Evaluation of U-Net Variants and Traditional Machine Learning Methods for Land Cover Classification Using High-Resolution Satellite Imagery

Aydin Ebrahimi[1*], Amirhossein Ghourkhanehchi Zirak[2], Amin Sedaghat[2], Nazila Mohammadi[2]

[1] Department of Photogrammetry and Remote Sensing, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran 19967-15433, Iran - a.ebrahimi3@email.kntu.ac.ir

[2] Department of Geomatics Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz 5166616471, Iran - a.ghourkhanehchizirak@gmail.com; a.sedaghat@tabrizu.ac.ir; n.mohammadi@tabrizu.ac.ir

Keywords: Remote sensing, Land cover classification, Deep neural networks, High-resolution imagery, Machine learning

Abstract

Recent advancements in deep learning have led to improvements across fields including computer vision, biomedical engineering, and geospatial analysis. Deep neural networks (DNNs) are prior at extracting complex spatial features from large-scale data, enabling accurate automated interpretation of remote sensing imagery. Land cover classification and segmentation are critical for urban development, environmental monitoring, and agricultural planning. As high-resolution satellite data become more accessible, demand for precise classification methods grows. This study investigates DNN performance for land cover classification and segmentation using high-resolution DigitalGlobe satellite imagery with three bands, specifically the DeepGlobe Land Cover Classification dataset derived from WorldView-3 imagery. Six land cover classes are examined: urban land, agricultural land, rangeland, forest land, water, and barren land. The dataset is divided into 70% training, 20% validation, and 10% testing subsets. Deep learning models (AttUnet, Unet++, and U-Net) are implemented, with evaluation metrics including accuracy, F1-score, and recall. We analyze test data using trained models and compare results to traditional algorithms to assess robustness and generalization. Results show Unet++ achieved an F1-score of 84.21%, accuracy of 90.32%, and recall of 87.46%, demonstrating superior performance. AttUnet and U-Net followed with F1-scores of 82.12% and 77.08%. SVM and RF performed good but lower, with F1-scores of 65.64% and 70.21%. Unet++ showed better performance in classifying water and agriculture, excelling at identifying boundaries. This makes Unet++ the most effective model, especially in heterogeneous regions with complex landscapes. The comparative analysis highlights its advantages in achieving higher segmentation accuracy and spatial consistency.

1. Introduction

In recent years, deep learning has transformed remote sensing and image classification. It provides new ways to analyze satellite data with impressive precision and scale (Ahmadian et al., 2024). One of the most significant developments is the U-Net architecture, which has become essential for semantic segmentation in areas like land-cover classification (Zang et al., 2021). Initially created for medical imaging, U-Net has shown great effectiveness for pixel-level classification, capturing both fine spatial details and broader patterns (Ronneberger et al., 2015). Its encoder and decoder structure maintains spatial context while learning detailed features. This makes it well-suited for interpreting high-resolution satellite images. Land-cover classification is a key task in remote sensing and supports practical applications in urban planning, agricultural management, and environmental monitoring (Li et al., 2022; Ulmas and Liiv, 2020). With the increasing availability of high-resolution imagery, analysts can extract more detailed information about the Earth's surface. However, this added detail also brings new challenges. Complex spatial patterns and mixed land types make it hard for traditional machine-learning methods to maintain accuracy. Techniques like Support Vector Machines (SVMs) and Random Forests (RFs) have been widely used (Liu et al. 2017; Rissati et al. 2020)., yet they often struggle to identify subtle boundaries or interactions among similar land-cover classes. Deep learning models, particularly those based on U-Net, have solved many of these problems. Recent versions like U-Net++ and Attention U-Net (AttUnet) build on the original framework to model more complex spatial relationships and improve segmentation quality (Kumar and Chaudhari 2024; Li et al. 2024). U-Net++ adds nested skip pathways to enhance feature flow between the encoder and decoder. AttUnet uses attention mechanisms that direct the model toward the most relevant areas in the image. Together, these features make U-Net derivatives especially effective for high-resolution remote sensing, where small differences between land-cover types can be hard to notice. This study compares three popular U-Net-based architectures: U-Net, U-Net++, and AttUnet, using high-resolution imagery from the DeepGlobe dataset. This dataset includes six land-cover categories: urban, agriculture, rangeland, forest, water, and

[*] Corresponding author

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-205-2026

barren. All models were trained under the same conditions, using a 70/20/10 train-validation-test split. They were evaluated with the same preprocessing and augmentation methods. U-Net acts as the baseline, U-Net++ improves feature flow through more skip connections, and AttUnet adds attention modules to focus on key spatial features. Performance was measured using accuracy, recall, and F1-score. To provide context, two traditional classifiers, SVM and Random Forest, were included as baseline models. While deep learning methods usually outperform traditional techniques on complex imagery, comparing them is useful for understanding the extent of improvement and the trade-offs in computational cost and resource use. By including both traditional and deep learning approaches, this study offers a balanced, reproducible comparison. It aims to illustrate when advanced architectures provide significant benefits and when simpler models may still be effective. Ultimately, the goal is to clarify how widely used U-Net variants perform under consistent experimental conditions and to provide practical insights for land-cover mapping in diverse urban and agricultural areas.

