Ingénierie des Systèmes d'Information
Vol. 31, No. 2, February, 2026, pp. 371-379
Journal homepage: http://iieta.org/journals/isi

Interpretable Land-Use Classification Using ResNet18 and SmoothGradCAM++: A Study on the UC Merced Dataset

S Jayanthi[1*], M. A. Josephine Sathya[2], B. Nathan[3], Karthik Karmakonda[4], Muthuvel Laxmikanthan[5], Manojkumar V[6]

[1] Department of Artificial Intelligence and Data Science, Faculty of Science and Technology (IcfaiTech), The ICFAI Foundation for Higher Education, Hyderabad 501203, Telangana, India
[2] Department of Computer Science and Applications, Christ Academy Institute for Advanced Studies, Bangalore 560100, Karnataka, India
[3] Department of Computer Science Engineering, Dhaanish Ahmed Institute of Technology, Coimbatore 641105, Tamil Nadu, India
[4] Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad 501510, Telangana, India
[5] Department of Artificial Intelligence and Data Science, Dhaanish Ahmed Institute of Technology, Coimbatore 641105, Tamil Nadu, India
[6] Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India

Corresponding Author Email: drsjayanthicse@gmail.com

https://doi.org/10.18280/isi.310205

Received: 31 August 2025
Revised: 15 November 2025
Accepted: 15 February 2026
Available online: 28 February 2026

Keywords:
Land-use classification, ResNet18; SmoothGradCAM++, interpretable machine learning, remote sensing imagery, convolutional neural networks

ABSTRACT

Land-use classification from remote sensing imagery is essential for applications like urban planning, environmental monitoring, and resource management. This study improves land-use classification by integrating deep learning and interpretability techniques. Using the UC Merced Land Use Dataset, we trained a ResNet18-based model to classify 21 land-use categories. To enhance model robustness, we applied data preprocessing techniques such as augmentation (random flips, rotations, and color jittering) and normalization. SmoothGradCAM++, an advanced variant of Grad-CAM, was used to generate class activation maps, providing visual interpretability by highlighting key regions influencing model predictions. The results show that preprocessing significantly improved classification accuracy, increasing it from 40.8% to 91.0%, with precision reaching 93.0% and an F1-score of 91.4%. SmoothGradCAM++ successfully localized important features, validating the model's decision-making process and confirming its interpretability. This study emphasizes the importance of preprocessing in remote sensing transfer learning and demonstrates how high accuracy can be achieved while ensuring model transparency. By focusing on both preprocessing and interpretability, our approach overcomes the limitations of traditional black-box models and enhances the reliability of land-use classification systems. This combined approach provides valuable insights for improving remote sensing applications in diverse fields, contributing to more transparent and effective land-use mapping.

1. INTRODUCTION

Over the past decade, land use classification has become a fundamental requirement for different applications in remote sensing imagery, such as urban planning, agricultural management, environmental monitoring, and disaster response, to name a few [1, 2]. Remote sensing offers a unique, holistic perspective on large spatial extents which makes it an invaluable method for understanding historical land use patterns. However, the complex variability in land-cover types, seasonal shifts, and variations in image resolution makes it challenging to achieve high accuracy in land-use classification. Traditional approaches to classification often fail to satisfy these requirements, especially concerning the interpretability of the results, which is key for validating model predictions and making them practically valuable for decision-making processes [3, 4].

The remote sensing domain has been progressively developed by increasing applications ranging from monitoring natural resources, forecasting urban developments, and precision agricultural analysis. The process of accurately classifying land uses remains challenging despite its importance. Conventional approaches, including manual annotation and rule-based classification, require substantial human effort and are prone to inconsistencies due to human bias. Also, most traditional machine learning approaches rely on hand-engineered features, which can be weak at addressing complex spatial hierarchies in image data. These restrictions and shortcomings also urge more automated and precise classification approaches that facilitate interpretability to the

end-user [5, 6].

In recent years, deep learning, especially convolutional neural networks (CNNs), has been gaining traction in remote sensing applications as it provides the capability of learning hierarchical and complex features directly from images without the need for domain-specific generated features. In this study, ResNet18 was selected due to its optimal trade-off between model depth and computational efficiency [7, 8]. By utilizing transfer learning with weights pre-trained on ImageNet, the model can be effectively adapted to identify nuanced class-specific spatial patterns in the dataset. Despite their high performance, deep learning models remain black boxes. This limits their application in critical areas where interpretability is crucial. To confront this, techniques like Grad-CAM and its improved version, SmoothGradCAM++, have been developed. These tools create class activation maps that emphasize the visual areas highlighted by the model, building trust in its predictions.

