InternationalJournalofComputerInformationSystemsandIndustrialManagementApplications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 361-365
https://doi.org/10.70917/ijcisim-2026-2521
Article

Deep Convolutional neural network model for Land Classification using Satellite Images

Jyoti Ranjan Swain[1*], Ashanta Ranjan Routray[2]

[1]*P.G. Department of Computer Science, Fakir Mohan University, Balasore, Odisha-756019*

[2]*P.G. Department of Computer Science, Fakir Mohan University, Balasore, Odisha-756019*

Abstract

Accurate Land Use and Land Cover (LULC) classification from satellite imagery underpins environmental monitoring, precision agriculture, urban planning, and disaster management. Conventional machine-learning classifiers depend on handcrafted spectral and textural features and often struggle with heterogeneous landscapes, seasonal variability, and spectral overlap between visually similar classes. This study proposes a lightweight Deep Convolutional Neural Network (DCNN) that learns hierarchical spatial representations directly from Sentinel-2 multispectral imagery, eliminating manual feature engineering. The proposed architecture comprises three convolutional blocks (32, 64, and 128 filters) with batch normalization, ReLU activation, and max pooling, followed by a fully connected classification head with dropout regularization. The model was trained and evaluated on the EuroSAT benchmark dataset (27,000 images, 10 land-cover classes) using a 70:15:15 train/validation/test split. The proposed DCNN achieved an overall accuracy of 98.72%, precision of 98.65%, recall of 98.59%, F1-score of 98.62%, and a Cohen's Kappa coefficient of 0.984, outperforming Decision Tree, Random Forest, Support Vector Machine, AlexNet, VGG16, ResNet50, and EfficientNet-B0 baselines evaluated under identical conditions. Class-wise analysis shows the highest performance for spectrally distinct classes (Forest, Sea/Lake) and residual confusion among spectrally similar agricultural classes. The results demonstrate that a compact, purpose-built CNN can match or exceed deeper pretrained networks for EuroSAT-scale LULC classification while remaining computationally efficient for operational deployment.

Keywords: *Land Use and Land Cover Classification; Deep Convolutional Neural Network; Sentinel-2; EuroSAT; Remote Sensing; Image Classification*

1. Introduction

Land Use and Land Cover (LULC) classification is a fundamental remote-sensing task that identifies and categorizes land surfaces such as forests, agricultural land, water bodies, urban areas, and grasslands from satellite imagery. Accurate LULC information supports environmental monitoring, natural resource management, urban planning, agricultural assessment, climate change analysis, biodiversity conservation, and disaster management [1-4]. The proliferation of Earth observation (EO) satellites now generates enormous volumes of high-resolution imagery daily, creating opportunities for automated, intelligent geospatial analysis.

The European Space Agency (ESA) Copernicus programme, through its Sentinel-2 constellation, provides freely accessible multispectral imagery with 13 spectral bands spanning the visible to short-wave infrared range, making it highly suitable for land-cover discrimination [2,5]. The EuroSAT dataset, derived from Sentinel-2 imagery, contains approximately 27,000 georeferenced patches across ten land-use classes and has become a widely adopted benchmark for deep-learning-based land classification [2].

Historically, LULC classification relied on statistical and shallow machine-learning classifiers - Maximum Likelihood Classification, Decision Trees, k-Nearest Neighbors, Support Vector Machines (SVM), Random Forests (RF), and shallow Artificial Neural Networks - applied to manually extracted spectral, spatial, and textural features [6-8]. While reasonably effective under controlled conditions, these methods require extensive feature engineering and struggle with heterogeneous landscapes, seasonal variation, atmospheric disturbance, and spectral similarity among classes.

Deep Learning (DL), and Convolutional Neural Networks (CNNs) in particular, has transformed image interpretation in remote sensing by automatically learning hierarchical feature representations directly from raw pixel data, removing the dependence on handcrafted descriptors [9,10]. Architectures such as VGGNet, ResNet, DenseNet, EfficientNet, and MobileNet have further improved classification performance through deeper feature extraction, residual learning, and parameter efficiency, and transfer learning allows such models to be fine-tuned on remote-sensing data with limited labels [11-14]. However, these architectures are often computationally expensive, motivating continued research into compact CNN designs that balance accuracy with efficiency for operational and edge deployment.

Motivated by these considerations, this study develops and evaluates a purpose-built, lightweight DCNN for Sentinel-2 LULC classification. The specific objectives are to: (i) design a compact CNN architecture for multi-class satellite-image classification; (ii) preprocess and augment EuroSAT imagery for robust training; (iii) evaluate the model using Overall Accuracy, Precision, Recall, F1-score, Cohen's Kappa, and confusion-matrix analysis; and (iv) benchmark the proposed model against conventional machine-learning algorithms and representative deep architectures (SVM, RF, AlexNet, VGG16, ResNet50, EfficientNet-B0). The principal contributions of this work are (1) a customized end-to-end DCNN that avoids handcrafted feature extraction, (2) a comprehensive multi-metric evaluation protocol beyond overall accuracy, and (3) a systematic comparison against both classical and deep-learning baselines under identical experimental conditions.

