A Deep CNN model for Landuse Landcover Classification for 4 Band Visible and NIR Datasets

Pranavi Chachra
Amity University
Noida, India
pranavi.chachra@gmail.com

Aparna Tiwari
Banasthali Vidyapith
Tonk,India
aparnatiwari0907@gmail.com

Minakshi Kumar
Indian Institute of Remote Sensing (ISRO)
Dehradun,India
minakshi@iirs.gov.in

Keywords: Land Use Land Cover (LULC) Classification, Convolutional Neural Network (CNN), Multi-Spectral Remote Sensing, Deep Learning, Image Classification.

Abstract

Accurate classification of Land Use and Land Cover (LULC) from satellite imagery is vital for environmental monitoring, sustainable urban development, and resource management. With the increasing availability of multi-spectral data from Earth observation missions such as Sentinel-2, deep learning provides powerful solutions for automating LULC classification. In this study, we present a lightweight Convolutional Neural Network (CNN) architecture tailored for 4-band satellite imagery. Unlike conventional approaches that rely solely on RGB inputs, our model incorporates Red, Green, Blue, and Near-Infrared (NIR) bands to capture a broader range of surface and vegetation characteristics. The architecture combines stacked convolutional blocks with batch normalization, pooling layers, and dropout regularization, ensuring both strong accuracy and efficient computation. Training was further enhanced through data augmentation strategies such as rotation, flipping, and zooming. Using the EuroSAT dataset (27,000 images across 10 classes), the model achieved a test accuracy of 96% and a macro-averaged F1-score of 0.96, with excellent performance in challenging categories such as Residential, SeaLake, and Forest. The compact design of the model makes it highly suitable for deployment in time-sensitive or resource-limited scenarios, including monitoring of city growth, assessing agricultural productivity, and supporting rapid response to environmental hazards.

1. Introduction

The classification of land use and land cover (LULC) is a fundamental task in the field of remote sensing, with applications spanning agriculture, forestry, urban development, climate modeling, and disaster management. Traditional methods such as maximum likelihood classifiers and decision trees, while effective in certain scenarios, are often limited by their reliance on handcrafted features and poor generalization across heterogeneous landscapes.

The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized image analysis tasks by enabling automatic feature extraction and robust spatial pattern recognition. Remote sensing datasets such as EuroSAT, which originate from the Sentinel-2 mission, offer valuable spectral and spatial data that can be utilized to train CNNs for precise LULC classification. The Sentinel-2 satellite provides 13 spectral bands at different spatial resolutions, with the Red, Green, Blue, and Near-Infrared (NIR) bands proving particularly effective for detecting vegetation, water bodies, and urban features.

The inclusion of the NIR band significantly enhances the model's capability to distinguish between different vegetation types due to its responsiveness to chlorophyll levels and leaf structure.

In this work, we propose a lightweight CNN architecture tailored to process 4-band (RGB + NIR) images extracted from the EuroSAT dataset. The model is trained to classify images into ten distinct land use categories, including both natural (e.g., River, Forest) and anthropogenic (e.g., Residential, Industrial) classes. To improve generalization and reduce overfitting, we employ data augmentation techniques such as rotation, translation, flipping, and zooming during the training process. Furthermore, training stability is ensured through the use of batch normalization and dropout layers, while the learning rate is adaptively reduced based on validation loss stagnation. Our experiments demonstrate that the proposed CNN model achieves high accuracy and robustness, even when working with a relatively low-resolution image size (64×64 pixels). Through quantitative evaluation metrics such as precision, recall, F1-score, and confusion matrix analysis, we show that the model not only performs well in terms of overall accuracy but also maintains consistency across all classes. The success of this approach highlights the value of integrating multi-band satellite data with deep learning models for efficient and scalable LULC classification.

2. Backgound and Related work

Land use and land cover (LULC) classification using satellite imagery has been an active area of research for decades. Initial methods primarily utilized conventional machine learning techniques such as Support Vector Machines (SVM), Random Forests, and Decision Trees, which required manual feature extraction from multispectral images. These methods, while effective in some scenarios, were limited by their dependence on domain expertise for feature engineering and their inability to capture complex spatial patterns present in diverse landscapes.

