, red-edge etc., that provide hidden underlying information for effective image classification. DL-based methods are well known to be integrated with multispectral image data. CNNs are known to perform better in classification tasks using multispectral data compared to RGB data [13]. Motivated by the rapid expansion of remote sensing technology and DL methods, in this work, we have analyzed EuroSAT RGB and multispectral datasets using state-of-the-art DL models for benchmarking purposes. The aim is to demonstrate the impact of augmenting the dataset with additional bands on classification accuracy. To compare the dataset containing blue, green, red and near-infrared (RGB-NIR) dataset with the red, green, blue

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

Send correspondence to S. Dev: `soumyabrata.dev@ucd.ie`

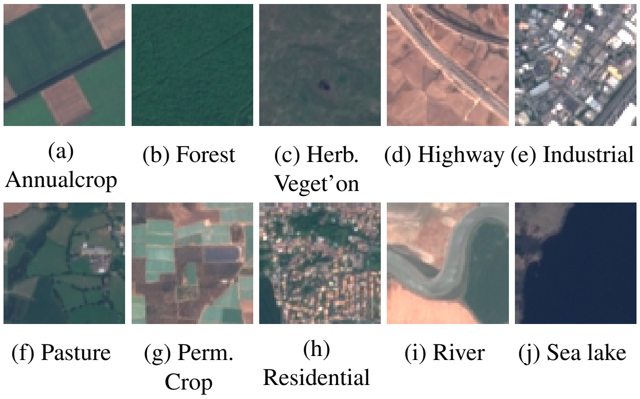
(RGB) dataset, the former was obtained by incorporating both RGB and NIR bands from the 13-band dataset. Similarly, to evaluate the performance of the multispectral dataset against a dataset with fewer bands, the 13-band dataset was utilized. This paper is organized into sections as follows. Section 2 discusses the contents of the dataset and the models used. Section 3 analyses the obtained results, and provides information on the metrics used to evaluate the models followed by future work and conclusion.

The contributions of this paper [1] are as follows:

- We have implemented and compared the results of local-based CNN models based on Resnet and global-based Vision Transformer on the EuroSAT dataset.
- We examined the impact of combining different spectral bands in the context of image classification.
- All the models are trained from scratch in PyTorch and the code is made available to the public.

2. DATASET & METHODOLOGY

EuroSAT dataset is a popular remote sensing dataset that consists of 10 classes such as Annual Crop, Forest, Herbaceous Vegetation, Highway, Industrial buildings, Pasture, Permanent Crop, Residential buildings, River and Sea and Lake with 2,000 to 3,000 images per class, adding up to 27,000 images. The dataset covers 13 different spectral bands such as red, green, blue, near-infrared (NIR), red-edge etc., captured by Sentinel-2 satellite. Sample belonging to each category is shared in Figure 1. Data augmentation is a commonly used technique to enhance the generalization ability of neural networks to hidden data and to augment the training dataset.


(a) Annualcrop (b) Forest (c) Herb. Veget'on (d) Highway (e) Industrial

(f) Pasture (g) Perm. Crop (h) Residential (i) River (j) Sea lake

Fig. 1: Representation of classes in the Sentinel-2 EuroSAT RGB image dataset.

In this work, we compared the performance of the state-of-the-art convolutional neural network based models such as Resnet-50, Resnet-101, Resnet-152 [14] and Vision Transformer [15] were implemented on the EuroSAT RGB dataset.

In order to ensure fairness, identical hyper-parameters were employed to train the ViT model on all datasets. To improve generalization, several data augmentation techniques, such as resizing, normalization, flipping, and rotation, were applied uniformly across the datasets. The performance and accuracy of satellite image classification are strongly influenced by the choice of model. The ViT model segments the images into position-embedded patches that are then processed by the transformer encoder, enabling the model to capture both local and global features of the image.

3. RESULTS & DISCUSSION

3.1. Model selection and configuration

Resnet: Resnet [14] became popular as it introduced the concept of residual connections, which enabled the training of exceptionally deep neural networks. ResNet mitigated the vanishing gradient problem and facilitated the flow of gradients throughout the network through residual connections. It exhibited superior performance and achieved lower error rates on benchmark image classification datasets, such as ImageNet [16].

