Automated Vegetation Segmentation in Mountainous Mediterranean Terrain Using U-Net and UAV Imagery: A Case Study in Dir El Ksiba, Morocco

Jennaoui, I.,[1*] Amadou, M. M. S.,[1] Bachaoui, E. M.,[1] El Ghmari, A.,[1] Biniz, M.,[2] El Harti, A.[1] and Badrouss, S.[1]

[1]Geomatics, Georesources and Environment Laboratory, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni Mellal, Morocco
E-mail: ilias.jennaoui@usms.ma,* ORCID ID: https://orcid.org/0009-0005-6564-7053*
saniamadou.mahamadoumansour@usms.ac.ma, ORCID ID: https://orcid.org/0009-0006-9045-7830
m.BACHAOUI@usms.ma, ORCID ID: https://orcid.org/0000-0003-4163-6307
a.elghmari@usms.ma, ORCID ID: https://orcid.org/0000-0002-2773-6048
a.elharti@usms.ma, ORCID ID: https://orcid.org/0000-0003-3976-4588
Sara.badrouss@usms.ma, ORCID ID: https://orcid.org/0009-0000-2675-4810

[2]Information Processing and Decision Support Laboratory (TIAD), Multidisciplinary Faculty of Beni Mellal, Sultan Moulay Slimane University, Beni Mellal, Morocco
E-mail: mohamedbiniz@gmail.com, ORCID ID: https://orcid.org/0000-0002-9448-6165
**Corresponding Author*

DOI: https://doi.org/10.52939/ijg.v22i5.4991

Abstract

Accurate vegetation mapping in mountainous Mediterranean terrain presents significant challenges due to topographical complexity and accessibility constraints. This study implements a U-Net architecture with MobileNetV2 encoder for automated vegetation segmentation using high-resolution UAV imagery in Dir El Ksiba, Morocco. The methodology leveraged 34 strategically selected images from challenging mountainous terrain, generating 9,996 standardized patches for optimized model training. Our approach targeted five vegetation classes: three phenological stages of Quercus ilex, Pinus halepensis, and Tetraclinis articulata. The U-Net+MobileNetV2 model demonstrated robust performance with validation metrics outperforming training metrics: 82.0% accuracy and mean Intersection over Union (mIoU) of 0.585. Class-specific IoU values revealed highest performance for Pinus halepensis (0.634), followed by Quercus ilex Stage 1 (0.598), Tetraclinis articulata (0.588), Quercus ilex Stage 2 (0.571), and Quercus ilex Stage 3 (0.556). The lightweight architecture (3.5M parameters) enables efficient deployment while maintaining competitive performance. This research establishes a methodological framework for applying deep learning-based vegetation segmentation in topographically complex Mediterranean ecosystems, addressing practical constraints of mountainous terrain data acquisition while maintaining scientific rigor.

Keywords: Deep learning, MobileNetV2, U-Net, UAV Remote Sensing, Vegetation Segmentation

1. Introduction

Mediterranean mountain ecosystems present significant challenges for vegetation mapping due to terrain complexity and accessibility constraints [1]. These ecosystems harbor exceptional plant biodiversity concentration, creating urgent needs for detailed vegetation segmentation to support conservation planning and territorial management [2]. However, conventional segmentation methods struggle with the spatial resolution, temporal efficiency, and cost-effectiveness required for comprehensive vegetation analysis in rugged mountainous terrain [3]. Traditional methods face limitations in mountainous terrain due to accessibility and resolution constraints [4]. Recent studies have highlighted the dynamic nature of vegetation cover in these mountainous environments, emphasizing the need for systematic monitoring approaches to understand temporal changes in vegetation distribution patterns [5]. Traditional ground-based botanical surveys, while scientifically robust, cannot provide the spatial coverage and systematic documentation required for

comprehensive territorial analysis [6]. Satellite imagery lacks the spatial resolution necessary for detailed vegetation segmentation in topographically complex terrain, creating significant gaps in knowledge of Mediterranean mountain ecosystems [7]. UAV technology enables systematic documentation of inaccessible areas with centimeter-level resolution [8] and [9], while deep learning methodologies, particularly CNNs, provide automated classification capabilities at unprecedented scales [10] and [11].

The U-Net architecture has emerged as a leading approach for vegetation segmentation applications due to its encoder-decoder structure with skip connections that enable precise boundary localization essential for accurate classification [12]. Recent applications in vegetation segmentation have demonstrated the effectiveness of U-Net architectures, with implementations achieving remarkable performance in complex terrain analysis. The viability of U-Net architectures with ResNet50 backbones for segmenting aromatic plant species in Morocco's High Atlas Mountains was established, achieving mIoU values of 73.99% and pixel accuracy of 91.75% [13].