2. Methodology

2.1 Dataset Description

In this study, the DeepGlobe Land Cover Classification dataset was used for analyzing different land-cover types. The dataset is derived from WorldView-3 satellite imagery and provides high-resolution RGB scenes that include three color bands red, green, and blue. These bands make it possible to separate and identify a variety of surface materials and land-cover categories. Six key classes were considered: urban land, agricultural areas, rangeland, forested land, water, and barren ground. The satellite imagery focuses on the Salt Lake City region in Utah, one of the key sites featured in the DeepGlobe Land Cover Classification Challenge (2018) (Demir et al., 2018), as shown in Figure 1. The area includes a mix of urban development, farmland, and natural terrain, making it an effective test site for models designed to classify complex landscapes. Each image is paired with pixel-level ground-truth labels that define the six categories. Before model training, the imagery was pre-processed for alignment and scaling to maintain consistency across the dataset. Approximately 1,800 RGB ortho-tiles from the Salt Lake City region were analyzed. To prevent spatial overlap between subsets, the split was made at the image level rather than at the patch level. Around 70 % of the data (1,260 images) were used for training, 20 % (360 images) for validation, and 10 % (180 images) for testing. This division provided a well-balanced setup for developing and tuning the models, as well as for assessing performance in an unbiased way.

Figure 1. Satellite image Overview of Salt Lake City, UT, USA

2.2 Model Selection

This study compares three deep learning models U-Net, U-Net++, and Attention U-Net that all trace their design back to the original U-Net framework. The U-Net architecture is well known for its encoder–decoder layout, which links corresponding layers through skip connections to retain spatial detail during pixel-level classification. U-Net++ modifies this design by adding nested skip pathways, allowing features to flow more effectively through the

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-205-2026 |

network and improving segmentation performance, especially around irregular or complex boundaries. Attention U-Net takes a different step by using attention modules that help the model focus on key spatial regions, leading to better results in areas where class boundaries are unclear or visually subtle. In remote sensing, U-Net–style networks remain a popular choice because they can accurately capture object boundaries, train well on limited datasets, and remain computationally efficient. To ensure an even comparison, all three models in this study were trained under the same conditions. More advanced segmentation models, such as DeepLabV3+, PSPNet, and transformer-based variants, were intentionally left out. This decision keeps the focus on interpretability and reduces computational demands while still allowing a fair analysis within one architectural family. These points are revisited later in the Discussion section.

2.3 Data Preprocessing

Before training, each image was cut into smaller 256 × 256-pixel RGB tiles using a stride of 256, so that no patches overlapped. This process produced 1,260 training samples, 360 for validation, and 180 for testing. Each patch was then normalized by color channel using simple min–max scaling, which mapped pixel values into the [0, 1] range and kept input distributions consistent across all models. To improve robustness and limit overfitting, only the training set underwent data augmentation. In this case, we applied random horizontal and vertical flips (each with a 50 % chance), rotations of 0°, 90°, 180°, or 270° (also with 50 % probability), and slight random rescaling within the range of about 0.9 to 1.1. Because the dataset is naturally imbalanced urban and agricultural areas appear more often than the other classes the sampling strategy was adjusted. Uniform patch sampling was combined with targeted augmentations to help balance class representation. The ground-truth labels were stored as single-channel, integer-encoded images, where pixel values from 0 to 5 correspond to the six land-cover categories: urban, agriculture, rangeland, forest, water, and barren land.

2.4 Model Training

The three models U-Net, U-Net++, and AttUnet were trained using roughly seventy percent of the full dataset. Training relied on the sparse categorical cross-entropy loss function, which works well for multi-class problems with integer-encoded labels. Optimization was handled by the Adam algorithm, starting with a learning rate of 0.001. As training progressed, the learning rate was gradually reduced to help the models settle into better minima and to lower the risk of overfitting. Each network was trained for as many as eighty epochs, though the process could stop earlier if the validation loss stopped improving. This early-stopping rule prevented the models from continuing once learning had stabilized. When the validation accuracy or loss remained unchanged for several rounds, the learning rate was also reduced so the optimizer could make smaller adjustments. A batch size of sixteen images was used, which offered a reasonable balance between training stability, GPU memory, and processing time.