Vision Transformer: ViT [15] has emerged as a prominent architecture in the field of deep learning, specifically for image classification tasks. Unlike traditional CNNs, which rely on convolutional layers for spatial feature extraction, ViT adopts a transformer-based architecture inspired by the success of transformers in natural language processing tasks. It uses a self-attention mechanism that enables the model to capture context and long-range dependencies across the entire image, facilitating a more holistic understanding of visual information. It can effectively handle both local and global image features without the need for handcrafted design choices such as filter sizes or pooling operations which makes ViT suitable for images of varying sizes and resolutions.

We followed 80-20% training-to-validation ratio. We trained and validated the data with CNN Based models such as Resnet-50, Resnet-101, Resnet-152 and Vision Transformer (ViT). The experimental setup is maintained consistent across all three experiments (RGB, RGB-NIR, and Multispectral). The experiments were conducted using PyTorch framework with a batch size of 32. Cross entropy loss function and Adam optimizer were used with a learning rate of 0.001. Local based CNN models such as Resnet-50, Resnet-101, Resnet-152 and global based Vision Transformer were trained and validated. The input image is 64 x 64. In case of ViT, the image was broken down to 16 patches hence, the number of transformer layers for ViT is taken as 16. All the models were trained from scratch.

[1] In the spirit of reproducible research, all the code and dataset used in this paper can be accessed here: `https://github.com/tejasri19/EuroSAT_data_analysis`

Table 1: Performance analysis of models on EuroSAT multispectral dataset.

Model	RGB				RGB-NIR			
	Train Accu.	Train loss	Val Accu.	Val. loss	Train Accu.	Train loss	Val. Accu.	Val. loss
ResNet-50	**98.33%**	**0.051**	88.537%	0.460	97.903%	0.061	89.130%	0.491
ResNet-101	97.78%	0.071	86.074%	0.548	**98.620%**	**0.041**	**92.037%**	**0.324**
Resnet-152	96.54%	0.098	87.778%	0.5398	98.241%	0.052	89.333%	0.435
ViT	69%	0.863	66.13%	0.962	74.056%	0.835	71.006%	0.813

Table 2: Performance analysis of models on EuroSAT multispectral dataset.

Model	Multispectral			
	Train Accu.	Train loss	Val Accu.	Val Loss
ResNet-50	98.069%	0.061	91.130%	0.418
ResNet-101	98.876%	0.233	95.276%	0.218
Resnet-152	**99.120%**	**0.026**	**96.630%**	**0.147**
ViT	71.62%	0.801	73.574%	0.779

The training and validation results are tabulated in Tables 1 and 2. It can be noted that Resnet-50 performed better with 0.051 training loss and 88.537% validation accuracy. In case of dataset with RGB and NIR bands, Resnet-101 achieved 0.041 training loss and 92.037% validation accuracy. Resnet-152 achieved superior results compared with other variants and ViT when all the bands are used, with 0.026 training loss and 96.630% validation accuracy.

3.2. Impact of band combination on classification performance

The RGB bands of an image contain valuable information about its vibrant color features, while the NIR (Near Infrared) band provides details about the sharp edges present. This highlights the significance of including additional bands in a dataset, as it enhances the amount of information available and consequently improves the performance of machine learning models.

In order to assess the performance of the model, several metrics such as accuracy, precision, recall, and F1 score provided in equations 1, 2 and 3 were utilized. These metrics provide valuable insights into the classification errors that may have occurred. The validation data from each class were used to calculate these metrics. The performance of models is tabulated in Tables 1, 2.

In Table 3, it is demonstrated that, for most classes, the combination of all bands of EuroSAT data yielded superior results with precision of 0.9654 and recall of 0.9651 with the Resnet-152 model. This indicates that by using all images, we can achieve enhanced classification accuracy compared to using RGB and RGB-NIR images alone. With ViT, it can be noted that precision and recall are highest when all the bands were used.

$$Recall = \frac{TP}{(TP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$F1 - score = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \tag{3}$$

where TP denotes true positives, FN corresponds to false negatives and FP to false positives. It can be inferred that, with increase in number of layers and quality and quantity of data, there is improvement in the performance of DL models. From Table 4, it can be noted that training time and parameters of the DL model proportionally increases to the number of bands included in the dataset and number of layers of the model used in case of Resnet.