We propose a ResNet18 model with SmoothGradCAM++ visualization for land-use classification in remote sensing to improve the accuracy and interpretability. Our contributions are two-fold: we demonstrate that a ResNet18-based transfer learning framework and efficient preprocessing can result in a high classification accuracy rate on multi-class benchmark remote sensing dataset. Second, we validate our model with interpretable tools to gain insights into the model decision process, allowing for meaningfulness and practical validation. This interpretable approach addresses the limitations of classical classification techniques alongside the black-box nature of classifiers. It paves the way for future interpretable and accurate land-use classification studies for remote sensing problems.

Some critical gaps in knowledge are:

- A systematic evaluation of preprocessing impact on transfer learning performance in remote sensing classification.
- Integration of optimization strategies including partial fine-tuning, dropout regularization, adaptive learning rate scheduling, and early stopping.
- Class-level performance analysis using confusion matrices to examine misclassification behavior.
- Interpretability validation using SmoothGradCAM++, demonstrating alignment between model attention and semantically meaningful spatial features.

This study addresses these limitations by leveraging deep learning capability and interpretability.

2. RELATED WORKS

Current research shows that deep learning has evolved into a robust solution for LULC tasks. Aljebreen et al. [9] enhanced classification accuracy by employing the River Formation Dynamics Algorithm alongside deep learning to better capture geographical patterns. Huang et al. [10] used AI-based clustering techniques for optimizing rural land use that showed the possibility of multi-industry land management. Alem and Kumar [11] assessed the efficiency and scalability of various deep learning models on remote-sensing data in 2022.

Hybrid modeling strategies have also gained attention. Karakose [12] combined Fuzzy Cognitive Maps with enhanced loss functions to improve satellite image classification accuracy. Fan et al. [13] demonstrated the effectiveness of segmentation-based models for differentiating land-cover categories in resource survey applications. Dang et al. [14] applied CNNs to ALOS and NOAA satellite data for coastal area classification and demonstrated the versatility of deep learning across multiple satellite platforms. Other improvements include Irfan et al. [15], who proposed cascaded architectures with improved LULC classification, and Yu et al. [16], who applied deep transfer learning to the challenges of domain adaptation issues. Beyond remote sensing, Fukae et al. [17] show just how versatile CNNs can be by applying them to healthcare diagnostics. Our work contributes to LULC classification by combining high-performance deep learning with interpretability. This aspect makes our approach much more trustworthy and transparent by removing a key limitation in the literature [18-20].

This study is distinct from other studies by combining optimized transfer learning with systematic interpretability evaluation. While prior studies emphasize classification accuracy, fewer works provide a systematic evaluation of the impact of preprocessing, combined with interpretability validation.

Several challenges remain to affect land-use classification:

- Accurate categorization of diverse and challenging land uses.
- Although UC Merced is balanced, real-world remote sensing datasets often exhibit class imbalance, making the models biased.
- Dealing with noisy and incomplete data is necessary for proper classification.

This work addresses the above mentioned issues using a fine-tuned ResNet18 architecture and SmoothGradCAM++ for interpretability.

3. METHODOLOGY

3.1 Dataset overview

This study utilizes the UC Merced Land Use Dataset, a benchmark dataset for remote sensing scene classification [21]. It includes 21 land-use classes, each containing 100 images. Every image measures 256 × 256 pixels and is sourced from high-resolution 1-foot-per-pixel imagery of urban areas in the USGS National Map Urban Area Imagery collection.

To ensure a robust evaluation, we implemented a rigorous experimental setup by partitioning the dataset into 70:15:15. It produced 1,470 images for training, 315 for validation, and 315 for independent testing. The balanced class distribution facilitates an unbiased evaluation of the model's generalization ability in diverse geospatial features.

3.2 Preprocessing and data augmentation

During training, a range of data augmentation techniques was applied to strengthen the model's robustness and simulate diverse perspectives.

Data Augmentation techniques include:

- Horizontal flipping with a probability of 0.5 to simulate varying scene orientations.
- Random rotations with ± 15° to enhance robustness to aerial viewpoint variations in 2D Euclidean space.
- Random color jitter, such as brightness, contrast, saturation, and hue, to cope with different illumination of input images.
- The transformation matrix for a rotation by an angle θ is

defined in Eq. (1), where θ represents the rotation angle.

$$T = \begin{matrix} cos\theta & -\sin\theta & 0 \\ \sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{matrix} \quad (1)$$

These augmentation strategies increase spatial and photometric diversity within the training set. It enables the model to learn invariant and robust feature representations.

After augmentation, the images were resized to 224 × 224 pixels to meet the input requirements of ResNet18. For effective transfer learning, we used the mean (μ) and standard deviation (σ) of the ImageNet dataset [0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively. This standardization ensures alignment between the input data distribution and the ImageNet pre-trained weights, thereby enabling faster convergence and improved accuracy. The normalization process is described in Eq. (2).