This previous research focused on relatively accessible terrain, whereas our study addresses the significantly more challenging topographical conditions of Dir El Ksiba's rugged mountainous environment. This prior investigation demonstrated the viability of combining UAV imagery with deep learning for automated vegetation segmentation and highlighted the importance of proper data augmentation and loss function selection for achieving optimal segmentation performance. However, segmentation applications in mountainous terrain present unique challenges requiring architectural considerations that balance classification accuracy with computational efficiency for practical workflows [14]. MobileNet architectures, utilizing depthwise separable convolutions, offer compelling advantages for segmentation applications by providing computational efficiency while maintaining feature extraction capability, making them particularly suitable for operational scenarios with resource constraints [15]. This study addresses critical gaps in segmentation methodology for Mediterranean mountainous terrain by evaluating the application of U-Net architecture with MobileNetV2 encoder for high-altitude UAV-based vegetation segmentation. The approach focuses on the challenging terrain of Dir El Ksiba in Morocco's Atlas Mountains, where conventional segmentation approaches are impractical due to topographical constraints and accessibility limitations. From extensive high-altitude UAV flights conducted under optimal conditions, we systematically selected 34 images representing the highest quality documentation achievable in this challenging mountainous environment. Importantly, vegetation segmentation applications differ fundamentally from other computer vision tasks in that each image contains hundreds to thousands of vegetation objects, enabling robust model training through systematic patch extraction from relatively few high-quality images. This characteristic distinguishes vegetation segmentation from object detection applications and validates the methodological approach for comprehensive ecosystem documentation using carefully selected imagery [16].

The primary research objectives are to evaluate the performance of U-Net+MobileNetV2 for vegetation classification in mountainous terrain, demonstrate the effectiveness of high-altitude UAV imagery for documenting vegetation in previously inaccessible terrain, assess the capability of deep learning models to distinguish between phenological stages within species using high-resolution imagery, quantify segmentation accuracy across different vegetation classes in challenging topographical conditions, and establish methodological approaches for automated vegetation segmentation in mountainous Mediterranean environments.

2. Study Area and Data Acquisition

2.1 Geographic Setting

The study was conducted in Dir El Ksiba, a mountainous commune located in the Béni Mellal-Khénifra region of central Morocco (32°30′N, 5°52′W) as shown in Figure 1. This area represents a transitional zone between the Middle Atlas Mountains and the Tadla Plain, with elevations ranging from 1,379 to 1,463 meters above sea level within the study area. The region is characterized by a Mediterranean climate with hot, dry summers and mild, wet winters, receiving an annual precipitation of approximately 357.5 mm with high interannual variability (Köppen climate classification: Csa). The topographical complexity of the study site exemplifies the challenges faced in vegetation segmentation, featuring steep slopes, varied exposures, and diverse microhabitats that create conditions representative of Mediterranean mountain ecosystems where advanced segmentation techniques are essential [17]. The detailed elevation profile (Figure 2) illustrates the pronounced terrain variations and identifies the UAV takeoff location used for aerial documentation.