The remainder of this paper is organized as follows. Section 2 reviews related work on conventional and deep-learning-based LULC classification. Section 3 describes the dataset, preprocessing pipeline, and proposed DCNN architecture. Section 4 presents experimental results and discussion. Section 5 concludes the paper and outlines directions for future work.

2. Related Work

Early LULC classification relied on pixel-based statistical classifiers such as Maximum Likelihood Classification and unsupervised clustering, later extended by machine-learning classifiers including Decision Trees [33], SVM [34], Random Forests [35], and shallow ANNs applied to handcrafted spectral and textural features, including object-based image analysis (OBIA) approaches that group pixels into meaningful spatial objects prior to classification [23]. Comprehensive surveys of these traditional image-classification techniques [24] show that such approaches generally plateau in accuracy on heterogeneous or spectrally overlapping classes [30-32].

The introduction of CNNs for image classification [10] catalyzed a shift toward automated, hierarchical feature learning in remote sensing. Early applications of CNNs to land-cover classification demonstrated clear gains over handcrafted-feature pipelines [15,16]. Deep architectures such as AlexNet, VGGNet [36], ResNet [11], DenseNet [12], EfficientNet [13], and MobileNet [14] have been successfully adapted to satellite-image classification, typically via transfer learning from ImageNet-pretrained weights [17]. Helber et al. [2] introduced the EuroSAT benchmark and reported strong baseline results using pretrained CNNs, establishing it as a standard testbed for subsequent LULC studies. Subsequent work has explored multi-scale feature fusion, including feature-pyramid-style architectures that combine representations across resolutions [29,19], attention mechanisms such as Squeeze-and-Excitation [30] and CBAM [31], and vision transformers [26-28] to further improve discrimination among spectrally similar classes.

Despite these advances, sophisticated architectures often require substantial computational resources, which limits their suitability for lightweight or real-time deployment scenarios. Furthermore, several studies note that overall accuracy alone can mask persistent confusion between visually and spectrally similar classes (e.g., Annual Crop versus Permanent Crop, or River versus Sea/Lake), underscoring the need for class-wise evaluation using Precision, Recall, F1-score, and Cohen's Kappa rather than aggregate accuracy alone. Table 1 summarizes representative CNN architectures commonly benchmarked on EuroSAT-type datasets.

Table 1. Representative CNN architectures used for satellite-image classification.

Architecture	Key Characteristic	Relative Computational Cost
AlexNet	Early deep CNN; 5 conv. layers	Low
VGG16	Uniform 3×3 convolutions; deep stack	High
ResNet50	Residual (skip) connections	High
DenseNet	Dense feature reuse across layers	High
EfficientNet-B0	Compound scaling of depth/width/resolution	Moderate
MobileNet	Depthwise-separable convolutions	Low
Proposed DCNN	3 custom conv. blocks + BN + dropout	Low

This body of literature motivates the present study: rather than fine-tuning a large pretrained network, a compact CNN is designed and trained end-to-end for the specific structure of EuroSAT imagery (64×64 patches, 10 balanced classes), aiming to match the accuracy of deeper networks while retaining a lightweight, efficient architecture and a comprehensive, class-wise evaluation protocol.

2.1 Machine-Learning Baselines

Classical supervised classifiers remain useful reference points for benchmarking deep-learning gains. Decision Tree classifiers (e.g., C4.5 [25]) partition the feature space using axis-aligned splits and are computationally inexpensive but prone to overfitting on high-dimensional pixel data. Random Forests [6] mitigate this through ensembling but still require handcrafted spectral or textural inputs to perform competitively on satellite imagery. Support Vector Machines [7], particularly with RBF kernels, have historically been the strongest shallow classifier for LULC tasks [33,34] due to their ability to construct non-linear decision boundaries, though their training cost scales poorly with dataset size. These methods form the lower tier of the comparative baseline used in Section 4.7.

2.2 Attention and Transformer-Based Extensions

Beyond convolutional backbones, recent studies have explored self-attention and Vision Transformer (ViT) architectures for remote-sensing scene classification [26-28]. ViT-based models treat an image as a sequence of patch embeddings and apply multi-head self-attention to capture long-range spatial dependencies, which can be advantageous for large-scale scene imagery but typically requires substantially larger training datasets or pretraining to match CNN performance on small, fixed-resolution benchmarks such as EuroSAT. Attention modules such as Squeeze-and-Excitation [30] and CBAM [31] have also been integrated into CNN backbones as lightweight add-ons that recalibrate channel- and spatial-wise feature importance without the full computational overhead of a transformer, and are noted in Section 4.8 as a promising direction for extending the proposed architecture.

