COMPARATIVE ANALYSIS OF CONVOLUTIONAL NEURAL NETWORKS (CNN) FOR LAND USE CLASSIFICATION BASED ON AGRICULTURAL SATELLITE IMAGES

Manuscript Info

……………………….

Manuscript History

Received: xxxxxxxxxxxxxxxx
Final Accepted: xxxxxxxxxxxx
Published: xxxxxxxxxxxxxxxx

Abstract

………………………………………………………………

The use of satellite imagery has grown considerably in recent years for land cover detection, particularly for agricultural land classification. This article aims to address these challenges by applying several machine learning algorithms to multispectral data from the Sentinel-2 satellite to obtain accurate land classification models. This study presents a comparative analysis of five convolutional neural networks (CNN) architectures for the automatic classification of land use from satellite images. The models evaluated include ResNet50, MobileNet, VGG16, GoogleNet and EfficientNet. These were applied to a dataset comprising four land use classes: Annual Crop, Forest, Permanent Crop and Residential. The results show that MobileNet and GoogleNet perform best with a validation accuracy of 99%, whereas ResNet50 is limited to 66%. This underlines the importance of using advanced machine learning techniques, including MobileNet or GoogleNet, to accurately classify changes in agricultural land use.

Introduction: -

Faced with global population growth, food security has become a major concern for policymakers and researchers. Ensuring a reliable and sustainable food supply requires a multidimensional approach that considers various factors, including agricultural practices, supply chain management, policy frameworks, economic factors, and factors incorporating new information technologies. Indeed, satellite imagery has become indispensable for classifying land cover and land use in various fields.

It plays a crucial role in natural resource management [1] and has a significant impact on sectors such as forestry, agriculture, land-use planning, and biological resources, while also providing essential data on water, the atmosphere, and soil erosion [2], [3], [4], [5]. This helps us better understand and monitor changes in our natural environment [6], [7]. This organized categorization facilitates the understanding and management of different land-use patterns. Thanks to advances in deep learning technologies, this classification is becoming more effective. Classification using the CNN algorithm derived from deep learning offers valuable insight into Earth's environmental dynamics by systematically organizing its surface into different soil types, thus creating a detailed map of the various land covers. Automatic land-use classification from satellite imagery represents a significant challenge for environmental monitoring, land-use planning, and natural resource management. The increasing availability of high-resolution satellite data allows deep learning techniques, and in particular convolutional neural networks (CNNs), to prove their effectiveness [8] and [9]. This research article aims to compare the performance of five common CNN architectures for land cover classification into four categories: annual crops, forests, permanent crops, and residential areas. The objective is to identify the most suitable architecture for this specific application and to propose and evaluate the best solution for classifying agricultural practices and land cover from satellite imagery. First, we focus on identifying this solution. Then, we process satellite data containing numerous images and complex parameters to improve the quality of the input data.

The article is structured in four distinct parts. The first, the introduction, describes the need for this research. The theoretical foundations of the research are presented in the second part. In the third part, we describe the research materials and methods, as well as the experimental setup. The research results are detailed in the fourth part, which offers commentary and discussion.