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (2)$$

- x: Input pixel value
- μ: Mean of the dataset
- σ: Standard deviation of the dataset

The pixel values are scaled by normalizing, enabling the model to leverage pre-trained weights effectively. Sample images from the dataset, both with and without preprocessing, are shown in Figures 1 and 2, respectively.

Figure 1. Sample UC Merced dataset without preprocessing

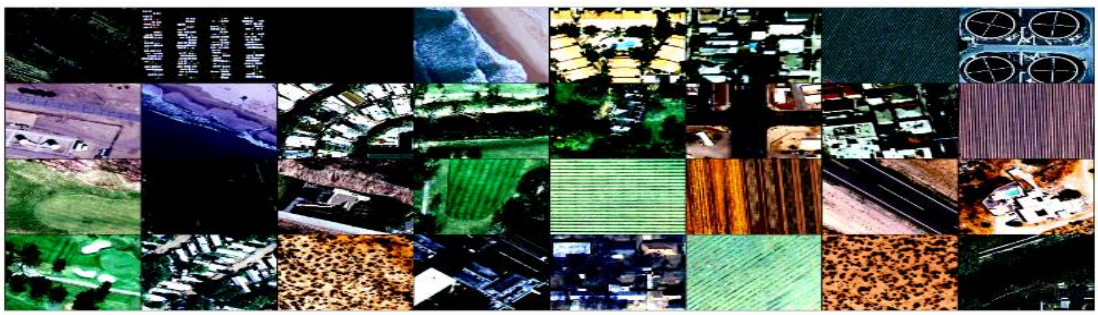


Figure 2. Sample UC Merced dataset with preprocessing

4. THE PROPOSED DEEP LEARNING MODEL

The proposed classification framework is built upon the ResNet18 architecture, pretrained on the ImageNet-1K (V1) dataset. The model employs residual blocks with identity shortcut connections that facilitate stable gradient propagation and mitigate vanishing gradient effects in deep convolutional networks.

To adapt the model architecture for classifying the dataset's 21 classes, we modified the final fully connected layer to output 21 logits. Furthermore, we implemented a fine-tuning strategy where the initial layers were frozen, and only the final residual block (layer 4) and the new classifier head were updated to specialize the model for geospatial feature extraction.

The model was trained using the AdamW optimizer, which decouples weight decay from gradient updates to improve regularization and generalization performance. Important hyperparameters of the model were set as follows:

- Learning Rate: 0.001; The learning rate was selected to solve the convergence and stability speed optimally.
- Batch Size: 32, chosen according to hardware limit and accurate to provide good robustness of gradient updates for the epochs with the weight decay: 1×10^{-4}
- Epochs: 10, enough for convergence, considering data set and model complexity
- Loss Function: Cross-entropy loss for multi-class classification tasks to train the model correctly for all 21 classes and
- Early stopping with a patience of 3 epochs

To further enhance convergence stability, a CosineAnnealingLR scheduler was employed. This scheduler gradually reduces the learning rate following a cosine function, allowing larger parameter updates during early training and finer adjustments toward later epochs.

Using Eq. (3), this measures the difference between the correct labels and the calculated probabilities. It ensures the model learns how to assign a high probability to correct

classes.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{21} y_{i,c}\log(p_i, c) \quad (3)$$

where, N denotes the number of samples, $y_{i,c}$ refers to the binary indicator (1 if class c is the correct classification for sample i, otherwise 0), p_i, c refers to the predicted probability for class c of sample i.

4.1 Interpretability with Grad-CAM

To address the interpretability limitations of deep neural networks, SmoothGradCAM++ was employed. This method extends Grad-CAM++ by incorporating higher-order gradient information and averaging over multiple noisy input perturbations, resulting in more stable and spatially precise activation maps.

The resulting heatmap was then interpolated to the size of the input image and superimposed on the original image, creating an interpretable overlay that points out areas on the image that contributed most to that particular prediction of the model. This process would produce qualitative and quantitative explainability about what the model has used to decide, making classification more explainable and enabling verification of classification outputs.

SmoothGradCAM++ was applied to visualize parts of the input image that more heavily contribute towards model classification using Eq. (4) for interpretability improvement.

$$L^{c}_{\text{Grad-CAM}} = ReLU(\sum_k \alpha_k^c A^k) \quad (4)$$

A^k denotes the activation map for the k-th feature channel in the selected convolutional layer, α_k^c (refer Eq. (5)) denotes the weights computed as the mean of gradients over spatial dimensions:

$$\alpha_k^c = \frac{1}{z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (5)$$

with Z denoting the total number of spatial locations.