With the rise of deep learning, Convolutional Neural Networks (CNNs) have become the preferred method for image classification tasks because of their ability to learn features hierarchically. CNNs have been successfully applied to remote sensing problems, outperforming classical approaches by automatically learning spatial-spectral features directly from the input data. Works such as Basu et al. (2015) and Marmanis et al. (2016) demonstrated the superior performance of deep networks in urban and rural land classification tasks.

The evolution of LULC classification parallels the broader progress in remote sensing technology — from manual interpretation of aerial photographs to fully automated classification using high-resolution multispectral satellite imagery. These advancements have enabled finer-scale mapping of ecological and urban landscapes with consistent spatial and temporal coverage. Furthermore, open-access satellite programs like Copernicus Sentinel-2 and Landsat have democratized access to Earth observation data, empowering research groups worldwide to develop reproducible, data-driven environmental models without proprietary barriers.

Numerous datasets have been created to assess deep learning techniques in remote sensing. One notable benchmark is the EuroSAT dataset, introduced by Helber et al. (2019), which consists of Sentinel-2 imagery featuring 10 LULC classes spread across 13 spectral bands. EuroSAT has facilitated the training and assessment of deep learning models on uniform satellite data. Most previous studies, however, have primarily focused on using either RGB imagery or all 13 bands without isolating specific spectral subsets for targeted tasks.

The importance of Near-Infrared (NIR) data in vegetation monitoring is well established in the literature. The NIR band provides enhanced separability between vegetation and non-vegetation classes due to its sensitivity to chlorophyll and plant structure. For example, vegetation typically reflects strongly in the NIR range, whereas water bodies absorb it, allowing clearer distinction in mixed land types. By combining NIR with the standard visible RGB bands, researchers can strike a balance between spectral diversity and computational tractability.

In addition, the task of multi-class LULC classification presents challenges such as overlapping spectral signatures, seasonal changes, and the presence of mixed pixels at land-use boundaries. These challenges are particularly pronounced in agricultural or semi-natural landscapes where boundaries between land cover types are not clearly delineated. Therefore, it becomes essential to use not only robust model architectures but also well-curated input bands to reduce ambiguity.

More recently, transformer-based models and spectral attention mechanisms have been explored for satellite image classification. While promising, these methods are often computationally intensive and require large training datasets, making them less practical for lightweight or edge-deployed applications.

In light of these developments, our study contributes by isolating and utilizing the four most informative bands (RGB + NIR) from the EuroSAT dataset and designing a CNN architecture optimized for this configuration. Unlike generic pre-trained models, our approach is tailored for spectral satellite inputs and benefits from targeted data augmentation, making it more suitable for large-scale environmental monitoring tasks. The proposed method also prioritizes scalability and deployment efficiency, aiming to support real-time or resource-constrained geospatial applications.

3. Methodology

3.1 Dataset Description

The dataset used in this study is derived from the EuroSAT collection, which consists of Sentinel-2 satellite imagery categorized into 10 land use and land cover (LULC) classes. Each image has a spatial resolution of 64×64 pixels and is captured using the Sentinel-2 multispectral instrument. Originally, each image includes 13 spectral bands, covering visible, near-infrared, and shortwave infrared wavelengths.

For this study, we selected four key bands for their proven effectiveness in distinguishing between vegetative and non-vegetative surfaces:
(1) Band 2 (Blue)
(2) Band 3 (Green)
(3) Band 4 (Red)
(4) Band 8 (Near Infrared, NIR)

These four bands were extracted and stacked to form 4-channel images. This selection provides a balanced spectral composition that enhances vegetation detection, surface reflectance analysis, and land use separation. In particular, the inclusion of the NIR band improves the model's ability to distinguish vegetation types based on chlorophyll reflectance, aiding in the classification of spectrally similar classes such as Forest, Pasture, and HerbaceousVegetation.

The images are stored in GeoTIFF format, preserving geospatial metadata. All images were normalized during preprocessing using percentile-based contrast stretching to ensure consistency across samples and improve feature scaling for model input.