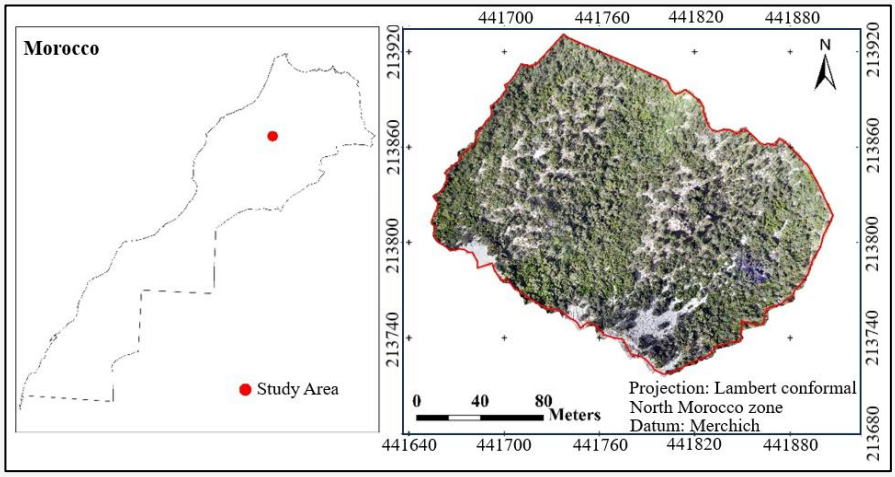

Figure 1: Study area map of Dir El Ksiba

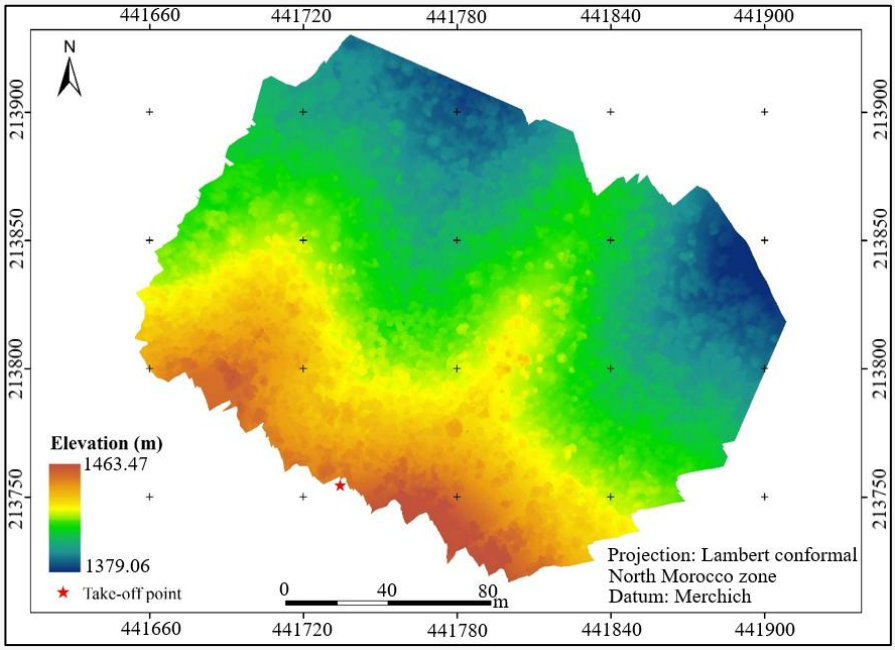

Figure 2: Digital elevation model of the study area

2.2. Target Vegetation Species

Five vegetation classes were identified for segmentation (Figure 3). Quercus ilex (Holm Oak) was subdivided into three phenological stages: Stage 1 (dark green) representing full vegetative vigor (Figure 3(a)), Stage 2 (light green) indicating moderate canopy density (Figure 3(b)), and Stage 3 (yellowish) representing environmental stress or senescence (Figure 3(c)). The phenological subdivision enables automated monitoring of vegetation stress responses detectable in RGB imagery [18] and [19]. Pinus halepensis (Aleppo Pine) was included as a drought-adapted coniferous species with distinctive needle morphology and branching patterns (Figure 3d). Tetraclinis articulata (Barbary Thuja) was selected as an endemic North African coniferous species with distinctive spectral signatures for conservation monitoring (Figure 3(e)).

2.3 UAV Data Acquisition

Aerial imagery was acquired using a DJI Phantom 4 RTK equipped with a 1-inch CMOS sensor providing 20-megapixel RGB images and integrated Real-Time Kinematic (RTK) positioning system for centimeter-level georeferencing accuracy. The UAV flights were conducted on April 28, 2025, under optimal weather conditions (clear skies, wind speed <10 km/h) during mid-morning hours to minimize shadows and ensure consistent illumination across the challenging mountainous terrain.

Flight parameters were specifically optimized for high-altitude mountainous terrain documentation, including altitude of 25 meters above the takeoff point, forward overlap of 80%, side overlap of 70%, and nadir camera orientation. These parameters ensured sufficient image overlap for photogrammetric processing while maintaining the spatial resolution necessary for individual plant identification in topographically complex terrain [8]. The RTK system provided positional accuracy within ±3 cm horizontally and ±5 cm vertically, eliminating the need for ground control points in inaccessible areas.

2.4 Dataset Characteristics

From high-altitude flight campaigns, 34 high-resolution images were systematically selected representing optimal quality documentation as shown in Table 1. This selective approach reflects practical constraints of challenging terrain including weather conditions, accessibility limitations, and strict quality requirements. Vegetation imagery exhibits extraordinary object density (519 objects per image average), enabling robust deep learning training from 34 images containing 17,644 vegetation objects. Combined with systematic 256×256 pixel patch extraction, this generated 9,996 training patches providing comprehensive spatial and spectral information for model training. Each image (5472×3648 pixels) provides 1.5 cm ground sampling distance optimal for individual plant identification. The 256×256 pixel patches represent 3.84×3.84 meters of ground coverage, capturing individual plant canopies and neighborhood context while accommodating scale variations in vegetation patterns.

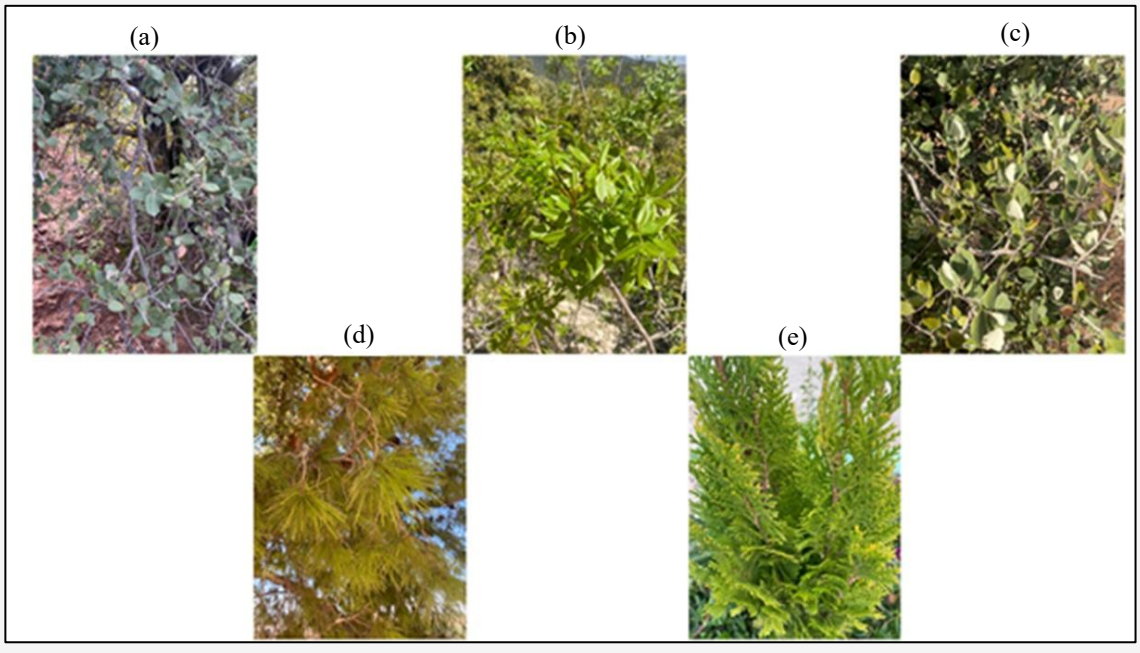

Figure 3: Target vegetation classes: (a) Quercus ilex Stage 1 (dark green), (b) Quercus ilex Stage 2 (light green), (c) Quercus ilex Stage 3 (yellowish), (d) Pinus halepensis, and (e) Tetraclinis articulata