Related work: -

Land cover classification from satellite imagery has been the subject of much research, and various methods have been developed to address the challenges posed by the diversity of land cover classes. Traditional approaches and advanced artificial intelligence techniques have been used. Studies by Abdorreza and al. [10] and Thisanke and al. [11] offer a detailed analysis of Vision Transformers (ViTs) as a high-performance alternative to traditional CNNs for semantic segmentation, presenting their applications, strengths, limitations, and performance on different reference datasets. Deep learning, and more specifically CNNs, is the most efficient method for classifying remote sensing images according to Amer and al. [12]. These models achieve high accuracy with high-resolution satellite data by leveraging multi-scale information and hybrid techniques. The research by Boonpook and al. in [13] introduce LoopNet, a novel convolutional neural network architecture designed for the semantic segmentation of land cover types from Landsat 8 imagery. Trained on a six-class dataset (background, agriculture, forest, miscellaneous, urban, and water), LoopNet significantly outperforms reference models for capturing complex land cover features, despite classification challenges for urban and miscellaneous areas. This research demonstrates the effectiveness of deep learning techniques for improving remote sensing accuracy.
Automated land use classification has been a popular research topic for several decades. Traditional approaches have relied on supervised classification methods, such as support vector machines (SVMs) and random forests, which are applied to spectral indices (e.g. NDVI and NDWI) extracted from multispectral images. Zhang and al. demonstrated the effectiveness of SVMs for land use classification using Landsat images, achieving 85% accuracy across six classes [14]. However, these approaches are limited by the need for manual feature engineering and are sensitive to atmospheric conditions. In the context of remote sensing, [15] Nogueira et al. demonstrated that a ResNet50 model pre-trained on ImageNet could achieve an accuracy of 95% on the UC Merced Land Use dataset, surpassing traditional methods. Simonyan and Zisserman proposed VGG, a simple yet efficient architecture incorporating 3×3 convolutions. Castelluccio et al. applied in [16] VGG16 to satellite scene classification, demonstrated the importance of transfer learning for small remote sensing datasets. Andrew et al. [17] introduced MobileNet, which uses depth wise separable convolutions to reduce the number of parameters. Kussul and al. [18] demonstrated MobileNet's effectiveness for crop mapping in Ukraine, achieving 94% accuracy and reducing processing time by 75%. Helber et al. in [19] evaluated the performance of EfficientNet on the EuroSat dataset and demonstrated that it outperforms traditional architectures.

Materials and Methods: -

The study was conducted using supervised computing and learning environments, combined with deep learning methods. The equipment used for the calculations consisted of a GPU with 64 GB of memory and an NVIDIA GeForce GTX 1070 Ti graphics card, capable of handling the training of neural networks. Tools such as Anaconda, Python, Jupyter, and Notebook were used for data exploration, model development, and experimentation:

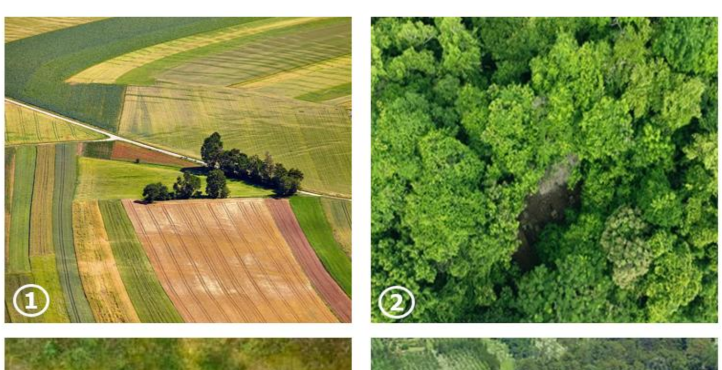

Figure 1: - shows sample images from the dataset: (1) Annual Crop, (2) Forest, (3) Herbaceous Vegetation, (4) Permanent Crop.

We also employed deep learning algorithms such as TensorFlow and Keras. All these experiments were performed on Windows 11, an operating system that supports all these software applications.
The EuroSat dataset is designed for land cover classification and is comprised of images taken by the Sentinel-2 satellite as part of the European Space Agency's (ESA) Copernicus program [19] It contains around 27,000 images, each measuring 64 × 64 pixels, which cover ten different land surface types. In Figure 1. some example images are presented.

During the data preparation phase, which involved segmenting the initial images into smaller areas, the dataset was divided into subsets for training and testing. We randomly selected 20% of the images for the test set and subsampled the remaining 80% because the initial dataset was unbalanced. To correct this imbalance, we subsampled the training set, selecting images with a more balanced distribution of classes, as appropriate subsampling can optimize model performance. Specifically, only image-mask pairs were included in the training dataset if at least four of the five categories contained at least 1,000 pixels. Before starting training, we performed several preprocessing steps, including normalization. The transformation methods applied included random rotation and image resizing with the following ratios: 0.5, 0.75, 1.0, 1.25, 1.5 and 1.75.

Subsequently, we used four classes to evaluate the performance of five image classification models: ResNet50, VGG16, GoogleNet and EfficientNet. These models were trained with the same baseline hyperparameters and evaluated using accuracy and loss metrics on both the training and validation sets. A summary of the crop images is shown in Table 1.