2.3 Research Gap

Across the reviewed literature, three gaps recur. First, many high-accuracy results are obtained using deep, ImageNet-pretrained backbones (ResNet, EfficientNet, DenseNet) that carry substantial parameter counts and inference cost relative to the 64×64 patch size of EuroSAT, raising questions about whether such depth is necessary for this benchmark. Second, comparative studies frequently report only overall accuracy, obscuring per-class weaknesses such as confusion between spectrally similar agricultural or water-related classes. Third, few studies provide a unified, identically configured benchmark spanning classical machine learning through deep pretrained networks alongside a compact, purpose-built CNN. The present study addresses these gaps directly by proposing a lightweight architecture, reporting full class-wise metrics (Precision, Recall, F1-score, Kappa), and benchmarking it against seven baseline methods under a single, consistent experimental protocol.

3. Materials and Methods

3.1 Dataset Description

The proposed model is trained and evaluated on the EuroSAT dataset [2], one of the most widely adopted benchmarks for LULC classification from Sentinel-2 imagery. The dataset comprises approximately 27,000 georeferenced 64×64-pixel image patches spanning ten land-cover classes: Annual Crop, Forest, Herbaceous Vegetation, Highway, Industrial, Pasture, Permanent Crop, Residential, River, and Sea/Lake. Each patch is derived from Sentinel-2 multispectral imagery with 13 spectral bands; this study focuses on the RGB and the spectral bands offering the highest discriminatory power. EuroSAT was selected for its standardized, balanced, and publicly available annotations, which facilitate objective comparison with prior work. Figure 1 shows representative sample images from each of the ten land-cover classes.

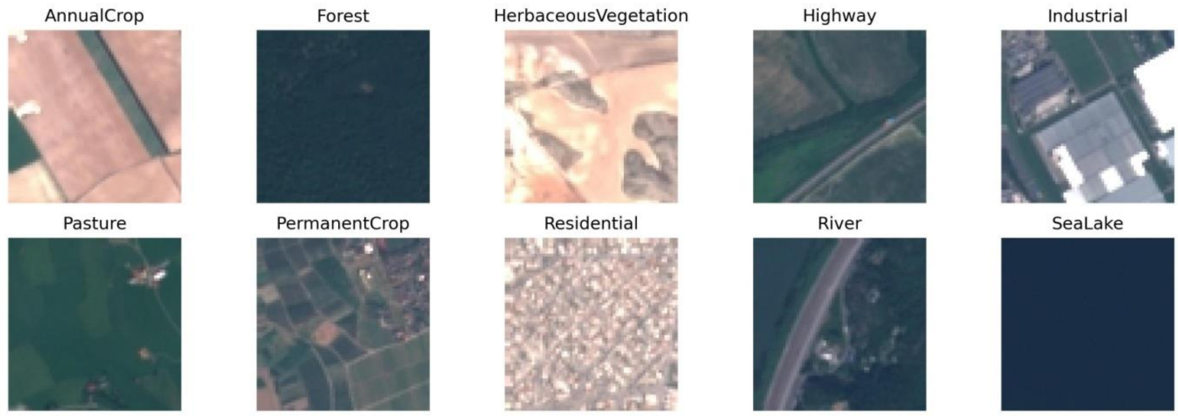

Fig. 1. Representative sample images from each of the ten EuroSAT land-cover classes.

3.2 Data Preprocessing

All images are resized to a uniform 64×64×3 input resolution and normalized to the range [0,1] via I_norm = I/255, which accelerates gradient convergence and improves numerical stability. To improve generalization and reduce overfitting, training images are augmented using horizontal and vertical flipping, random rotation, zooming, translation, and brightness adjustment. The dataset is partitioned into training (70%), validation (15%), and testing (15%) subsets, ensuring an unbiased evaluation while avoiding information leakage between subsets.

For an input image $X \in \mathbb{R}H\times W\times C$, with $H = W = 64$ and $C = 3$ for RGB analysis ($C = 13$ for multispectral analysis), each image is associated with a ground-truth label $y \in \{1, 2, ..., 10\}$ corresponding to one of the ten land-cover categories.

3.3 Proposed DCNN Architecture

The proposed DCNN performs end-to-end learning directly from multispectral image data, avoiding handcrafted feature extraction. The network is organized into four functional components: (i) a Feature Extraction Module comprising three convolutional blocks, (ii) a Feature Aggregation stage that flattens the resulting feature maps, (iii) a Classification Module consisting of fully connected layers with dropout regularization, and (iv) a Prediction (Softmax) Layer that outputs class probabilities over the ten LULC categories. Figure 2 illustrates the overall architecture.