Table 1: Dataset characteristics and spatial information density

Component	Count	Percentage (%)	Description
Selected High-Quality Images	34	100	Optimal documentation from mountainous terrain
Total Annotated Objects	17,644	N.A.	Dense vegetation object representation
Average Objects per Image	519.5	N.A.	Demonstrating spatial information density
Total Patches Generated	9,996	100	256×256 pixel segmentation units
Training Patches	6,997	70	Model training dataset
Validation Patches	1,499	15	Model validation dataset
Ground Sampling Distance	1.5 cm	N.A.	Spatial resolution achievement
Average Vegetation Coverage	50.24%	N.A.	Per image vegetation representation
Terrain Coverage	1.63 hectares	N.A.	Documented mountainous area

3. Methodology

3.1 Data Preprocessing and Patch Extraction

Raw UAV imagery underwent systematic preprocessing to address the unique challenges associated with high-altitude documentation in topographically complex environments. Initial processing included radiometric calibration to standardize pixel values across varying illumination conditions typical of mountainous terrain, and geometric correction using the integrated RTK positioning data to ensure spatial accuracy despite terrain-induced distortions. Images were systematically divided into 256×256 pixel patches using a grid-based partitioning approach with no overlap to maximize the number of independent training samples while avoiding spatial autocorrelation issues common in complex terrain documentation. The 256×256 pixel patch size was selected based on computational efficiency for batch processing, adequate spatial context for vegetation pattern recognition in complex topography [20], and compatibility with deep learning architectures adapted for these applications. However, patch-based training introduces a fundamental scale mismatch between training and full-resolution UAV inference, a critical deployment challenge in high-resolution vegetation mapping [21].

3.2 Expert Annotation Protocol

Ground truth annotations were created using Supervisely platform [22] by expert botanists with Mediterranean mountain vegetation knowledge. Quality control included cross-validation between multiple experts and field verification where accessible. The annotation workflow (Figure 4) shows the transition from UAV imagery to semantic segmentation masks. Each polygon was converted to pixel-level masks with numeric identifiers: Quercus ilex stages (1, 2, 3), Tetraclinis articulata (4), Pinus halepensis (5), and background (0).

3.3 Overall Methodology Workflow

The complete methodology workflow for the U-Net+MobileNetV2 vegetation segmentation approach demonstrates the systematic process from raw UAV imagery acquisition to final segmentation output. The workflow encompasses data preprocessing, expert annotation, model training, and evaluation phases. The methodology begins with raw UAV images captured at 5472×3648 pixel resolution, which undergo systematic splitting to create training (70%), validation (15%), and test datasets (15%). A parallel test set using full images is maintained for comprehensive model evaluation under real-world conditions. The training and validation datasets are further processed through image partitioning to generate 256×256 pixel patches, optimizing computational efficiency while preserving spatial context. The U-Net+MobileNetV2 architecture processes these standardized patches through systematic training and validation cycles, ultimately producing automated segmentation capabilities for five distinct vegetation classes in challenging mountainous terrain conditions. Table 2 presents vegetation object density analysis across all five classes. Quercus ilex Stage 2 represents the most abundant class with 7,087 objects across 33 images (214.76 objects per image, 28.42% coverage). The dataset contains 17,644 total objects across 34 images, averaging 519.5 vegetation objects per image.

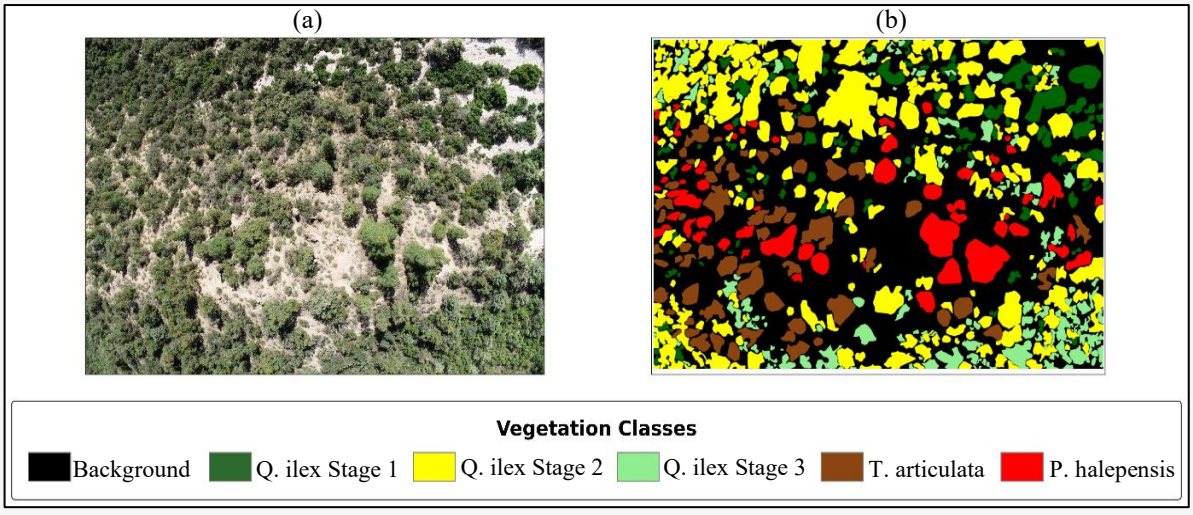

Figure 4: Annotation workflow: (a) original UAV imagery and (b) ground truth segmentation masks