The resulting heatmap is interpolated to the input image size and superimposed as an overlay. This provides qualitative evidence of the model's decision-making logic, allowing for the verification of whether the network focuses on semantically relevant geographic features (e.g., runways for the 'airplane' class) or background noise. The step-by-step implementation of the proposed deep learning model for land-use classification is shown in Figure 3.

First, it loads the dataset containing the training and testing images. Then, if preparation is used, it undergoes some pretreatment stages. Otherwise, it uses a simple data loader. In either case, it undergoes fundamental changes.

The model is based on a pre-trained ResNet18 architecture. Its fully connected layer is then adapted to the needs of land-use classification, such as changing the number of output neurons to correspond to the intended classification categories. The AdamW optimizer and the CrossEntropy loss function are used to train the modified model.

There are many performance metrics, such as accuracy, precision, recall, F1 score, and the confusion matrix, that measure the quality of a trained model. Performance metrics are compared, and conclusions are made on how preprocessing affects the model's accuracy and other performance metrics.

The workflow of the suggested deep learning model is clearly illustrated in this flowchart, which outlines the essential phases of data preparation, using a pre-trained ResNet18 architecture, model training, evaluation, and result analysis.

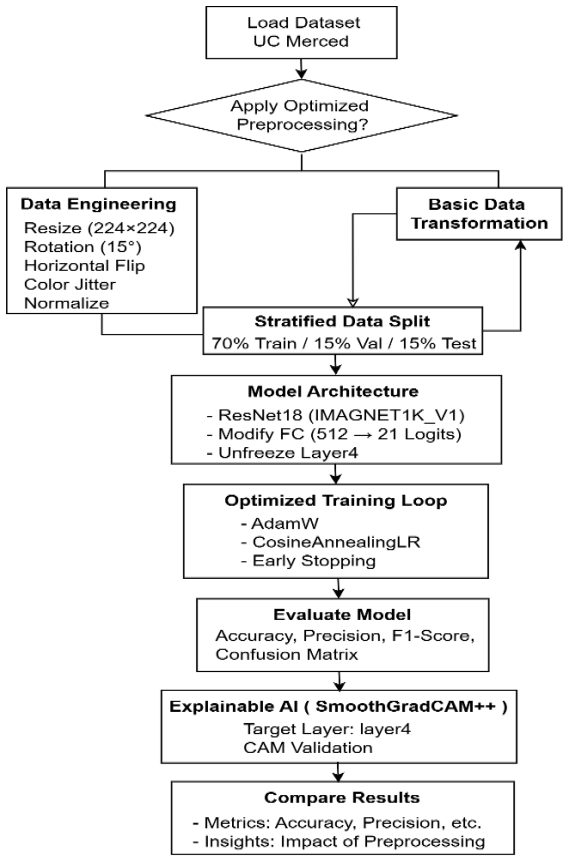


Figure 3. The process diagram for land-use classification

To understand the model's decision-making process and gain insights into which image features contribute most to its predictions, we employed SmoothGradCAM++. This technique generates class activation maps that highlight the regions of the input image with the greatest influence on the model's classification for a specific class.

Algorithm: ResNet18 Framework for Land-Use Classification with Grad-CAM

Input:

- Dataset: UCMerced Land Use dataset (21 classes)
- Base Model: ResNet18 pretrained on ImageNet-1K (V1).

Output:

- Performance metrics for classification (Accuracy, Precision, Recall, F1-score, Confusion Matrix)
- SmoothGradCAM++ visual explanations

Step 1: Dataset Preparation

- Split the dataset into training, validation, and testing sets (as defined in the experimental setup).
- Perform data augmentation techniques (e.g., random rotations, flips, Color jitter transformations) to the training dataset.
- Resize images to 224 × 224 pixels.
- Normalize pixel values using ImageNet statistics for

optimal model initialization.
- Convert images to tensor format for model input.

Step 2: Core Model Training
- Train ResNet18 as outlined in the standard process.
- Store intermediate results, including learned weights and feature maps.

Step 3: Explainability via Grad-CAM
- Select the final convolutional block (layer4) for activation mapping.
- Generate class-discriminative localization maps using SmoothGradCAM++.
- Upsample the activation maps to match input image resolution.
- Overlay heatmaps on the original images for visualization.
- Analyze attention regions for both correctly classified and misclassified samples.

Step 4: Model Evaluation
- Assess performance using evaluation metrics

Evaluate interpretability with SmoothGradCAM++ visualizations.

5. MODEL PERFORMANCE AND ANALYSIS

5.1 Experimental setup

Training and evaluation were performed using Google Colab with NVIDIA Tesla T4 GPU (16GB memory).

This implementation makes use of the following software libraries:
- Creating, training, and evaluating deep learning models with PyTorch.
- TorchCAM is used to produce SmoothGradCAM++ visualizations of the model's predictions.
- Using OpenCV for image processing tasks during data preparation and visualization.