Table 1: -the crop images

Categories	**Number of Images**
AnnualCrop	3000
Forest	3000
PermanentCrop	500
HerbaceousVegetation	3000
Total	9500

The deep learning architecture is shown in Figure 2 below.

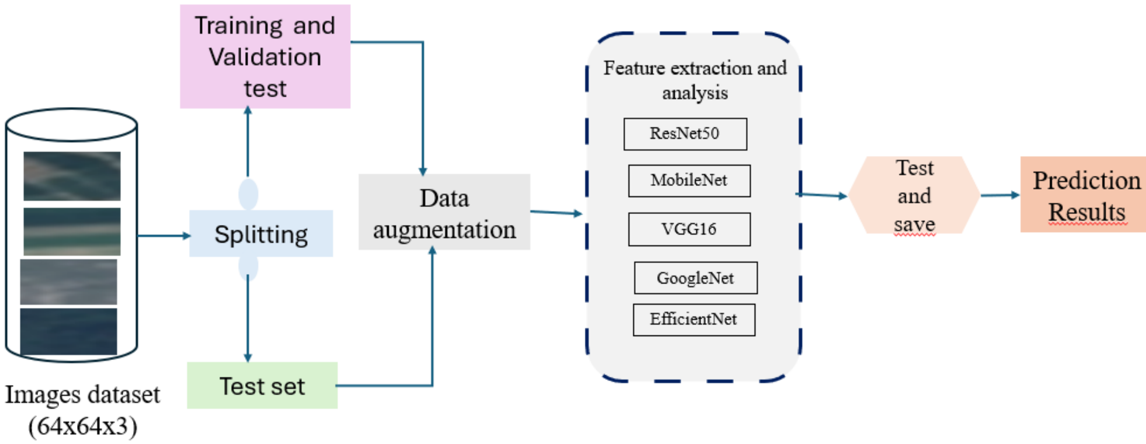

Figure 2: - Architecture of the 5 proposed models.

Figure 1 presents five proposed models. The first step of the architecture involves acquiring image data with dimensions of 64 × 64 × 3, where 64 × 64 represents the spatial resolution and 3 represents the RGB color channels. This data is then divided into sets: Training Data (1) and Validation Data (2). (1) The network learns to recognize image features on this set. Set (2) is reserved for evaluating the model's performance during training, enabling us to detect overfitting and adjust hyperparameters. After this step, we move on to data augmentation, which involves applying various transformations to the original images, such as brightness adjustments. This increases the size of the data set and improves the model. The models were trained in over 20 epochs with a batch size of 64 for both the training and validation datasets. Once training is complete, the model undergoes a testing and saving phase, during which is performance is evaluated on the dataset. The validated model is then saved and used for prediction.

Results: -

In recent years, the evolution of image recognition in remote sensing technologies for basic tasks, in addition to more complex missions such as denoising or categorization, has led to the development of a diverse range of classification methods concerning land cover and land use, see [20].
This section presents the outcomes of data preprocessing, training, and model evaluation. It also includes visualizations of the results related to the research topic.

Model performance

To evaluate the proposed system, we used the following metrics: accuracy, precision, recall and F1-score. The mathematical definitions of these evaluation criteria are presented below:

- $\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$
- $\text{Precision} = \frac{TP}{TP + FP}$
- $\text{Recall} = \frac{TP}{TP + FN}$
- $\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

Where: The evaluation metrics employed in this study include the pixel accuracy, precision, recall and F1-score. Detailed explanations of each metric are provided below, based on the terms false positive (FP), false negative (FN), true positive (TP), and true negative (TN).

Furthermore, their performance will be evaluated using the following indicators: precision, recall and F1-score. Training the system over a total of 20 epochs resulted in improved accuracy. Table 2 shows the comparisons.

Table 2: - Results of the different models.

Model	Accuracy Train (%)	Accuracy Val (%)	Convergence
ResNet50	66.7	66.7	Fast but limited
MobileNet	97.8	99.0	Excellent
VGG 16	85.8	87.2	Progressive
GoogleNet	97.8	99.0	Excellent
EfficientNet	66.7	66.7	Limited

Figures 3, 4, 5, 6, and 7 allow us to visualize the learning curves of the models.