Table 2: Vegetation object density analysis from supervisely platform

Vegetation Class	Images	Total Objects	Objects per Image (avg)	Coverage per Image (%)
Quercus ilex Stage 2	33	7,087	214.76	28.42
Quercus ilex Stage 3	33	4,638	140.55	7.13
Quercus ilex Stage 1	33	3,140	95.15	5.97
Tetraclinis articulata (thuya)	33	1,794	54.36	5.33
Pinus halepensis (pin d'Alep)	32	985	30.78	3.39
Total	**34**	**17,644**	**519.5**	**50.24**

3.4 U-Net Architecture with MobileNetV2 Encoder

The U-Net architecture incorporates MobileNetV2 as the encoder backbone, utilizing depthwise separable convolutions for computational efficiency in field deployment scenarios. MobileNetV2 was selected for its reduced computational complexity (8-9 times lower than standard convolutions), minimal memory requirements (3.5M vs 98M parameters for ResNet50), and faster inference speed (3-4 times faster than ResNet50) while maintaining competitive accuracy [23] and [15]. The architecture demonstrates proven effectiveness for vegetation segmentation tasks with operational advantages [14]. To evaluate the architectural advantages of the proposed MobileNetV2 encoder, two additional architectures were trained under identical experimental conditions: a U-Net with ResNet34 encoder and a DeepLabV3+ with ResNet50 backbone. All three models were trained on the same dataset split (70/15/15), with identical augmentation strategies, batch size, learning rate, and number of epochs, ensuring fair comparison. The model configuration utilized pre-trained ImageNet weights, three RGB input channels, six output classes (five vegetation classes plus background), and 256×256 pixel input dimensions to optimize batch processing while preserving spatial context for vegetation boundary delineation.

3.5 Training Configuration and Optimization

The U-Net+MobileNetV2 model was trained with batch size 32, learning rate 0.001 (Adam optimizer), and 100 epochs using CrossEntropyLoss function for multi-class segmentation. Data augmentation included geometric transformations (horizontal/vertical flipping, random rotation ±30°, elastic deformation), photometric enhancements (brightness/contrast adjustment, Gaussian noise, color jittering), and grid distortion (probability 0.2-0.5). Training was conducted on Intel Core i9-14900K with NVIDIA RTX 4060 GPU, requiring approximately 6.2 hours for convergence.

3.6 Evaluation Metrics

Model performance was assessed using multiple complementary metrics providing comprehensive evaluation of segmentation quality specifically adapted for mountainous terrain challenges. Mean Intersection over Union (mIoU) serves as the primary metric for mountainous terrain segmentation evaluation, calculating intersection over union for each class and averaging across all classes while accounting for the spatial complexity inherent in topographically challenging environments. The mIoU metric as calculated using Equation 1:

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i}$$

Equation 1

Where N is the number of classes, TP_i represents true positives, FP_i represents false positives, and FN_i represents false negatives for class i. This metric provides comprehensive assessment of overlap accuracy essential for vegetation boundary delineation in complex topographical conditions. Pixel Accuracy measures the percentage of correctly classified pixels across all classes, providing overall assessment of classification performance under varied terrain conditions. Pixel accuracy as calculated using Equation 2:

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}\left(TP_i + FP_i + FN_i\right)}$$

Equation 2

Precision for each class i as calculated using Equation 3:

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$

Equation 3

F1-Score provides balanced assessment of precision and recall, as calculated using Equation 4:

$$F1\text{-}Score_i = \frac{2 \times Precisoin_i \times Recall_i}{Precision_i + Recall_i}$$

Equation 4

Where Recall for class *i* is defined as:

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

Equation 5

All metrics are computed per class and averaged across classes to provide comprehensive evaluation of model performance.

4. Results

4.1 Training Performance Analysis

The training analysis of the U-Net+MobileNetV2 architecture (Figure 5) reveals stable performance characteristics optimized for vegetation segmentation applications. The training dynamics demonstrate optimal convergence patterns with good generalization capability, as evidenced by validation metrics maintaining consistent performance throughout the training process without significant overfitting or instability. The loss evolution analysis demonstrates stable training characteristics specifically adapted for the complexity of mountainous terrain vegetation segmentation.

The U-Net+MobileNetV2 achieved steady training dynamics with training loss decreasing from approximately 1.5 to 0.52 over 100 epochs, while validation loss stabilized around 0.61, indicating adequate generalization capability essential for challenging terrain applications. The loss curves reveal rapid initial improvement within the first 20 epochs, followed by steady convergence throughout the remaining training period, indicating effective adaptation to the complexity characteristics of mountainous terrain vegetation patterns. Notably, the validation loss consistently tracked training loss without significant divergence, demonstrating adequate generalization capability essential for practical applications in challenging terrain environments where model robustness is critical for operational deployment.

4.2 Performance Metrics Evolution

The performance analysis illustrated in (Figure 6) reveals consistent characteristics across primary evaluation metrics optimized for mountainous terrain applications. The U-Net+MobileNetV2 demonstrated steady performance evolution with adequate generalization capability across accuracy and mIoU metrics throughout the training process. The performance metrics analysis demonstrates consistent characteristics optimized for complex topographical conditions.