The dataset includes the following 10 LULC classes:
['SeaLake', 'Forest', 'HerbaceousVegetation', 'Residential', 'River', 'PermanentCrop', 'AnnualCrop', 'Pasture', 'Highway', 'Industrial']

Initially, a subset of 50 images per class was used to prototype the model architecture and data processing pipeline. Once validated, the dataset was expanded to include all available 4-band images, resulting in a final dataset of 27,000 samples. These were split into 80% for training (21,600 samples) and 20% for testing (5,400 samples). From the training data, 20% was further allocated for validation. Stratified sampling was used during all splits to maintain balanced class representation across subsets.

This curated and normalized 4-band version of EuroSAT forms the basis for training and evaluating the proposed CNN model in a controlled and balanced setting.

3.2 CNN Architecture

We propose a custom convolutional neural network (CNN) architecture tailored to process 8-bit 4-channel multispectral satellite imagery. The input shape is fixed at 64×64×4 to accommodate the RGB + Near-Infrared bands. The network comprises four convolutional blocks, each designed to progressively extract higher-level spatial and spectral features from the input image.

Each block includes:
(1) A 2D convolutional layer with a (3×3) kernel and ReLU activation
(2) Batch normalization to stabilize and accelerate training

https://doi.org/10.5194/isprs-annals-X-5-W2-2025-65-2025 | © Author(s) 2025. CC BY 4.0 License.

(3) A 2×2 max pooling layer to reduce spatial resolution
(4) Dropout regularization to prevent overfitting

The convolutional layers use an increasing number of filters (32, 64, 128, and 256) to capture hierarchical patterns, from low-level edges and textures to high-level semantic features. After the final convolutional layer, a Global Average Pooling (GAP) layer is applied to compress the spatial dimensions, thus reducing the number of parameters and enabling the network to focus on dominant features. This is followed by a fully connected dense layer with 128 units and ReLU activation, a dropout layer (rate 0.5), and a final softmax output layer for classifying into one of the 10 land cover categories.

Key architectural choices include:

(1) Dropout rates between 0.25 and 0.5 for strong regularization
(2) Batch Normalization after each convolution and dense layer to improve generalization and training convergence
(3) Adam optimizer with a learning rate of 0.001 for stable, adaptive updates
(4) Total trainable parameters: approximately 424,812, making the model lightweight and well-suited for efficient training on standard hardware and potential deployment in edge environments

The architecture balances complexity and efficiency, enabling it to capture intricate spatial–spectral relationships without overfitting, even with moderately sized datasets. Moreover, the use of GAP instead of flattening minimizes parameter count while preserving representational power, supporting the model's generalization on unseen test data.

3.3 Data Augmentation

To improve generalization and model robustness, we employed online data augmentation using TensorFlow's ImageDataGenerator. Augmentations were chosen to mimic common variations in satellite imagery due to seasonal changes, viewing angle, and atmospheric conditions. API during training. Augmentation strategies include:
(1) Random rotation (±15°)
(2) Translation along width and height (10%)
(3) Horizontal flip
(4) Zooming (±10%)
(5) Filling new pixels using the nearest strategy

These transformations increase the diversity of the training data and help the model learn invariant representations.

3.4 Training Strategy

The model is trained using sparse categorical cross-entropy loss and monitored using validation accuracy. We employ:
(1) EarlyStopping to terminate training if the validation accuracy does not improve for 15 consecutive epochs
(2) ReduceLROnPlateau to adaptively reduce the learning rate on plateau
(3) ModelCheckpoint to save the best model weights

The model is trained for up to 40 epochs using a batch size of 32. Validation data is split from the training set (20%), resulting in a 64–16–20 train–validation–test split. Stratified sampling is used to preserve class balance across all subsets.
All learning is optimized using the Adam optimizer with an initial learning rate of 0.001.

This setup ensures stable convergence, prevents overfitting, and enables the model to generalize well on unseen data.