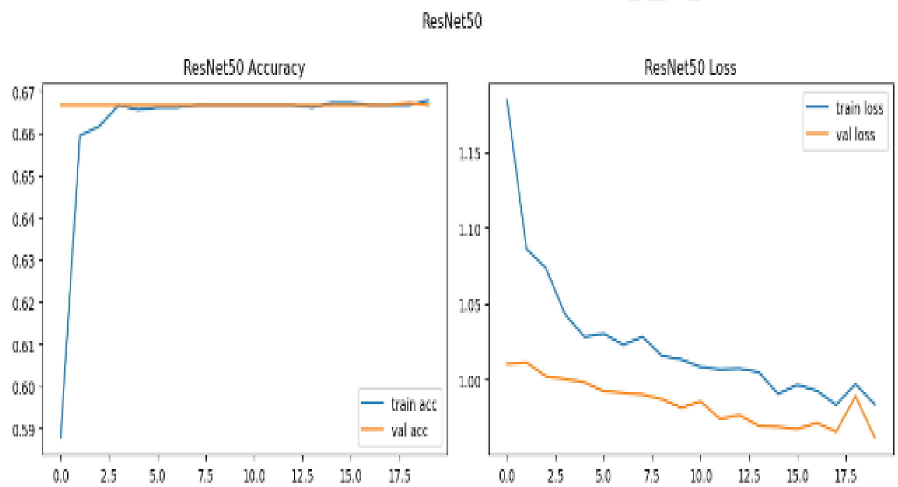

Figure 3: -(A) ResNet50

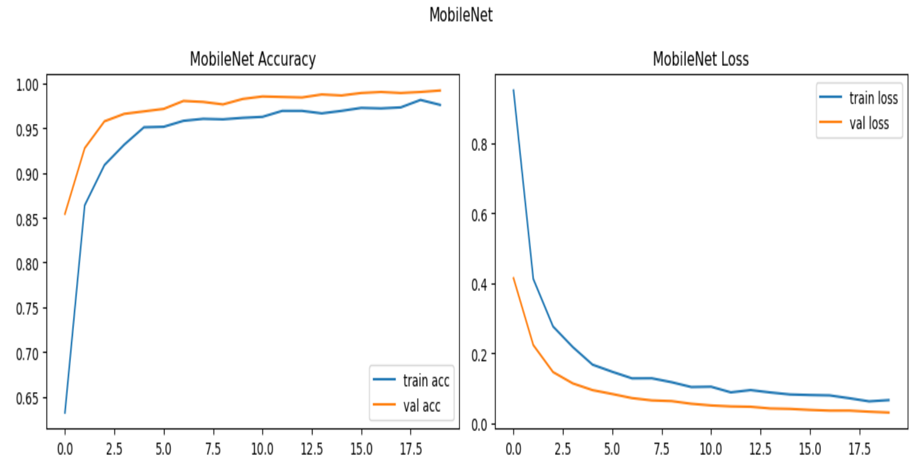

Figure 4: -(B) MobileNet

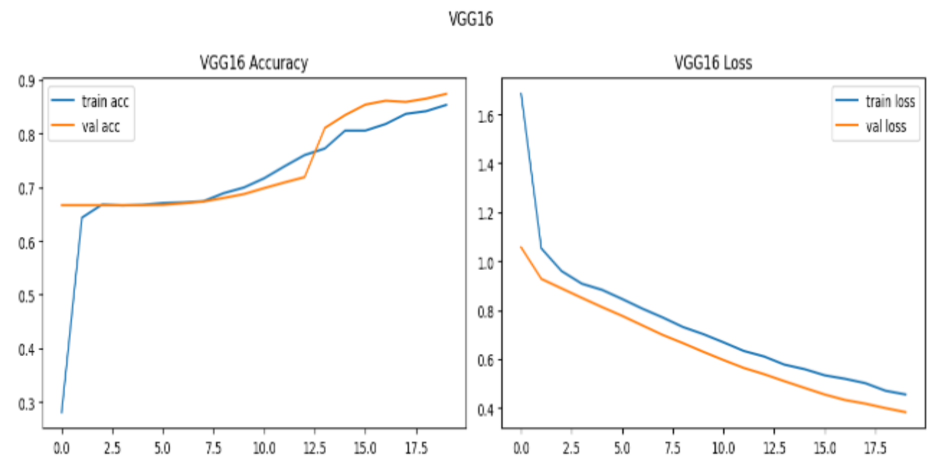

Figure 5:- (C) VGG 16

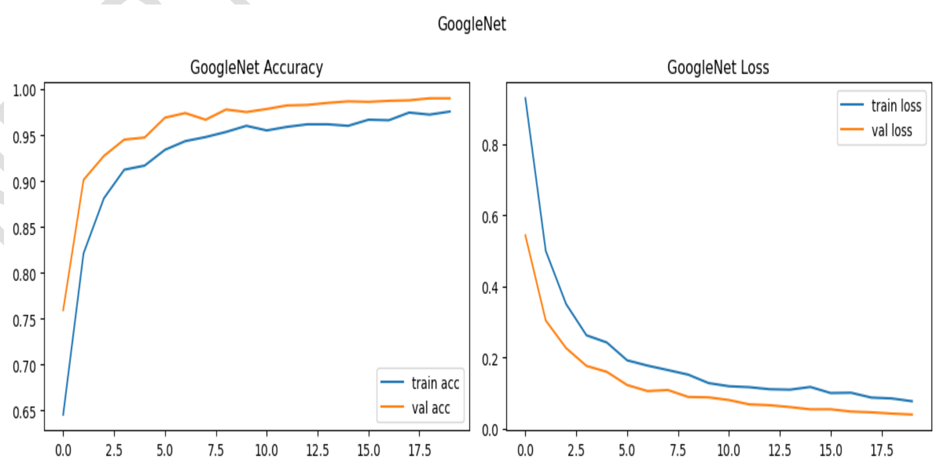

Figure 6: -(D) GoogleNet

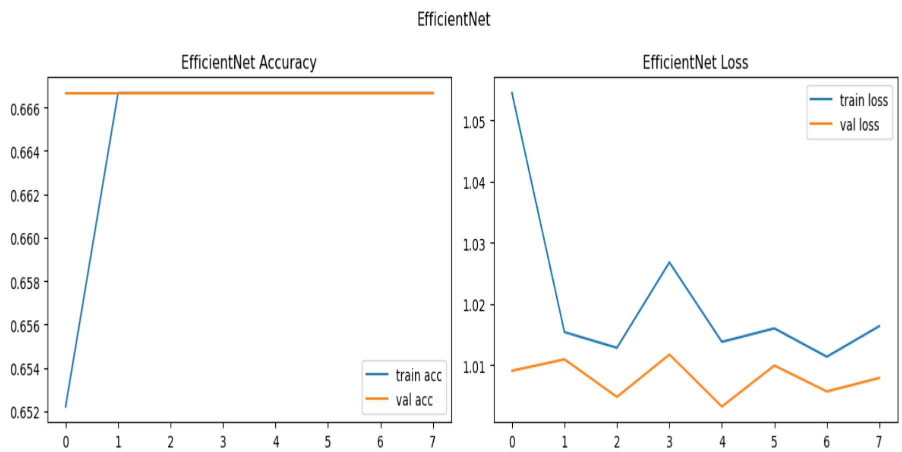

Figure 7: - (E) EfficientNet

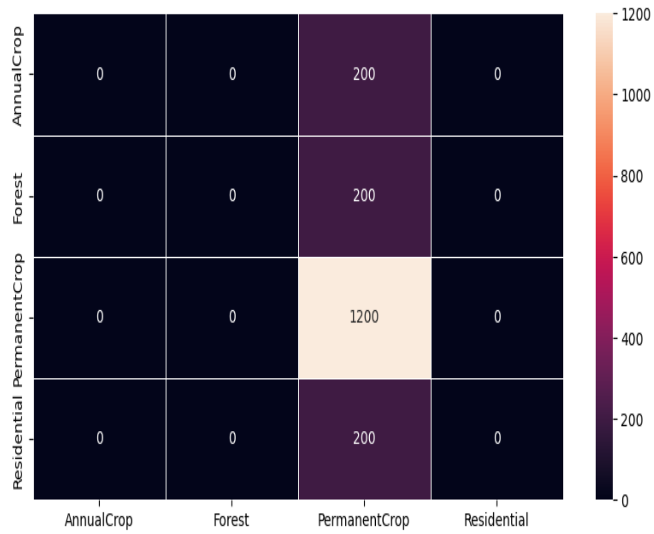

Figure 8: - (F) Confusion matrix of each model

This study is based on the EuroSAT dataset, which consists of multispectral images from Sentinel-2 satellite, encompassing 10 categories from a set of 27,000 labeled and georeferenced images. Sentinel-2A is one of two satellites in the Sentinel-2 constellation, which also includes Sentinel-2B. These two spacecrafts are identical and are tasked with monitoring landmasses. The Sentinel-2A and Sentinel-2B satellites were successfully launched in June 2015 and March 2017, respectively. These two sun-synchronous satellites image the Earth's surface globally using a multispectral imager (MSI) that covers 13 distinct spectral bands. Both satellites have a fuel reserve that ensures their operation for up to 12 years, thus providing the possibility of extending their service life. Two satellites in orbit provide almost total coverage of the Earth's surface every five days, meaning they record virtually every point in the covered area approximately every five days.

Detailed analysis by architecture: The ResNet50 model converges quickly but only achieves 66.7% accuracy on the validation set with high validation loss. This suggests that the model is struggling to generalize and is likely underfitted. In contrast, the MobileNet model performs excellently, converging to 99% accuracy on the validation set with a very low loss of around 0.05. Similarly, the VGG 16 model gradually improves, reaching 87.2% accuracy. Furthermore, training continues without any obvious signs of overfitting, suggesting good capacity. Like MobileNet, the GoogleNet model demonstrates exceptional performance, achieving 99% accuracy and fast convergence. This suggests excellent capacity. However, the EfficientNet model's performance is limited, similar to ResNet50's, with only 20 epochs of training and oscillations in the loss.