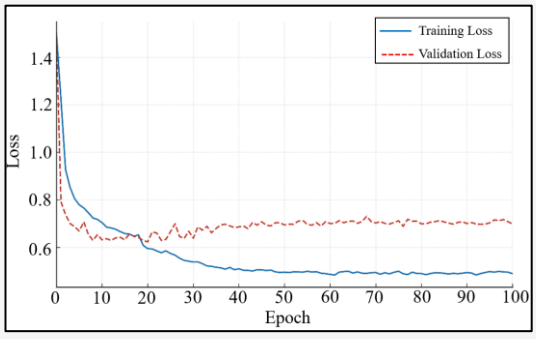

Figure 5: Training and validation loss evolution over 100 epochs

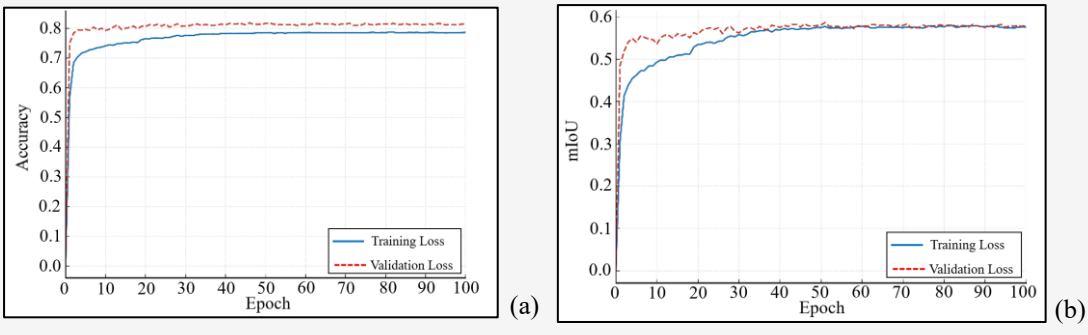

Figure 6: Performance evolution: (a) accuracy and (b) mIoU progression over 100 epochs

Table 3: Comprehensive performance metrics for enhanced architecture in mountainous terrain

Performance Category	Training	Validation	Performance Advantage
Pixel Accuracy (%)	78.3	82.0	+3.7%
mIoU	0.568	0.585	+0.017
Precision	0.803	0.841	+0.038
F1-Score	0.793	0.825	+0.032

The architecture achieved training accuracy of approximately 78.3% and validation accuracy of 82.0%, demonstrating the notable characteristic where validation performance exceeds training performance by 3.7 percentage points, indicating good architectural design adapted for mountainous terrain complexity. The mIoU progression demonstrates effective segmentation performance with the architecture achieving training mIoU of approximately 0.568 and validation mIoU of 0.585, representing adequate segmentation quality with the notable pattern of validation performance exceeding training performance. This pattern validates the architecture's capability to generalize beyond training data while maintaining optimal segmentation accuracy under challenging terrain conditions. The precision analysis demonstrates comprehensive performance in accurate positive identification of vegetation classes with the architecture achieving training precision of approximately 80.3% with validation precision reaching 84.1%, while the F1-score analysis provides comprehensive evaluation of the architecture's balanced performance with the model achieving training F1-score of approximately 79.3% and validation F1-score of 82.5%. This consistent performance across all metrics validates the architecture's balance between precision and comprehensive detection capability while maintaining good generalization capability essential for practical deployment in diverse mountainous terrain environments.

4.3 Comprehensive Performance Summary

Comprehensive evaluation of the U-Net+MobileNetV2 architecture as presented in Table 3 reveals consistent performance characteristics. The model demonstrates adequate performance across all evaluation metrics while providing computational efficiency essential for practical deployment in challenging terrain environments. The validation performance across all metrics demonstrates the architecture's ability to generalize beyond training data while maintaining adequate performance. This performance characteristic validates the architectural implementation for challenging terrain applications and establishes reasonable benchmarks for vegetation segmentation in topographically complex environments.

4.4 Per-Class Performance Analysis

The individual vegetation class performance analysis reveals varying segmentation accuracy across species and phenological stages under topographical constraints. Table 4 presents detailed per-class metrics, providing insights into classification difficulty and species-specific model performance. Pinus halepensis achieved the highest segmentation accuracy (IoU: 0.634), demonstrating the model's ability to leverage distinctive needle morphology and branching patterns for robust identification. The consistent performance hierarchy among Quercus ilex stages (Stage 1: 0.598, Stage 2: 0.571, Stage 3: 0.556) demonstrates the model's capability to detect subtle physiological variations. This phenological sensitivity represents a significant advancement for automated vegetation health monitoring, enabling assessment of vegetation stress patterns and seasonal dynamics critical for ecosystem monitoring in inaccessible terrain. Tetraclinis articulata achieved adequate performance (IoU: 0.588), demonstrating feasibility for automated monitoring of this endemic North African species and supporting conservation applications in previously inaccessible environments.