5.2 Evaluation metrics

We assessed the model performance on multiple metrics:

Accuracy represents the percentage of correctly identified samples and is calculated using the formula in Eq. (6).

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{TTotal Number of samples}} \quad (6)$$

For multi-class classification, precision, recall, and F1-score were computed using weighted averaging across all 21 classes. Precision evaluates the ratio of true positive predictions to all positive predictions with Eq. (7).

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall indicates the proportion of true positive predictions relative to the total number of actual positives using Eq. (8).

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

The F1-score provides a balance between precision and recall by taking their harmonic mean using Eq. (9).

$$F1 - score = 2\frac{Precision\ .\ Recall}{Precision\ +\ Recall} \quad (9)$$

where, TP, FP, and FN denote true positives, false positives, and false negatives, respectively, computed per class.

The Confusion Matrix provides a class-wise breakdown of the predictions. This can highlight areas of misclassification and allows us to analyze class-wise performance.

5.3 Results

We compared performance for two model configurations: training without data preprocessing and with data preprocessing (augmentation). The following table (Table 1) presents evaluation metrics for both configurations.

The performance evaluation values obtained in the empirical study are depicted in Figure 4. Suppose the model is trained.

Without preprocessing, the model performs poorly, achieving an accuracy of only 40.8%, with a precision of 43.54%, a recall of 41.8%, and an F1-score of 40.8%. These results suggest that the model is unable to effectively learn complex land-use patterns from the images. In contrast, the model with preprocessing achieves significantly better performance, with an accuracy of 91.0%, a precision of 93.0%, a recall of 91.0%, and an F1-score of 91.4%, highlighting the important role of preprocessing and data augmentation in remote sensing classification.

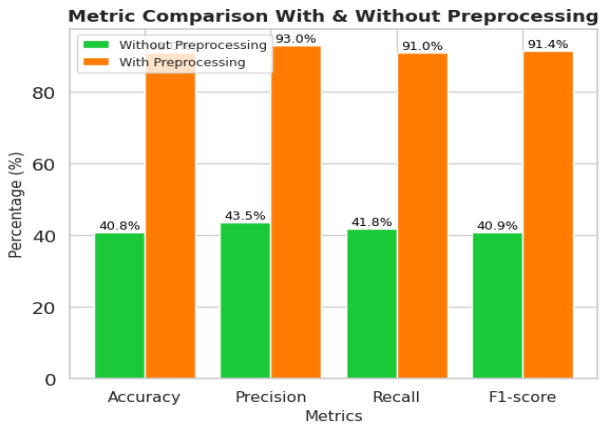


Figure 4. Performance evaluation of ResNet18

Table 1. Performance evaluation metrics of ResNet18 without and with preprocessing

Model Configuration	Accuracy	Precision	Recall	F1-Score
Without Preprocessing	40.8%	43.54%	41.8%	40.87%
With Preprocessing	91.0%	93.0%	91.0%	91.4%

5.3.1 Training loss
- The model, trained without preprocessing data, showed slow and unstable convergence with high loss values and marginal decreases in loss during epochs:
- Number of epochs,1-10 (Without Preprocessing): The average loss values went down from 4.67 to 3.45, but the overall accuracy was still relatively low (4.8%).
- The preprocessed model exhibited loss consistently

decreasing over epochs, signifying a steady training convergence:

- Epochs 1-10 (With Preprocessing): The loss values reduced constantly from 1.02 to 0.14, in line with the model's final accuracy near the high end.

5.3.2 Confusion matrix analysis

- Both configurations' confusion matrices provided detailed information on class-specific performance:
- Without Preprocessing: The standard matrix for the model corroborated this, showing widespread misclassifications across all categories, but having the most trouble distinguishing between visually similar categories such as "dense residential" and "medium residential".
- With Preprocessing: On the other hand, the confusion matrix reveals many correct classifications for the model with preprocessing.

Few classes were misclassified between visually overlapped categories (for example, between "forest" and "agricultural"). But despite this, the model was capable of differentiating at least the majority of land-use types. For instance, the classes "tennis court" and "airplane" had near-perfect classification, with 98 and 100 samples correctly predicted, respectively. The confusion matrix analysis without and with preprocessing is displayed in Figures 5 and 6, respectively. A sample of correctly classified images across Five Land-Use Classes is illustrated, and Loss over 10 epochs before and after preprocessing is shown in Figures 7 and 8, respectively.