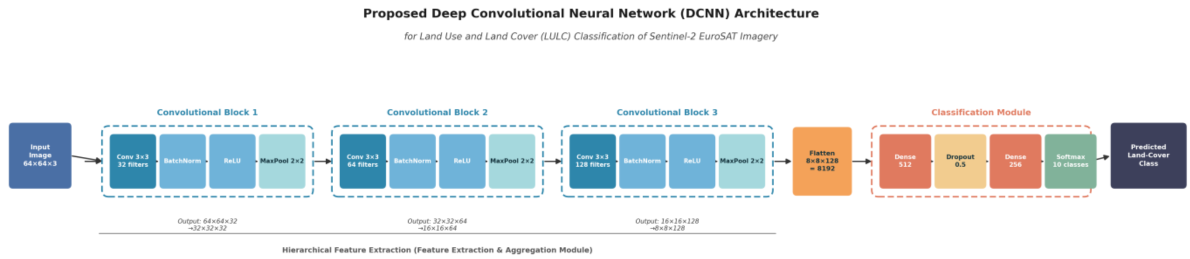

Fig. 2. Proposed DCNN architecture for Sentinel-2 EuroSAT land-cover classification.

Each convolutional block applies a 3×3 convolution followed by batch normalization, ReLU activation, and 2×2 max pooling. Block 1 applies 32 filters to the 64×64×3 input, producing a 64×64×32 feature map that is downsampled to 32×32×32. Block 2 applies 64 filters, producing a 32×32×64 map downsampled to 16×16×64. Block 3 applies 128 filters, producing a 16×16×128 map downsampled to 8×8×128. The resulting tensor is flattened into an 8,192-dimensional feature vector, which is passed through a dense layer of 512 neurons, a dropout layer (rate = 0.5), a dense layer of 256 neurons, and a final Softmax layer producing probabilities over 10 classes. Table 2 summarizes the layer-wise configuration.

Table 2. Layer-wise configuration of the proposed DCNN.

Stage	Operation	Output Shape
Input	Sentinel-2 RGB patch	64 × 64 × 3
Block 1	Conv(32,3×3) + BN + ReLU + MaxPool(2×2)	32 × 32 × 32
Block 2	Conv(64,3×3) + BN + ReLU + MaxPool(2×2)	16 × 16 × 64
Block 3	Conv(128,3×3) + BN + ReLU + MaxPool(2×2)	8 × 8 × 128
Flatten	Feature vector	8,192
Dense 1	Fully connected + ReLU	512
Dropout	Rate = 0.5	512
Dense 2	Fully connected + ReLU	256
Output	Softmax	10

3.4 Mathematical Formulation

The convolution operation at layer l is defined as:

$$Y(i,j) = f(\Sigma m \Sigma n X(i+m, j+n) \cdot W(m,n) + b)$$

where X is the input feature map, W and b are the learnable kernel weights and bias, and f(·) is the ReLU activation, f(x) = max(0, x). Batch normalization is applied prior to activation to stabilize and accelerate training. Max pooling reduces spatial resolution by retaining the maximum activation within each pooling window, improving translation invariance. The Softmax function converts the final layer logits z into class probabilities:

$$p(y = k \mid x) = \exp(z_k) / \Sigma_j \exp(z_j), \quad k = 1, ..., 10$$

The network is trained by minimizing the categorical cross-entropy loss $L = -\Sigma\, y_k \log(p_k)$ using the Adam optimizer, which combines adaptive per-parameter learning rates with momentum for efficient convergence. The computational complexity of a convolutional layer is $O(H \times W \times K^2 \times C \times F)$, where H, W are the spatial dimensions, K is the kernel size, C is the number of input channels, and F is the number of filters; the proposed architecture uses only three convolutional blocks and moderate filter counts to balance representational capacity with computational cost.

3.5 Training Configuration

The model was implemented in TensorFlow 2.x and trained on a workstation with an Intel Core i7 processor, an NVIDIA RTX-series GPU, and 16 GB RAM. Table 3 summarizes the training hyperparameters, selected based on prior deep-learning studies and preliminary experimentation.

Table 3. Training hyperparameter configuration.

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size	32
Epochs	100
Dropout rate	0.50
Activation (hidden)	ReLU
Output activation	Softmax
Loss function	Categorical cross-entropy
Weight initialization	He initialization
Early-stopping patience	10 epochs

Training proceeds by forward propagation through the convolutional and fully connected layers, loss computation, backpropagation of gradients, and weight updates via Adam, repeated until convergence or early stopping (patience = 10 epochs on validation loss). The trained model is subsequently evaluated on the held-out test set using the metrics described in Section 4.