2.5 Computational Resources

The models were trained on Google Colab Pro using a T4 GPU with 15 GB of RAM and 40 GB of GPU memory. This setup allowed for efficient training of the models, reducing the time required for training and enabling the use of deep learning models on a large dataset.

2.6 Model Evaluation

In this work, the trained models were tested on about ten percent of the dataset, which was kept aside to get an unbiased sense of how well each model performed. The evaluation mainly relied on three numerical measures: overall accuracy, recall, and the F1-score. Accuracy gives a general idea of how many pixels were correctly labeled across all land-cover types. Recall looks at how many of the true pixels for a given class were actually found by the model, which is useful for seeing how sensitive the model is to smaller or less common categories. The F1-score, which combines precision and recall through their harmonic mean, helps balance the trade-off between missing pixels and misclassifying them. It's particularly handy when the dataset is uneven across classes. All these metrics were calculated at the pixel level for six land-cover categories urban, agriculture, rangeland, forest, water, and barren. Their macro-averaged scores were then taken so that each class had equal influence on the final results. To understand differences among land types, I also compared the class-wise F1 and recall values. Lastly, we checked whether the differences between models were meaningful by running paired tests on the image-level F1-scores and accuracies. A p-value below 0.05 was treated as significant. This setup offered a straightforward and fair way to compare deep learning models with the traditional machine-learning ones.

3. Results

The models were evaluated based on three key metrics: overall accuracy**,** F1-score**,** and recall**.** These metrics were chosen to assess the models' ability to correctly classify land cover types and their robustness in handling different land cover classes. The Table 1 presents the results for each model. From the performance metrics, it is evident that U-Net++ outperforms both AttUnet and U-Net across all evaluation metrics. Specifically, U-Net++ achieved an F1-score of 84.21%, accuracy of 90.32%, and recall of 87.46%, demonstrating its superior capability in classifying land cover types accurately, especially in complex landscapes. In comparison, AttUnet and U-Net achieved F1-scores of 82.12% and 77.08%, respectively, with U-Net++ showing a clear edge in both overall performance and consistency. Traditional machine learning models such as SVM (F1-score of 65.64%, accuracy of 73.21%, recall of 74.72%) and RF (F1-score of 70.21%, accuracy of 76.54%, recall of 75.09%) performed significantly lower, particularly in handling complex and heterogeneous landscapes, further

Model	F1-Score	Accuracy	Recall
Unet++	84.21%	90.32%	87.46%
AttUnet	82.12%	87.65%	83.41%
U-Net	77.08%	82.94%	80.23%
SVM	65.64%	73.21%	74.72%
RF	70.21%	76.54%	75.09%

Table 1. Performance comparison of U-Net variants and traditional models on the DeepGlobe dataset (F1-score, accuracy, and recall).

emphasizing the superiority of deep learning models for this task. The performance of the models was also evaluated on a class-by-class basis. Unet++ demonstrated particularly strong performance in classifying challenging land cover types such as water and agriculture, where the distinction between classes is often subtle. Additionally**,** Unet++ showed better delineation of boundaries, particularly in areas where land cover types are interspersed, such as between urban areas and agricultural land. These results highlight Unet++'s ability to accurately identify fine-grained differences in the landscape. In terms of model stability, Unet++ exhibited consistent improvement throughout the training process. The model showed a steady increase in accuracy and recall, with no significant signs of overfitting, indicating its robustness to variations in the dataset. On the other hand, while both AttUnet and Unet showed improvements in early epochs, their performance plateaued sooner than Unet++**,** suggesting that Unet++ may benefit from a more effective feature extraction and propagation mechanism, allowing for better convergence during training.

To complement the quantitative metrics, we present visual comparisons of the predicted land cover classifications for the three models, as depicted in Figure 2. The following figure illustrates the segmentation results for a sample region from DeepGlobe dataset.

The figure clearly demonstrates that Unet++ provides more accurate and detailed segmentation, particularly in regions with complex land cover boundaries, such as urban vs agricultural land. Unet++ also excels in correctly classifying small and scattered features like water bodies and barren land. The segmentation from AttUnet and U-Net shows more apparent misclassifications, especially along class boundaries where the models struggled to define precise distinctions. As depicted in Figure 3, a visual comparison of the land cover segmentation results from the U-Net, U-Net++, and AttUnet models is shown. The top row presents the original satellite images, and the following rows display the predicted classifications. The superimposed map clearly visualizes that U-Net++ outperforms the other models, particularly in terms of boundary delineation and accurate classification of complex land cover types.