4. CONCLUSION & FUTURE WORK

In this paper, we have compared the performance of CNN-based neural networks and ViT trained on RGB, RGB-NIR and multispectral data. The results demonstrated that augmenting the number of bands enhances the performance of the model, highlighting the significant impact of incorporating spectral bands on neural networks. In future works, we extend to propose a novel method to select the efficient bands of multispectral data of the EuroSAT dataset for better classification accuracy. A potential contribution will be a training strategy capable of trading some performance off in comparison to the state-of-the-art for more advantages in terms of training speed.

5. REFERENCES

[1] C. O'Sullivan, S. Coveney, X. Monteys, and S. Dev, "Analyzing water body indices for coastal semantic segmentation," in *Proc. Photonics & Electromagnetics Research Symposium (PIERS)*. IEEE, 2023.

[2] J. Wu, F. Orlandi, D. O'Sullivan, and S. Dev, "Linkclimate: An interoperable knowledge graph platform for climate data," *Computers & Geosciences*, vol. 169, p. 105215, 2022.

Table 3: Classification results of DL models on EuroSAT multispectral dataset.

Model	**RGB**			**RGB & NIR**			**Multispectral**		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
ResNet-50	**0.8881**	**0.8860**	**0.8870**	0.9036	0.8884	0.8959	0.9285	0.9049	0.9164
ResNet-101	0.8620	0.8577	0.8598	**0.9205**	**0.9181**	**0.9192**	0.9519	0.9523	0.9520
Resnet-152	0.8791	0.8733	0.8761	0.9054	0.8895	0.8973	**0.9654**	**0.9651**	**0.9652**
ViT	0.5583	0.5283	0.5427	0.7352	0.7345	0.7348	0.8282	0.8048	0.8162

Table 4: Parameter size and Training time of DL models.

Model	**RGB**		**RGB-NIR**		**Multispectral**	
	Train time (s)	Param. size (mb)	Train time (s)	Param. size (mb)	Train time (s)	Param. size (mb)
Resnet-50	764.86	113.21	774.03	113.23	868.23	113.48
Resnet-101	1291.04	197.34	1324.77	197.37	1418.04	197.62
Resnet-152	1840.37	271.46	2141.19	271.48	1935.55	271.73
ViT	1655	355.99	1668.81	356.76	1746.69	363.65

[3] J. Wu, F. Orlandi, D. O'Sullivan, and S. Dev, "Publishing climate data as linked data via virtual knowledge graphs," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 4090–4093.

[4] J. Wu, F. Orlandi, I. Gollini, E. Pisoni, and S. Dev, "Uplifting air quality data using knowledge graph," in *2021 Photonics & Electromagnetics Research Symposium (PIERS)*. IEEE, 2021, pp. 2347–2350.

[5] C. O'Sullivan, S. Coveney, X. Monteys, and S. Dev, "Interpreting a semantic segmentation model for coastline detection," in *Proc. Photonics & Electromagnetics Research Symposium (PIERS)*. IEEE, 2023.

[6] G. Kaplan and U. Avdan, "Mapping and monitoring wetlands using sentinel-2 satellite imagery," 2017.

[7] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[10] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 173, pp. 24–49, 2021.

[11] G. U. Sai, N. Tejasri, A. Kumar, and P. Rajalakshmi, "Deep learning based overcomplete representations for paddy rice crop and weed segmentation," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 6077–6080.

[12] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?" in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 12 116–12 128.

[13] P. J. Navarro, L. Miller, A. Gila-Navarro, M. V. Díaz-Galián, D. J. Aguila, and M. Egea-Cortines, "3DeepM: An ad hoc architecture based on deep learning methods for multispectral image classification," *Remote Sensing*, vol. 13, no. 4, p. 729, 2021.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 248–255.