To complement the quantitative metrics, confusion matrices were generated for the U-Net+MobileNetV2 and U-Net+ResNet34 architectures on the test set (Figure 7). The normalized matrices reveal that Q.ilex Stage 3 exhibits the highest misclassification rate, with 0.15 confusion toward both Stage 1 and Stage 2, reflecting inherent spectral similarity between phenological stages under challenging mountainous terrain conditions. Tetraclinis articulata and Pinus halepensis achieve the strongest discrimination (0.91 and 0.79 respectively for MobileNetV2), leveraging their morphologically distinctive characteristics. Both architectures demonstrate consistent confusion patterns, confirming the reliability of misclassification trends across encoder configurations. Comparative analysis across three architectures (Table 5) reveals that while DeepLabV3+ achieves the highest validation mIoU (0.598), it requires 39.6M parameters more than eleven times the computational cost of the proposed architecture (3.5M). The U-Net+MobileNetV2 surpasses the ResNet34-based U-Net (24.4M

parameters) across all metrics while requiring 7 times fewer parameters.

4.5 Segmentation Results Visualization

The U-Net+MobileNetV2 architecture's segmentation capabilities are further demonstrated through comprehensive visualization across multiple representative examples from the mountainous terrain dataset. To validate the quantitative performance metrics, two distinct examples showcase the model's consistent performance across varying terrain complexity and vegetation density scenarios as illustrated in (Figure 8).

Table 4: Per-Class performance analysis for mountainous terrain vegetation

Vegetation Class	IoU	Precision	F1-Score
Pinus halepensis	0.634	0.842	0.794
Quercus ilex (Stage 1)	0.598	0.823	0.772
Tetraclinis articulata	0.588	0.811	0.765
Quercus ilex (Stage 2)	0.571	0.798	0.754
Quercus ilex (Stage 3)	0.556	0.784	0.745

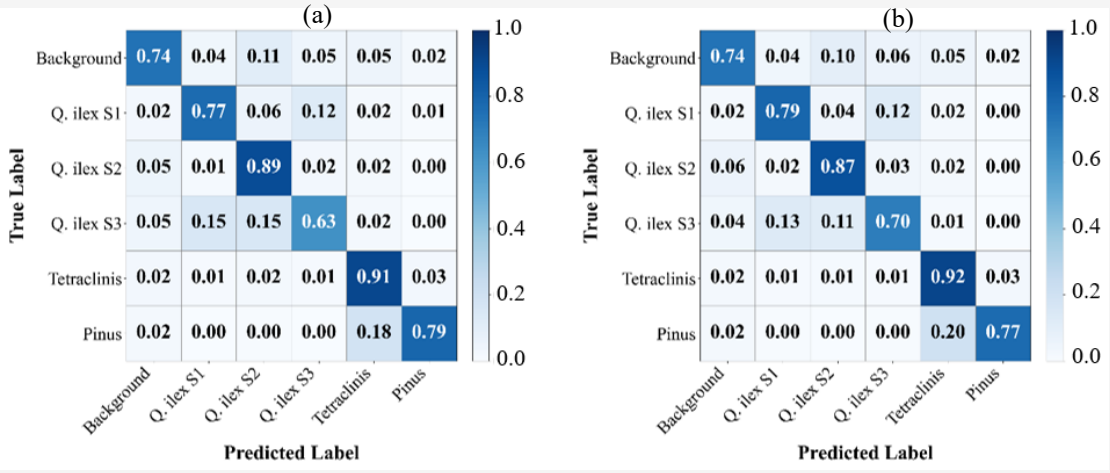

Figure 7: Normalized confusion matrices on the test set: (a) U-Net+MobileNetV2 and (b) U-Net+ResNet34

Table 5: Comparative performance of three segmentation architectures trained under identical experimental conditions

Architecture	Encoder	Val mIoU	Val Accuracy (%)	Val F1	Parameters
U-Net	ResNet34	0.568	79.2	0.799	24.4M
DeepLabV3+	ResNet50	0.598	83.3	0.831	39.6M
U-Net (ours)	MobileNetV2	0.585	82.0	0.825	3.5M

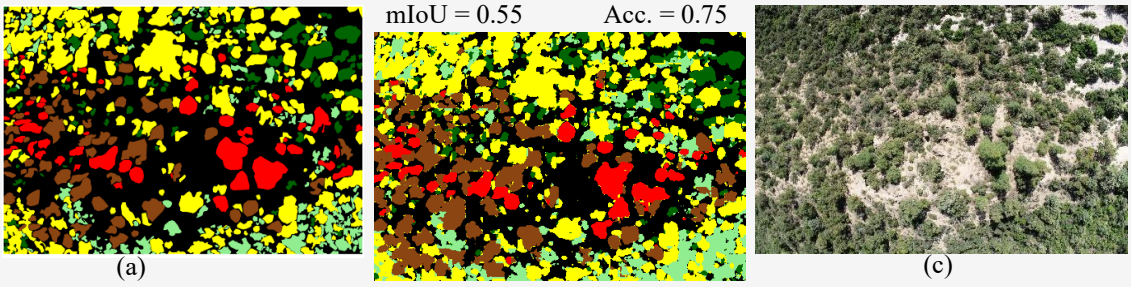

Figure 8: Full-image segmentation results showing ground truth, model predictions with performance metrics: (a) ground truth, (b) predictions with metrics, (c) original imagery

(continue next page)