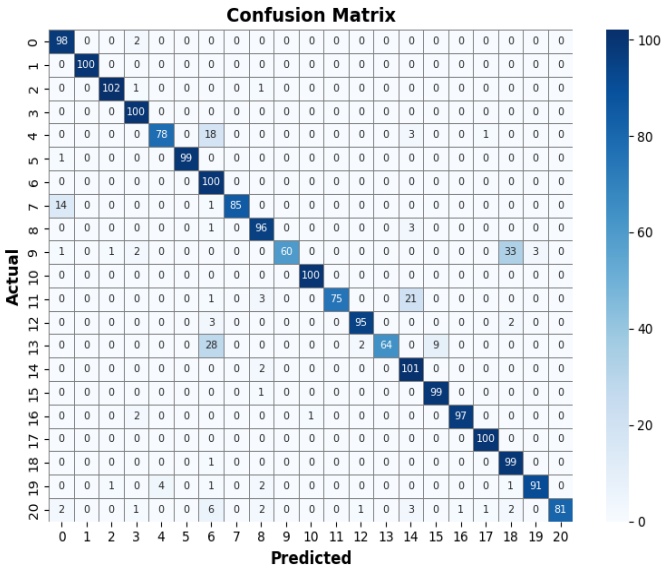


Figure 5. Confusion matrix without preprocessing

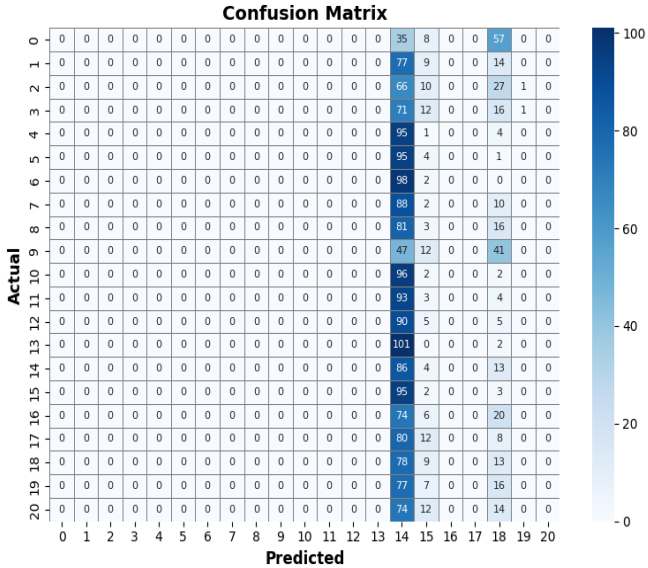


Figure 6. Confusion matrix with preprocessing

Figure 7. Sample of correctly classified images across Five Land-Use Classes

Figure 8. Loss across epochs

As expected, in classes such as "freeway" and "golf course," the highest SmoothGradCAM++ overlays highlighted appropriate visual regions such as roads and green areas. This suggests that the model picked up the correct visual regions for classification. For classification misclassifications, e.g., "medium residential" vs "dense residential" areas, the visualizations showed overlapping focus locations, likely due to similar features in these classes. These visualizations confirmed that the model's decision-making aligned well with key features in the imagery, especially after preprocessing, thereby validating the model's reliability and interpretability.

6. DISCUSSION

6.1 Analysis of results

The experimental results demonstrate that the proposed

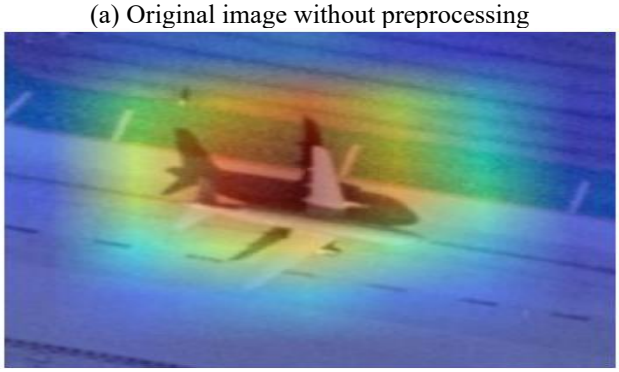
optimized ResNet18 framework significantly outperforms the baseline. The leap from 40.8% to 91.0% accuracy is attributed to the synergistic effect of advanced data augmentation and the AdamW optimizer with Cosine Annealing. As shown in our ablation study, augmentation methods such as random rotations (15°) and color jittering provided the model with the necessary spatial and spectral variance to generalize beyond the limited training samples.

Interpretability analysis using SmoothGradCAM++ provided insights into the model's decision-making process. For classes that had dedicated elements for their class, such as "airplane" and "tennis court, " our model focused closely on the appropriate structures, such as airplanes sitting on runways or the lines on a tennis court.

A key contribution of this work is the validation of model decisions through SmoothGradCAM++. For distinct classes such as "airplane" and "tennis court," the model achieved near-perfect accuracy (98–100%) by successfully localizing unique geometric structures.