4. Results and discussion

4.1 Evaluation Metrics

The performance of the proposed CNN model was evaluated using standard classification metrics including:
(1) Accuracy: The overall percentage of correctly predicted labels.
(2) Confusion Matrix: Provides a detailed breakdown of true vs. predicted labels across all 10 classes, highlighting inter-class confusion.
(3) Class-wise Precision, Recall, and F1-score Useful for understanding performance per category, especially in multi-class classification tasks where some classes may be harder to distinguish than others.

These metrics were computed on a held-out test set of 5,400 images.

4.2 Quantitative Results

The final CNN model achieved a test accuracy of 96%, demonstrating robust classification performance on the 4-band EuroSAT dataset. The training process showed steady improvement, with training accuracy reaching ~97.95%, and validation accuracy stabilizing at around 96.02% after several epochs. These results indicate excellent learning capacity with minimal overfitting, supported by the application of dropout, batch normalization, and data augmentation techniques.
Evaluation metrics calculated on the 5,400-image test set show that the model maintains balanced performance across all 10 land cover classes.
The macro-averaged precision, recall, and F1-score were all 0.96, reflecting strong generalization. Notably, classes such as Residential, SeaLake, and Forest achieved F1-scores ≥ 0.97, while even challenging categories like Pasture and Herbaceous Vegetation recorded F1-scores above 0.92

Metric	Value
Training Accuracy (final epoch)	~97.95%
Validation Accuracy (best)	~96.02%
Test Accuracy	96%
Test Set Size	5400 images

The use of data augmentation played a crucial role in achieving this level of generalization. It enabled the model to better handle intra-class variability due to seasonal, geographical, and atmospheric differences.

4.3 Confusion Matrix Analysis

The normalized confusion matrix (Figure 1) illustrates the detailed classification performance across all 10 land cover classes. Notable insights include:
(1) Excellent performance on SeaLake, Forest, and Residential classes, with F1-scores above 0.97, indicating that these classes have strong and distinctive spectral patterns, particularly in the NIR band.
(2) High precision and recall for Industrial, Highway, and AnnualCrop, each achieving F1-scores above 0.92, suggesting the model successfully captures built-up area characteristics and cropland textures.
(3) While confusion remained between classes like HerbaceousVegetation and PermanentCrop, and Pasture and

HerbaceousVegetation, the model's improvements in F1-scores—especially for Pasture (now 0.93) and PermanentCrop (0.96)—demonstrate stronger separation.
Despite these minor overlaps, most predictions were accurate, with macro-averaged precision, recall, and F1-score all exceeding 0.9, validating the model's ability to distinguish complex land cover patterns.

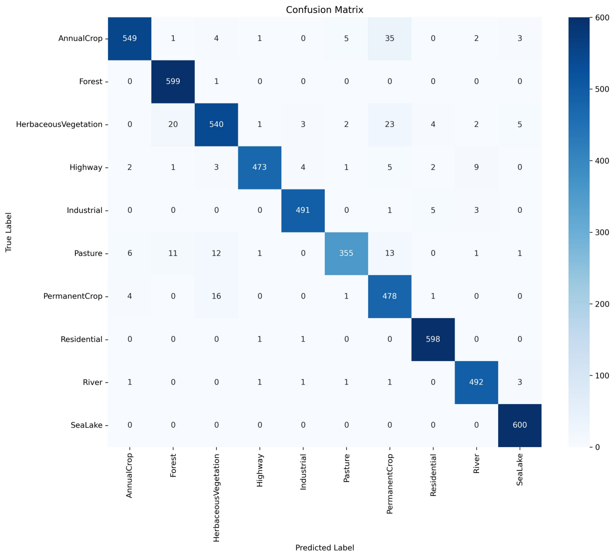

Figure 1. Confusion matrix of test set predictions.

4.4 Prediction Analysis

Figure 2 presents sample predictions across all classes, annotated with true labels, predicted labels, and associated confidence scores. The model correctly classifies a majority of the images, particularly those with distinct spectral or spatial patterns. High-confidence predictions are observed in classes such as SeaLake, Forest, Residential, and Highway, which exhibit strong visual contrast and well-defined structural or environmental features.