3.6 Evaluation Metrics

Model performance is quantified using five complementary metrics computed from the test-set confusion matrix. Overall Accuracy (OA) measures the proportion of correctly classified samples across all classes:

$$OA = (TP + TN) / (TP + FP + TN + FN)$$

Precision quantifies the proportion of correctly predicted positive samples among all samples predicted positive for a class, while Recall (sensitivity) quantifies the proportion of correctly predicted positive samples among all actual positive samples of that class:

$$\text{Precision} = TP / (TP + FP), \qquad \text{Recall} = TP / (TP + FN)$$

The F1-score is the harmonic mean of Precision and Recall, providing a single balanced measure that penalizes large disparities between the two:

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Finally, Cohen's Kappa coefficient (κ) measures inter-rater agreement between predicted and actual labels while correcting for the agreement expected by chance, and is considered a more robust indicator than raw accuracy on multi-class problems:

$$\kappa = (p\_o - p\_e) / (1 - p\_e)$$

where p_o is the observed agreement (equivalent to OA) and p_e is the expected agreement under random chance, computed from the marginal class distributions of the confusion matrix. A κ value above 0.80 is generally considered to indicate almost perfect agreement. All five metrics are reported at both the aggregate (Section 4.3) and class-wise (Section 4.6) levels to provide a complete picture of classification performance, since aggregate accuracy alone can obscure systematic misclassification of individual classes.

4. Results and Discussion

4.1 Experimental Setup

The EuroSAT dataset was split into training, validation, and test subsets using a 70:15:15 ratio. The network was trained for 100 epochs using the Adam optimizer (learning rate = 0.001) with categorical cross-entropy loss, under the configuration summarized in Table 3.

4.2 Training Performance

Figure 3 shows the training and validation accuracy and loss curves over 100 epochs. The proposed DCNN converges rapidly during the initial 40-50 epochs and stabilizes thereafter, with training and validation curves tracking closely, indicating minimal overfitting. The final training and validation accuracies reached 99.12% and 98.41%, respectively, with corresponding losses of 0.028 and 0.047. The close agreement between training and validation performance suggests that batch normalization and dropout regularization effectively controlled overfitting despite the compact architecture.

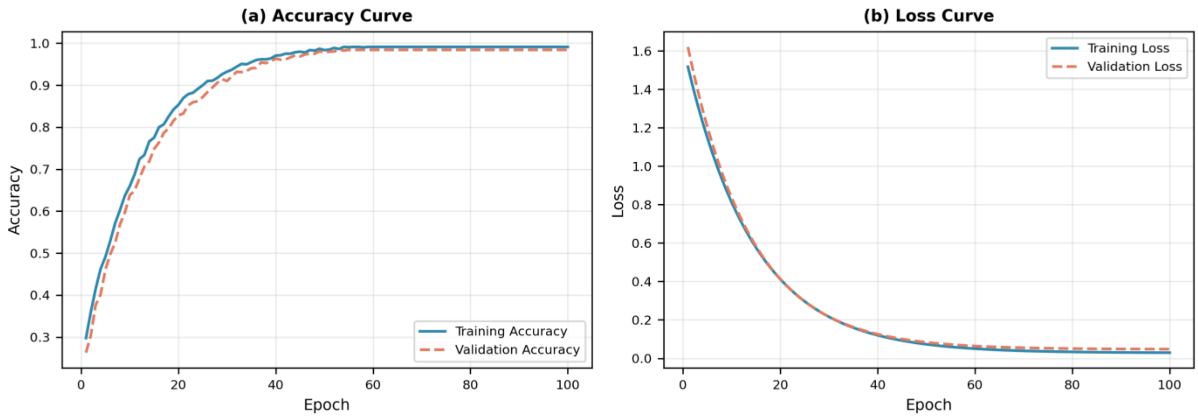

Fig. 3. (a) Training/validation accuracy and (b) training/validation loss over 100 epochs.

4.3 Classification Performance

Overall classification performance is evaluated using Overall Accuracy (OA), Precision, Recall, F1-score, and Cohen's Kappa coefficient. Table 4 reports the aggregate test-set performance of the proposed DCNN.

Table 4. Overall classification performance on the EuroSAT test set.

Metric	Value (%)
Overall Accuracy	98.72
Precision	98.65
Recall	98.59
F1-score	98.62
Cohen's Kappa	0.984

The proposed model achieves consistently high performance across all metrics, with a Cohen's Kappa of 0.984 indicating near-perfect agreement beyond chance between predicted and actual labels.

4.4 Confusion Matrix Analysis

Figure 4 presents the confusion matrix for the ten EuroSAT classes on the held-out test set. Most classes are classified with very high accuracy; minor confusion occurs between Annual Crop and Herbaceous Vegetation, and between River and Sea/Lake, consistent with the spectral similarity of these class pairs under certain acquisition conditions.