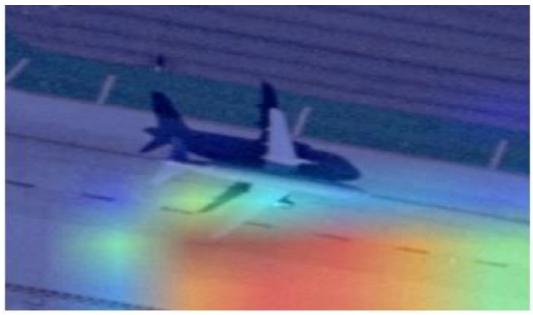
However, the Confusion Matrix (Figure 6) highlights a persistent challenge: the overlap between "dense residential" and "medium residential" classes. From a feature perspective, these categories share high-frequency textural patterns (individual rooftops) and similar spectral signatures (pavement and vegetation). Figure 9(a) shows the original aerial image without preprocessing. Our analysis revealed that in misclassified instances, the model's attention was dispersed across the entire residential block rather than identifying specific density-defining features. This suggests that while ResNet18 is highly capable of feature extraction, these visually similar classes may require higher-resolution imagery or additional contextual metadata to be fully resolved and Figure 9(b) depicts the same after resizing, normalization, and augmentation. Input normalization and preprocessing transformation steps are essential to access the ResNet18 model.

(a) Original image without preprocessing

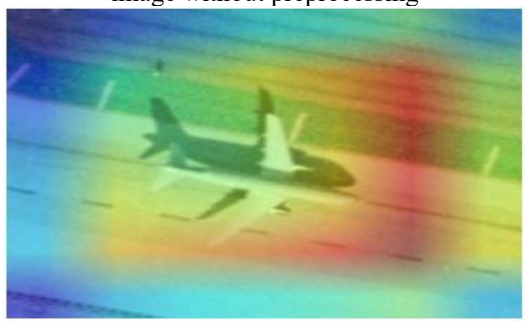


(b) Image after preprocessing

Figure 9. (a) Original image without preprocessing, (b) Image after preprocessing



(a) SmoothGradCAM++ heatmap overlay on the original image without preprocessing



(b) SmoothGradCAM++ heatmap overlay on the preprocessed image

Figure 10. (a) SmoothGradCAM++ heatmap overlay on the original image without preprocessing, (b) SmoothGradCAM++ heatmap overlay on the preprocessed image

6.2 Limitations and future directions

Unlike previous studies that often treat deep learning as a "black box," this research provides a systematic validation of interpretability on the UCMerced dataset. By employing SmoothGradCAM++, which utilizes higher-order derivatives, we produced more stable and precise activation maps than standard Grad-CAM. This interpretability is vital for high-stakes geospatial applications, such as urban planning or disaster response, where understanding the *why* behind a classification is as important as the accuracy itself.

Despite these gains, the model's performance is limited by the inter-class similarity of the residential sectors. Future work will investigate the use of Vision Transformers, which leverage global self-attention to capture broader spatial contexts that may better differentiate dense from medium residential layouts. Furthermore, implementing model ensembles or multi-scale feature fusion could refine the detection of fine-grained urban textures.

Future research directions include:

- Evaluating deeper convolutional architectures such as ResNet50 or EfficientNet to capture higher-level semantic representations.
- Investigating transformer-based models (e.g., Vision Transformers) to improve global contextual modeling.
- Incorporating multi-temporal imagery to leverage temporal dynamics for improved class differentiation.
- Integrating multispectral or LiDAR data to enhance spectral and structural discrimination.
- Exploring complementary interpretability techniques such as SHAP or Layer-wise Relevance Propagation (LRP) for enhanced explainability.
- Validating the framework on larger and more diverse remote sensing datasets to assess scalability and

generalization robustness.

- Addressing these aspects may further enhance classification performance and broaden real-world applicability.

7. CONCLUSIONS

Through this study, a fine-tuned ResNet18 model deep learning approach was thus used to achieve accurate and interpretable classification of land use successfully. The model reached an accuracy of 91.0% on the Land Use Dataset through the approach of using data augmentation transfer learning, and SmoothGradCAM++ interpretability, the proposed approach achieved a classification accuracy of 91.0%.

These findings highlight the prospects for deep learning and interpretability techniques to enhance development in remote sensing. Future research avenues may include developing more sophisticated networks, like Vision Transformers, to improve the classification performance further. Also, integrating more data, such as LiDAR or multispectral images, could enrich the information for more accurate land use classification. Addressing limited data availability and high class imbalance in large datasets will enable the design of better, more flexible land-use classification models.