ion.

Figure 9: - (G) Model prediction

The confusion matrix in Figure 8 reveals perfect classification for the majority class (PermanentCrop) (1,200 samples) and confusion for the minority classes (AnnualCrop, Forest and Residential), each of which has 200 samples. This class imbalance appears to strongly influence the results, leading to accurate classification of the majority class but causing difficulties with the minority classes. Qualitative analysis of predictions: Figure 9 below shows the successful and unsuccessful cases. Green labels indicate correct predictions, and red labels indicate errors. Accuracy scores are displayed for each prediction.

Permanente crop prediction

The image shows a collection of labeled satellite image thumbnails, presumably used for training or evaluating a machine learning model. Each image is labeled with an "Actual" category and a "Predicted" category. The categories include "Permanent Crop," "Annual Crop," "Residential," and "Forest." An "Accuracy" indicator is also provided for each prediction, illustrating the model's confidence in its projections.

The model appears to have classification issues, which are evident in situations where the prediction does not match reality (e.g., "Actual: Permanent Crop Predicted: Residential"). This information is characteristic of a supervised learning procedure, where an algorithm is trained to classify images based on labeled data.

Discussion: -

In this article, our study focused on machine learning models capable of optimizing the automatic classification of agricultural methods and land use from satellite imagery. Following a review of our work, theMobileNet and GoogleNet models play a crucial role in classifying arable land. The results have significant implications for future studies and approaches aimed at overcoming the challenges in the agricultural sector, particularly regarding the management of real-time data in relation to its processing within modeling. When processing real-time data, it is crucial to consider various aspects such as data quality and volume, as well as the computing power that will be used.

We have observed that some classification errors cannot always be attributed to the CNN algorithm and network. In many cases, they result from inaccuracies in the reference annotations. Indeed, the human eye can sometimes struggle with ambiguous or visually confusing images. In the outlook, we intend to study the real-time land cover aspect of arable land using the LSTM algorithm or a Transformer.

Conclusion: -

This article demonstrates that the use of deep learning models for land use classification presents multiple opportunities for research and applications. With access to large-scale remote sensing datasets and the rapid evolution of deep learning methods, we can expect steady progress in this area. By developing more automated and efficient land use classification techniques, we will be able to fully utilize these data to address important environmental and societal issues.

Furthermore, our comparative study reveals surprising results that shed new light on the performance of convolutional neural network (CNN) architectures for land cover classification. MobileNet and GoogleNet emerge as the optimal solutions, achieving 99% accuracy while offering superior efficiency. However, transfer learning from ImageNet is effective for satellite data and balancing the quality of the remaining classes achieves optimal performance. Future work will focus on data balancing to improve the representativeness of all classes and on the design of specialized architectures for satellite imagery, as well as on image labeling to improve the score.