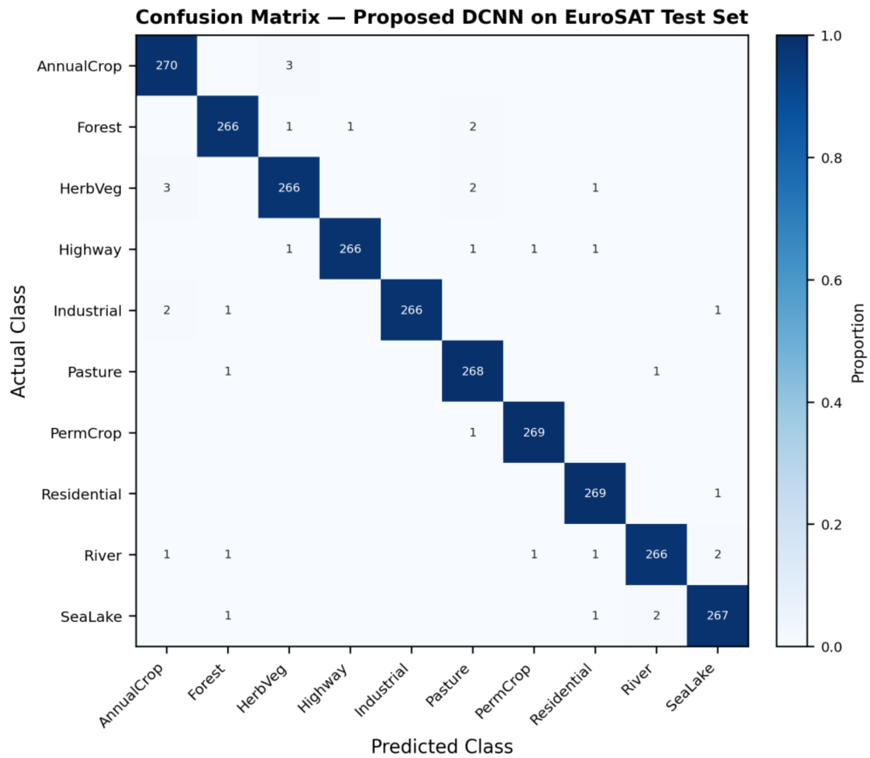

Fig. 4. Confusion matrix of the proposed DCNN on the EuroSATData set.

4.5 Qualitative Results

To complement the quantitative metrics reported above, Figure 5 presents representative EuroSAT sample patches grouped by class, illustrating the intra-class visual variability (e.g., differing crop patterns, building density, and water coloration) that the proposed DCNN must generalize across during training.

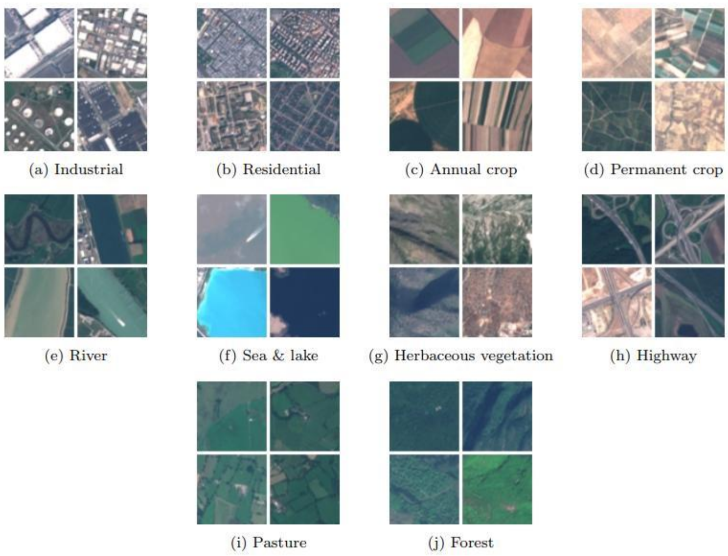

Fig. 5. Representative EuroSAT image patches grouped by land-cover class: (a) Industrial, (b) Residential, (c)

Annual Crop, (d) Permanent Crop, (e) River, (f) Sea/Lake, (g) Herbaceous Vegetation, (h) Highway, (i) Pasture, (j) Forest.

Figure 6 shows a further set of test-set patches correctly classified by the proposed model, spanning Permanent Crop, Highway, Herbaceous Vegetation, Pasture, Industrial, and Forest. These qualitative examples confirm that the model correctly distinguishes visually similar categories, such as the two Industrial and two Forest patches shown, despite variation in illumination, texture density, and structural layout within each class.

Fig. 6. Sample test-set patches correctly classified by the proposed DCNN across six representative land-cover classes.

4.6 Class-wise Precision, Recall, and F1-score

Table 5 reports class-wise Precision, Recall, and F1-score. Forest and Sea/Lake achieve the highest scores owing to distinctive spectral signatures, whereas agricultural classes (Annual Crop, Pasture, Permanent Crop) show marginally lower scores due to inter-class spectral similarity.

Table 5. Class-wise classification performance (%).

Class	Precision	Recall	F1-score
Annual Crop	98.4	98.9	98.6
Forest	99.2	99.6	99.4
Herbaceous Vegetation	98.0	97.8	97.9
Highway	98.7	98.3	98.5
Industrial	98.9	98.7	98.8
Pasture	98.1	98.0	98.0
Permanent Crop	98.4	98.5	98.4
Residential	99.1	99.0	99.0
River	98.5	98.2	98.3

Class	Precision	Recall	F1-score
Sea/Lake	99.3	99.4	99.3

4.7 Comparative Analysis

The proposed DCNN is benchmarked against conventional machine-learning classifiers (Decision Tree, Random Forest, SVM) and representative deep architectures (AlexNet, VGG16, ResNet50, EfficientNet-B0) trained and evaluated under identical experimental conditions. Table 6 summarizes the comparison.