4. Discussion

In this research, we compared several well-known models for land cover classification using high-resolution images from the DeepGlobe dataset. Three deep learning architectures U-Net, U-Net++, and AttUnet were evaluated along with two classical machine-learning algorithms, Support Vector Machines (SVM) and Random Forests (RF). Among all models, U-Net++ gave the strongest results, reaching an F1-score of 84.21%, an accuracy of 90.32%, and a recall of 87.46%. These results show that its nested skip connections help the model preserve spatial detail and handle the complex scene often present in real landscapes. Even though the deep learning models clearly outperformed the traditional ones, SVM and RF still produced meaningful outcomes. The SVM achieved an F1-score of 65.64%, accuracy of 73.21%, and recall of 74.72%. The RF model performed a bit better with an F1-score of 70.21%, accuracy of 76.54%, and recall of 75.09%. While these values are lower than those of U-Net++, they're reasonable when hardware or time is limited, or when a simpler model is easier to interpret and deploy. This study focused mainly on U-Net-based designs to keep the setup consistent and to control external factors. In future work, it would be worthwhile to test DeepLab architecture or transformer-based networks to see how they perform under similar conditions. Overall, U-Net++ appears to be the most reliable and precise among the tested models, but traditional approaches still hold value in practical, resource-constrained scenarios.

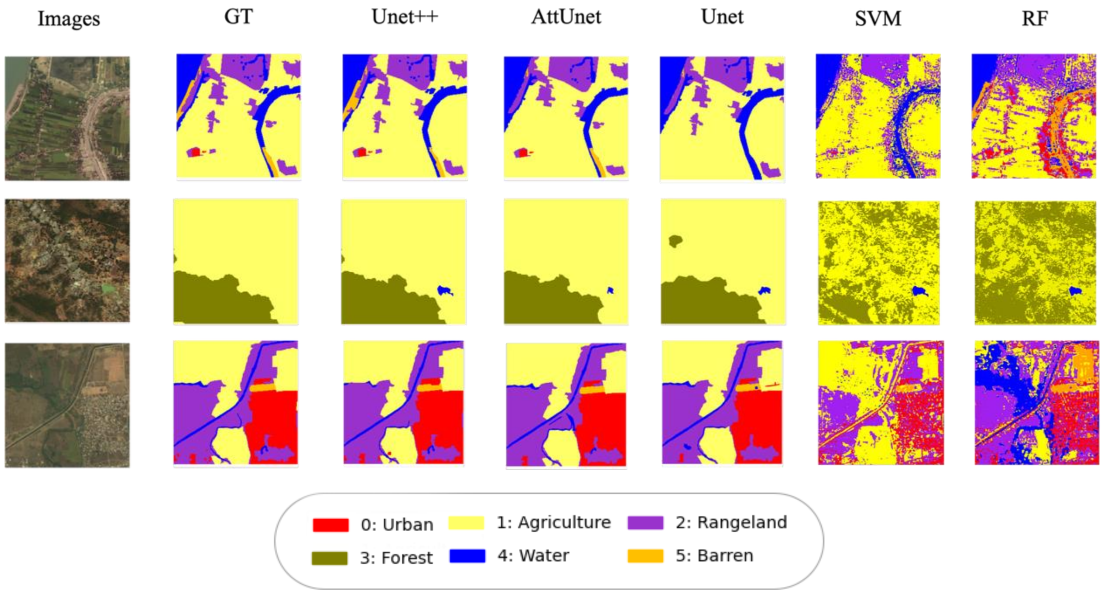

Figure 2. Visual comparison of land cover classification results by U-Net, Unet++, AttUnet, SVM, and RF