Misclassifications tend to occur in spectrally similar or visually ambiguous categories, such as PermanentCrop versus HerbaceousVegetation, or AnnualCrop versus PermanentCrop. These errors typically arise in edge regions, mixed land cover zones, or areas with low intra-class variance. Despite these overlaps, the model effectively captures meaningful spectral–spatial relationships and demonstrates consistent performance across diverse land cover types.

Figure 2. Sample predictions across land cover classes.

4.5 Comparison with Baselines

To evaluate the effectiveness of our proposed 4-band CNN model, we compare its performance with baseline approaches commonly used for LULC classification on the EuroSAT dataset. Traditional baselines include CNN architectures trained solely on RGB bands (Bands 2, 3, and 4), as well as more generic pre-trained models such as ResNet50 and VGG16 fine-tuned for satellite image classification.

RGB-only CNN models typically achieve classification accuracies in the range of 87% to 91% on EuroSAT, with macro F1-scores hovering between 0.85 and 0.89. These models, although effective at capturing basic spatial textures and color-based differences, are inherently limited in distinguishing vegetation types or water-related features due to the absence of Near-Infrared (NIR) information.

By contrast, our model achieves 96% accuracy and a macro F1-score of 0.96, marking a substantial improvement of 5–9 percentage points over RGB-only models. This performance boost is, largely due to:
(1) A custom CNN architecture tuned for multi-spectral satellite imagery
(2) Incorporating the NIR band
(3) Strong data augmentation strategies
(4) Batch Normalization and Dropout for regularization

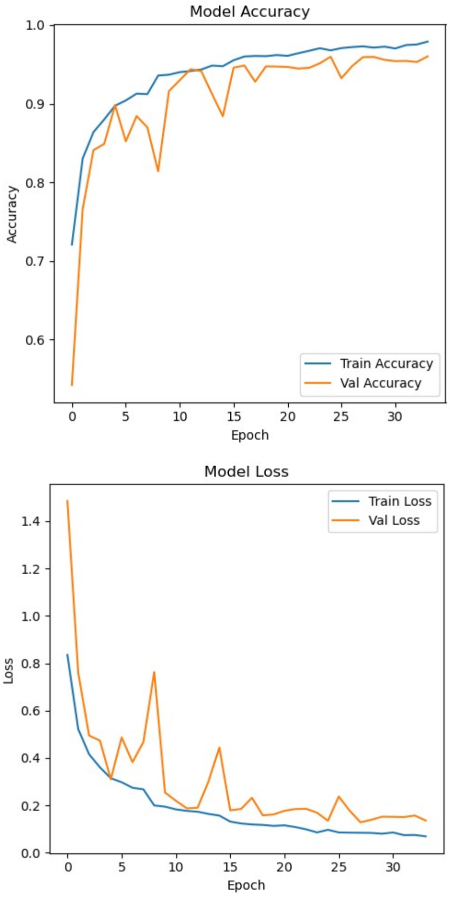

Figure 3. Training and validation accuracy graphs

5. Conclusion and Future Work

This study presents a robust and efficient CNN-based approach for land use and land cover (LULC) classification using a custom 4-band (RGB + NIR) version of the EuroSAT dataset. By leveraging the spectral richness of Sentinel-2 imagery and integrating it with a lightweight convolutional architecture, we achieved a high test accuracy of 96% and a macro-averaged F1-score of 0.96 across 10 diverse land cover classes.

The inclusion of the Near-Infrared (NIR) band proved essential for improving the separability of vegetation-related classes, while the use of targeted data augmentation and regularization techniques like dropout and batch normalization helped ensure strong generalization. The model performed especially well in spectrally complex categories such as Forest, SeaLake, and Residential, demonstrating its ability to capture both natural and human-made features from satellite imagery. Training and inference remained efficient due to the model's low parameter count and design tailored to 64×64 pixel multi-spectral inputs.