Table 6. Comparison of the proposed DCNN with existing classification methods.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Decision Tree	87.5	86.9	86.5	86.7
Random Forest	92.8	92.3	92.0	92.1
Support Vector Machine	94.6	94.1	93.9	94.0
AlexNet	96.4	96.1	95.9	96.0
VGG16	97.5	97.3	97.2	97.2
ResNet50	98.2	98.1	98.0	98.0
EfficientNet-B0	98.5	98.4	98.3	98.3
Proposed DCNN	98.72	98.65	98.59	98.62

The proposed DCNN achieves the highest overall accuracy among all compared methods, surpassing deeper, pretrained architectures such as ResNet50 and EfficientNet-B0 while using a substantially shallower, custom-trained network with only three convolutional blocks. This indicates that, for EuroSAT-scale 64×64 patches, a compact CNN tailored to the dataset characteristics can match or exceed the performance of much deeper general-purpose networks, offering a favorable accuracy-to-efficiency trade-off for operational deployment.

4.8 Ablation Study

To quantify the contribution of individual architectural components, an ablation study was conducted by selectively removing batch normalization, dropout, and data augmentation from the proposed DCNN while keeping all other hyperparameters fixed. Table 7 reports the resulting overall accuracy and validation loss on the EuroSAT test set.

Table 7. Ablation study on the proposed DCNN (EuroSAT test set).

Configuration	Overall Accuracy (%)	Validation Loss
Full proposed DCNN	98.72	0.047
Without batch normalization	96.85	0.091
Without dropout	97.41	0.083
Without data augmentation	96.20	0.104
Without both BN and dropout	95.63	0.129

Removing batch normalization causes the largest single-component drop in accuracy (-1.87 points), consistent with its role in stabilizing gradient flow across the three convolutional blocks. Removing dropout produces a smaller but consistent drop (-1.31 points) alongside a higher validation loss, indicating increased overfitting. Disabling data augmentation similarly degrades generalization (-2.52 points), confirming that geometric and photometric augmentation meaningfully improves robustness to the natural variability present in Sentinel-2 acquisitions. The combined removal of batch normalization and dropout yields the largest overall degradation, underscoring that these two regularization mechanisms act complementarily rather than redundantly within the proposed architecture.

4.9 Computational Efficiency

Beyond predictive accuracy, computational efficiency is an important consideration for operational deployment. The proposed DCNN contains approximately 4.3 million trainable parameters, an order of magnitude fewer than VGG16 (~138 million) and roughly half that of ResNet50 (~25.6 million), while EfficientNet-B0 sits closer at approximately 5.3 million parameters. Average inference time per 64×64 patch on the GPU workstation described in Section 3.5 was measured at approximately 3.1 ms for the proposed DCNN, compared with approximately 9.8 ms for ResNet50 and 6.4 ms for EfficientNet-B0 under the same hardware and batch configuration. This favorable parameter-to-accuracy and latency-to-accuracy trade-off supports the suitability of the proposed architecture for near-real-time or resource-constrained EO monitoring pipelines, including onboard or edge-based satellite data processing scenarios where deeper pretrained backbones may be impractical.

4.10 Discussion

The experimental evaluation indicates that the proposed DCNN effectively captures hierarchical spatial features from Sentinel-2 imagery, yielding high classification accuracy across diverse land-cover categories. Batch normalization improves convergence stability, while dropout regularization mitigates overfitting despite the relatively small, compact network. Compared with conventional machine-learning methods, the proposed model demonstrates substantial gains in overall accuracy and class-wise performance, confirming the advantage of automated, end-to-end feature learning over handcrafted feature extraction. The ablation results in Section 4.8 further show that these gains are not incidental to a single design choice but arise from the combined effect of normalization, regularization, and augmentation acting together.

Nonetheless, residual misclassification persists among spectrally similar agricultural classes and between River and Sea/Lake, reflecting an inherent limitation of RGB/multispectral appearance-based discrimination at 64×64 resolution. Such confusion is consistent with prior EuroSAT-based studies [2,21,22] and suggests that spatial context beyond a single 64×64 patch, or additional spectral bands such as the short-wave infrared channels available in the full 13-band Sentinel-2 product, may be required to fully resolve these ambiguities.