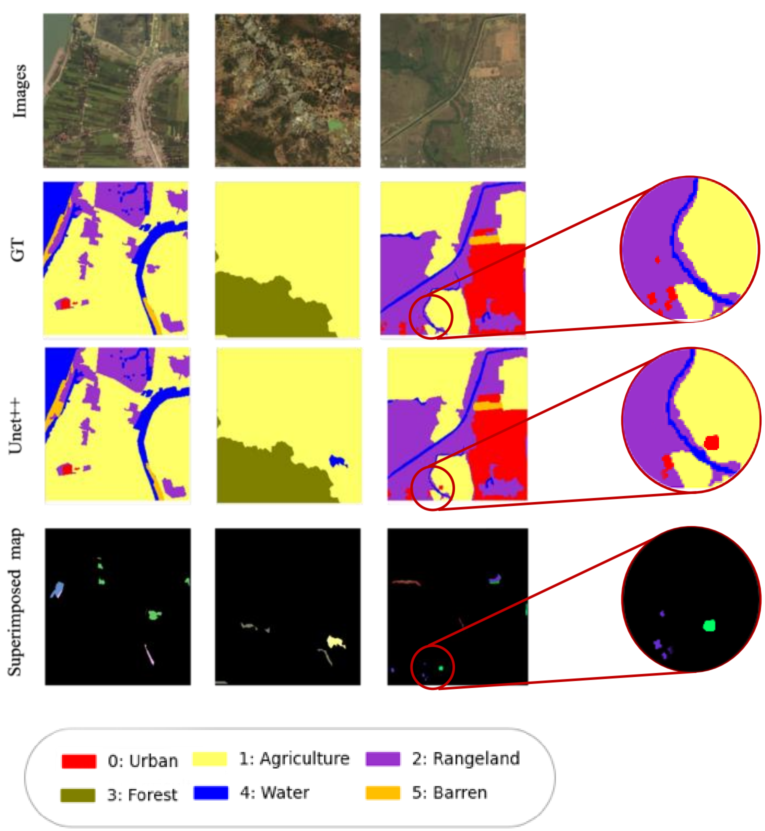

Figure 3. Visual comparison of land cover classification results for U-Net, U-Net++, and AttUnet models, showing the original satellite images and their corresponding predicted classifications.

5. Conclusion

The results of this study highlight the significant advantage of deep learning models, particularly U-Net++, in land cover classification using high-resolution satellite imagery. U-Net++ outperformed both AttUnet and U-Net, demonstrating superior classification accuracy, especially in complex landscapes with mixed land cover types. Its advanced architecture, with nested skip pathways, enabled better feature propagation and boundary delineation. While deep learning models offer higher accuracy, traditional machine learning methods like SVM and RF still have their place. SVM performed adequately in terms of recall but lagged behind deep learning models in classification accuracy. RF performed slightly better than SVM but still struggled with the spatial complexity of high-resolution data. These traditional methods are computationally less demanding and can be valuable when resources are limited or when interpretability is crucial.

Overall, U-Net++ is the best option for high-resolution remote sensing tasks, offering the best trade-off between accuracy and spatial consistency. However, SVM and RF remain useful alternatives, especially in resource-constrained environments or less complex datasets. Future work could explore hybrid models that combine deep learning with traditional methods to leverage the strengths of both. Future work could explore transfer learning and hybrid approaches that combine deep learning with traditional machine-learning methods, including transformer-based networks integrated with Kolmogorov models.

References

Ahmadian, N., Sedaghat, A., & Mohammadi, N, 2024: Building footprint extraction from Remote sensing images with residual attention Multi-scale Aggregation fully Convolutional Network. *Journal of the Indian Society of Remote Sensing, 52*, 2417-2429

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., & Raskar, R, 2018: Deepglobe 2018: A challenge to parse the earth through satellite images. In, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 172-181)

Kumar, D.G., & Chaudhari, S, 2024: Auxg: Deep feature extraction and classification of remote sensing image scene using attention unet and xgboost. *Journal of the Indian Society of Remote Sensing, 52*, 1687-1698

Li, K., Tan, M., Xiao, D., Yu, T., Li, Y., & Li, J, 2024: Research on road extraction from high-resolution remote sensing images based on improved UNet++. *IEEE Access, 12*, 50300-50309

Li, R., Zheng, S., Duan, C., Wang, L., & Zhang, C, 2022: Land cover classification from remote sensing images based on multi-scale fully convolutional network. *Geo-spatial information science, 25*, 278-294

Liu, P., Choo, K.-K.R., Wang, L., & Huang, F, 2017: SVM or deep learning? A comparative study on remote sensing image classification. *Soft Computing, 21*, 7053-7065

Rissati, J., Molina, P., & Anjos, C, 2020: Hyperspectral image classification using random forest and deep learning algorithms. In, *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)* (pp. 132-132): IEEE

Ronneberger, O., Fischer, P., & Brox, T, 2015: U-net: Convolutional networks for biomedical image segmentation. In, *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241): Springer

Ulmas, P., & Liiv, I, 2020: Segmentation of satellite imagery using u-net models for land cover classification. *arXiv preprint arXiv:2003.02899*

Zang, N., Cao, Y., Wang, Y., Huang, B., Zhang, L., & Mathiopoulos, P.T, 2021: Land-use mapping for high-spatial resolution remote sensing image via deep learning: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 5372-5391

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-205-2026