REFERENCES

[1] Temenos, A., Temenos, N., Kaselimi, M., Doulamis, A., Doulamis, N. (2023). Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. IEEE Geoscience and Remote Sensing Letters, 20: 1-5. https://doi.org/10.1109/LGRS.2023.3251652

[2] Bhosle, K., Musande, V. (2019). Evaluation of deep learning CNN model for land use land cover classification and crop identification using hyperspectral remote sensing images. Journal of the Indian Society of Remote Sensing, 47: 1949-1958. https://doi.org/10.1007/s12524-019-01041-2

[3] Douass, S., Ait Kbir, M. (2022). Deep learning approach for land use images classification. E3S Web of Conferences, 351: 01043. https://doi.org/10.1051/e3sconf/202235101043

[4] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 618-626. https://doi.org/10.1109/ICCV.2017.74

[5] de Camargo, T., Schirrmann, M., Landwehr, N., Dammer, K.H., Pflanz, M. (2021). Optimized deep learning model as a basis for fast UAV mapping of weed species in winter wheat crops. Remote Sensing, 13(9): 1704. https://doi.org/10.3390/rs13091704

[6] Zhao, S.Y., Tu, K.W., Ye, S.T., Tang, H., Hu, Y.C., Xie, C. (2023). Land use and land cover classification meets deep learning: A review. Sensors, 23(21): 8966. https://doi.org/10.3390/s23218966

[7] Helber, P., Bischke, B., Dengel, A., Borth, D. (2017). 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, 12: 2217-2226. https://doi.org/10.1109/JSTARS.2019.2918242

[8] Kadhim, S.H., Al-Jawari, S.M., Hasach, N.A.R. (2024). Analyzing earth's surface temperatures with relationship to land urban land cover (LULC) to enhance sustainability. International Journal of Sustainable Development and Planning, 19(1): 123-130. https://doi.org/10.18280/ijsdp.190110

[9] 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, 12: 11147-11156. https://doi.org/10.1109/ACCESS.2023.3349285

[10] Huang, Q., Xia, H.B., Zhang, Z.C. (2023). Clustering analysis of integrated rural land for three industries using deep learning and artificial intelligence. IEEE Access, 11: 110530-110543. https://doi.org/10.1109/ACCESS.2023.3321894

[11] Alem, A., Kumar, S. (2022). Deep learning models performance evaluations for remote sensed image classification. IEEE Access, 10: 111784-111793. https://doi.org/10.1109/ACCESS.2022.3215264

[12] Karaköse, E. (2024). An efficient satellite images classification approach based on fuzzy cognitive map integration with deep learning models using improved loss function. IEEE Access, 12: 141361-141379. https://doi.org/10.1109/ACCESS.2024.3461871

[13] Fan, Z.Y., Zhan, T., Gao, Z.C., Li, R., Liu, Y., Zhang, L.Z. (2022). Land cover classification of resources survey remote sensing images based on segmentation model. IEEE Access, 10: 56267-56281. https://doi.org/10.1109/ACCESS.2022.3175978

[14] Dang, K.B., Dang, V.B., Bui, Q.T., Nguyen, V.V., Pham, T.P.N., Ngo, V.L. (2020). A convolutional neural network for coastal classification based on ALOS and NOAA satellite data. IEEE Access, 8: 11824-11839. https://doi.org/10.1109/ACCESS.2020.2965231

[15] Irfan, A., Sun, G.M., Li, Y., Zhang, H.S. (2024). Cascaded deep learning model for accurate land use and land cover classification. In IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, pp. 3077-3080. https://doi.org/10.1109/IGARSS53475.2024.10641891

[16] Yu, F.C., Xiu, X.C., Li, Y.H. (2022). A survey on deep transfer learning and beyond. Mathematics, 10(19): 3619. https://doi.org/10.3390/math10193619

[17] Fukae, J., Isobe, M., Hattori, T., Fujieda, Y., et al. (2020). Convolutional neural network for classification of two-dimensional array images generated from clinical information may support diagnosis of rheumatoid arthritis. Scientific Reports, 10: 5648. https://doi.org/10.1038/s41598-020-62634-3

[18] Yifter T.T., Razoumny Y.N., Lobanov V.K. (2022). Deep transfer learning of satellite imagery for land use and land cover classification. Informatics and Automation, 21(5): 963-982. https://doi.org/10.15622/ia.21.5.5

[19] Natya, S., Manu, C., Anand, A. (2022). Deep transfer learning with RESNET for remote sensing scene classification. In 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, pp. 1-6.

https://doi.org/10.1109/ICDSIS55133.2022.9915967
[20] Kurian, V., Jacob, V., Kuruvilla, J. (2024). Approach of transfer learning in remote sensing image classification. In 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST), Kochi, India, pp. 1-3. https://doi.org/10.1109/ICTEST60614.2024.10576178
[21] Yang, Y., Newsam, S. (2010). Bag-of- visual-words and spatial extensions for and-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, California, pp. 270-279. https://doi.org/10.1145/1869790.1869829