References: -

1. Butt, M.H.F., Li, J.P., Ahmad, M., Butt, M.A.F., 2024. Graph-infused hybrid vision transformer: Advancing GeoAI for enhanced land cover classification. Int. J. Appl. Earth Obs. Geoinf. 129, 103773. [Online] Available : http://dx.doi.org/10.1016/j.jag.2024.103773

2. Aljebreen, M., Mengash, H.A., Alamgeer, M., Alotaibi, S.S., Salama, A.S., Hamza, M.A., 2024. Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images. IEEE Access
3. Amirgan, B., Erener, A., 2024. Semantic segmentation of satellite images with different building types using deep learning methods. Remote. Sens. Appl.: Soc. Environ. 34, 101176.
4. Li, J., Cai, Y., Li, Q., Kou, M., Zhang, T., 2024a. A review of remote sensing image segmentation by deep learning methods. Int. J. Digit. Earth 17 (1), 2328827.
5. Fayaz, M., Nam, J., Dang, L.M., Song, H.-K., Moon, H., 2024. Land-cover classification using deep learning with high-resolution remote-sensing imagery. Appl. Sci. 14 (5), 1844
6. Azedou, A., Amine, A., Kisekka, I., Lahssini, S., Bouziani, Y., Moukrim, S., 2023. Enhancing land cover/land use (LCLU) classification through a comparative analysis of hyperparameters optimization approaches for deep neural network (DNN). Ecol. Inform. 78, 102333
7. Ramos, L., Sappa, A.D., 2024. Multispectral semantic segmentation for land cover classification: An overview. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.
8. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
9. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. [Online] Available : https://doi.org/10.48550/arXiv.1704.04861
10. Abdorreza A. G. ,VahidH.1 , José J. M. M. , João M. R. S. Tavares, (2025), Land Cover Classification Model Using Multispectral Satellite Images Based on a Deep Learning SynergisticSemantic Segmentation Network, MDPI, *Sensors* 2025, *25*(7), 1988;[Online] Available : https://doi.org/10.3390/s25071988
11. Thisanke, H., Deshan, C., Chamith, K., Seneviratne, S., Vidanaarachchi, R., Herath, D., 2023. Semantic segmentation using vision transformers: A survey. Eng. Appl. Artif. Intell. 126, 106669.
12. Amer, I.M., Rahouma, K., Shahin, A.A.A., 2024. Remote sensing image processing, DEEP learning classification algorithms survey and review. J. Adv. Eng. Trends 43 (1), 239–267
13. Boonpook, W., Tan, Y., Nardkulpat, A., Torsri, K., Torteeka, P., Kamsing, P., Sawangwit, U., Pena, J., Jainaen, M., 2023. Deep learning semantic segmentation for land use and land cover types using landsat 8 imagery. ISPRS Int. J. Geo-Inf. 12 (1), 14.
14. Zang, N., Cao, Y., Wang, Y., Huang, B., Zhang, L., Mathiopoulos, P.T., 2021. Land-use mapping for high-spatial resolution remote sensing image via deep learning: A review. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 5372–5391.
15. Nogueira, K., Penatti, O. A. B., & Santos, J. A. D. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61, 539–556.
16. Castelluccio, M., Poggi, G., Sansone, C., &Verdoliva, L. (2015). Land use classification in remote sensing images by convolutional neural networks. Acta Ecologica Sinica, 28(2), 627–635.
17. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. [Online] Available : https://doi.org/10.48550/arXiv.1704.04861
18. N. Kussul, G. Lemoine, J. Gallego, S. Skakun, M. Lavreniuk, A. Shelestov,Parcel-based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2016. – vol. 9, no. 6. – P. 2500 – 2508. DOI: 10.1109/JSTARS.2016.2560141
19. P. Helber, B. Bischke, A. Dengel, D. Borth. EuroSAT: A Novel, Dataset and Deep Learning Benchmark for Land Use and Land Cover, (2019), Classification. arXiv preprint arXiv: 1709.00029
20. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 77-778).