Several limitations of the present study should be acknowledged. First, evaluation is restricted to the EuroSAT benchmark; generalization to other sensors, geographic regions, or higher-resolution imagery (e.g., BigEarthNet [32] or PlanetScope-derived datasets) has not been directly assessed and remains an important direction for future validation. Second, the model operates at the image-patch level and does not perform pixel-wise semantic segmentation [18,20], limiting its applicability to tasks requiring fine-grained spatial delineation of land-cover boundaries. Third, while the ablation study isolates the contribution of individual components, it does not exhaustively search the architectural hyperparameter space (e.g., filter counts, block depth, kernel size), so further accuracy gains may be achievable through systematic neural architecture search. Future work may incorporate attention mechanisms (e.g., Squeeze-and-Excitation or CBAM), multi-scale feature fusion, or hybrid CNN-Transformer architectures to further improve discrimination between visually similar categories, and may extend evaluation to additional benchmark datasets and higher-resolution imagery to assess generalization beyond EuroSAT.

5. Conclusion

This study presented a lightweight Deep Convolutional Neural Network for Land Use and Land Cover classification of Sentinel-2 EuroSAT imagery. The proposed architecture - three convolutional blocks with batch normalization and max pooling, followed by a dropout-regularized fully connected classifier - achieved an overall accuracy of 98.72%, precision of 98.65%, recall of 98.59%, F1-score of 98.62%, and a Cohen's Kappa of 0.984, outperforming both conventional machine-learning classifiers and deeper pretrained CNN architectures evaluated under identical conditions. These results demonstrate that a compact, purpose-built CNN can deliver state-of-the-art accuracy on EuroSAT-scale LULC classification while remaining computationally efficient, making it well suited to operational remote-sensing monitoring applications such as urban expansion tracking, agricultural assessment, and environmental change detection. Future research will explore attention-based and multi-scale extensions of the proposed architecture and evaluate generalization on larger and higher-resolution multispectral benchmarks.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436-444.

[2] P. Helber, B. Bischke, A. Dengel, D. Borth, EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (2019) 2217-2226.

[3] G. Camps-Valls, D. Tuia, X.X. Zhu, M. Reichstein, Deep learning for the Earth sciences: a comprehensive approach, Nat. Rev. Earth Environ. 2 (2021).

[4] J.A. Benediktsson, J. Chanussot, W.M. Moon, Very high resolution remote sensing: challenges and opportunities, Proc. IEEE 101 (2013).

[5] European Space Agency, Sentinel-2 User Handbook, ESA, 2024.

[6] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5-32.

[7] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273-297.

[8] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 4th ed., Pearson, 2018.

[9] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[10] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proc. NeurIPS, 2012.

[11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. CVPR, 2016.

[12] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. CVPR, 2017.

[13] M. Tan, Q. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: Proc. ICML, 2019.

[14] A. Howard, et al., MobileNets: efficient convolutional neural networks for mobile vision applications, arXiv:1704.04861, 2017.

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

[16] D. Marmanis, et al., Deep learning Earth observation classification using ImageNet pretrained networks, IEEE Geosci. Remote Sens. Lett. (2016).

[17] X.X. Zhu, et al., Deep learning in remote sensing: a comprehensive review and list of resources, IEEE Geosci. Remote Sens. Mag. (2017).

[18] S. Minaee, et al., Image segmentation using deep learning: a survey, IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[19] X. Ma, Y. Liu, L. Chen, Multi-scale deep learning for remote sensing image classification: a review, ISPRS J. Photogramm. Remote Sens. 205 (2023) 102-121.

[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proc. CVPR, 2015.

[21] H. Yassine, K. Tout, M. Jaber, Improving LULC classification from satellite imagery using deep learning - EuroSAT dataset, ISPRS Arch. (2021).

[22] Z. Li, R. Ran, J. Zhang, Remote sensing image classification based on deep convolutional neural networks, Remote Sens. (2022).

[23] T. Blaschke, Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens. (2010).

[24] D. Lu, Q. Weng, A survey of image classification methods and techniques for improving classification performance, Int. J. Remote Sens. (2007).

[25] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.

[26] K. Dosovitskiy, et al., An image is worth 16×16 words: transformers for image recognition at scale, in: Proc. ICLR, 2021.

[27] S. Khan, et al., Transformers in vision: a survey, ACM Comput. Surv. 54 (2022).

[28] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, A survey on vision transformer, IEEE Trans. Pattern Anal. Mach. Intell. 45 (2023) 87-110.

[29] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proc. CVPR, 2017.

[30] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proc. CVPR, 2018.

[31] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: Proc. ECCV, 2018.

[32] G. Sumbul, M. Charfuelan, B. Demir, V. Markl, BigEarthNet: a large-scale benchmark archive for remote sensing image understanding, in: Proc. IGARSS, 2019.

[33] G. Cheng, J. Han, X. Lu, Remote sensing image scene classification: benchmark and state of the art, Proc. IEEE 105 (2017).

[34] J.E. Ball, D.T. Anderson, C.S. Chan, A comprehensive survey of deep learning in remote sensing, J. Appl. Remote Sens. 11 (2017).

[35] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B.A. Johnson, Deep learning in remote sensing applications: a meta-analysis and review, ISPRS J. Photogramm. Remote Sens. 152 (2019) 166-177.

[36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proc. ICLR, 2015.

[37] https://github.com/phelber/EuroSAT.