Looking ahead, there are several directions for extending this work. First, the model can be adapted to support the full 13-band Sentinel-2 imagery using band-specific attention mechanisms or spectral fusion modules to exploit additional spectral information such as shortwave infrared (SWIR) and red-edge bands. This would enhance class separability in cases where RGB+NIR alone may be insufficient.
Second, integrating temporal satellite imagery—i.e., multi-date sequences—could enable dynamic LULC monitoring and improve performance in applications such as crop type classification, land change detection, and seasonal vegetation analysis. Time-series-aware models could further refine predictions by capturing phenological patterns.
Third, we plan to explore transfer learning using large-scale Earth observation datasets such as SEN12MS, BigEarthNet, or So2Sat. Leveraging pre-trained multi-spectral representations could significantly reduce training time and enhance generalization across different geographies and climates. Such an approach would also make the model adaptable to other satellite constellations with minimal tuning.

Overall, this work demonstrates that a carefully optimized, band-selective CNN can achieve state-of-the-art performance on LULC classification tasks while maintaining computational efficiency. The findings underscore the importance of combining spectral domain knowledge with deep learning strategies to build scalable, reliable, and interpretable remote sensing models.

References

Adegun, A.A., Viriri, S. and Tapamo, J.R., 2023. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. Journal of Big Data, 10(93). https://doi.org/10.1186/s40537-023-00772-x

Ali, K.; Johnson, B.A. Land-Use and Land-Cover Classification in Semi-Arid Areas from Medium-Resolution Remote-Sensing Imagery: A Deep Learning Approach. *Sensors* 2022, *22*, 8750. https://doi.org/10.3390/s22228750.

Arrechea-Castillo, D.A.; Solano-Correa, Y.T.; Muñoz-Ordóñez, J.F.; Pencue-Fierro, E.L.; Figueroa-Casas, A. Multiclass Land Use and Land Cover Classification of Andean Sub-Basins in Colombia with Sentinel-2 and Deep Learning. *Remote Sens.* 2023, *15*, 2521. https://doi.org/10.3390/rs15102521

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M. and Nemani, R., 2015. DeepSat: A learning framework for satellite imagery. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM.

Carranza-García, M.; García-Gutiérrez, J.; Riquelme, J.C. A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks. *Remote Sens.* 2019, *11*, 274. https://doi.org/10.3390/rs11030274

Castelluccio, M., Poggi, G., Sansone, C. and Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:1508.00092.

Dhingra, S. and Kumar, D., 2019. A review of remotely sensed satellite image classification. International Journal of Electrical and Computer Engineering (IJECE), 9(2), pp.1200–1210. ISSN: 2088-8708.

Feng, Q.; Liu, J.; Gong, J. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sens.* **2015**, *7*, 1074-1094. https://doi.org/10.3390/rs70101074

Helber, P., Bischke, B., Dengel, A. and Borth, D., 2019. "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification" in *IEEE Journal of*

Selected Topics in Applied Earth Observations and Remote Sensing, 12(7), pp.2217–2226. doi.org/10.1109/JSTARS.2019.2918242

Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M. and Stilla, U., 2016. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135, pp.158–172.

Papoutsis, I., Bountos, N.-I., Zavras, A., Michail, D., Tryfonopoulos, C., 2021. Benchmarking and scaling of deep learning models for land cover image classification. *arXiv preprint*, arXiv:2111.09451. https://doi.org/10.48550/arXiv.2111.09451

Rwanga, S.S. and Ndambuki, J.M. (2017) Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. International Journal of Geosciences, 8, 611-622. https://doi.org/10.4236/ijg.2017.84033.

Smith, A., 2019. A review of remotely sensed satellite image classification. International Journal of Electrical and Computer Engineering (IJECE), 9(2), ISSN: 2088-8708.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv preprint*, arXiv:1905.11946. https://doi.org/10.48550/arXiv.1905.11946

Yassine, H., Tout, K. and Jaber, M., 2021. Improving LULC classification from satellite imagery using deep learning – EuroSAT dataset. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2021, pp.369–376.

Zhao, S.; Tu, K.; Ye, S.; Tang, H.; Hu, Y.; Xie, C. Land Use and Land Cover Classification Meets Deep Learning: A Review. *Sensors* 2023, *23*, 8966. https://doi.org/10.3390/s23218966