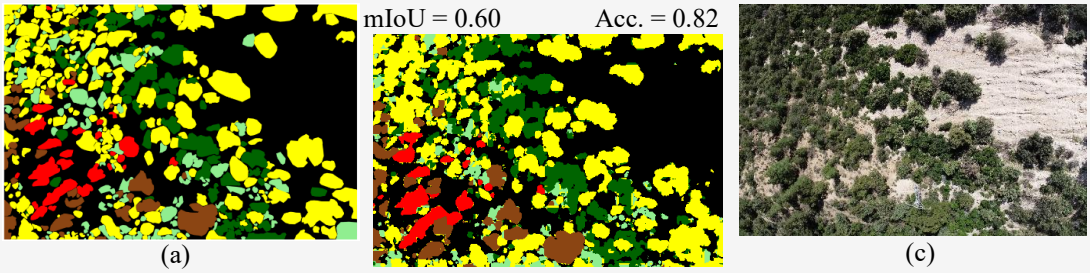

Figure 8: Full-image segmentation results showing ground truth, model predictions with performance metrics: (a) ground truth, (b) predictions with metrics, (c) original imagery
(Continue from previous page)

The representative segmentation results presented in (Figure 8) across three distinct mountainous terrain examples within the study zone demonstrate the architecture's consistent performance under varying topographical and vegetation density conditions. The individual performance metrics for each example (Example 1: mIoU 0.55, Accuracy 75%; Example 2: mIoU 0.60, Accuracy 82%) confirm adequate and consistent performance across varying terrain complexity levels. The U-Net+MobileNetV2 successfully maintains adequate boundary delineation and class discrimination across all examples, with the architecture demonstrating particular effectiveness in distinguishing between Quercus ilex phenological stages while robustly identifying distinctive species morphology. The visual comparison between ground truth annotations and model predictions validates the quantitative metrics and confirms the practical applicability of the architecture for detailed vegetation mapping in challenging mountainous environments. These results demonstrate that the patch-based training approach effectively generalizes to complete high-resolution imagery, maintaining spatial continuity and minimizing classification artifacts across diverse terrain conditions representative of Mediterranean mountainous ecosystems. The performance variation across examples (mIoU range: 0.52-0.60) reflects the natural complexity of mountainous terrain vegetation patterns, while the consistently adequate accuracy values (73-82%) validate the architecture's reliability for practical deployment in challenging topographical conditions.

5. Discussion

Our U-Net+MobileNetV2 demonstrates competitive performance in deep learning-based vegetation analysis. A previous study achieved 84% accuracy using U-Net for vegetation segmentation, showing that spatial patterns drive classification at high resolutions [24]. MIoU values of 82.47%, 73.44%, and 90.77% across three datasets were reported using MST-DeepLabv3+ with MobileNetV2, achieving parameter reduction (22.96MB vs our 3.5M) while maintaining high accuracy [25]. DeepLabv3+ with 512×512 pixel tiles performed optimally (mean F1-Score = 0.61) for tree species mapping in savanna ecosystems, noting challenges in temporal transferability [26]. These results compare with 73.99% mIoU achieved for aromatic plant species classification using U-Net+ResNet50 (98M parameters) at 50m flight altitude (GSD: 1.37cm) in relatively accessible terrain [13].

Our MobileNetV2 approach achieves competitive performance (58.5% mIoU) while operating at lower altitude (25m, GSD: 1.5cm) in significantly more challenging mountainous terrain and addressing the more complex task of phenological stage discrimination, with dramatically reduced computational requirements (3.5M parameters) suitable for resource-constrained deployments. Patch-based training introduces a fundamental scale mismatch between training and full-resolution UAV inference, a critical deployment challenge in high-resolution vegetation mapping [21].

The lightweight nature of the architecture makes it adaptable for potential future applications with Landsat data at regional scales, extending beyond the UAV-based approach demonstrated in this study, complementing local-scale UAV imagery studies. Our lightweight architecture (3.5M parameters) enables larger spatial deployments while maintaining accuracy comparable to parameter-intensive models. Future research could integrate our approach with fine-grained classification capabilities for targeted ecological monitoring [24] and [25]. Validation performance exceeding training performance (accuracy: 82.0% vs 78.3%, mIoU: 0.585 vs 0.568) demonstrates robust generalization and effective regularization through augmentation. This performance pattern is attributed to our intensive data augmentation strategy, which introduces geometric and photometric variations that make the training

task artificially more difficult. The validation set, evaluated without these augmentations, presents the model with cleaner, non-distorted images that are inherently easier to segment. This confirms the effectiveness of our augmentation approach as a regularization technique that enhances model generalization.

The successful differentiation of Quercus ilex phenological stages (IoU: 0.556-0.598) enables automated ecosystem health monitoring, with performance hierarchy (Stage 1: 0.598 > Stage 2: 0.571 > Stage 3: 0.556) reflecting sensitivity to vegetation stress patterns. Pinus halepensis achieved the highest accuracy (IoU: 0.634), leveraging distinctive morphological features for fire risk and drought monitoring applications. Tetraclinis articulata performance (IoU: 0.588) supports conservation efforts in challenging terrain. This study establishes that robust deep learning models can be trained through systematic patch extraction from high-quality images. The high vegetation object density (519.5 objects per image) enables effective training with fewer images, addressing mountainous terrain data acquisition constraints while maintaining scientific rigor.

6. Conclusion

This study demonstrates U-Net+MobileNetV2 effectiveness for vegetation segmentation in Mediterranean mountains using UAV imagery, achieving 82.0% accuracy and 0.585 mIoU with computational efficiency suitable for field deployment. The successful differentiation of Quercus ilex phenological stages and identification of Pinus halepensis (IoU: 0.634) and Tetraclinis articulata (IoU: 0.588) validates automated ecosystem monitoring in inaccessible terrain. The research establishes that robust deep learning models can be trained through systematic patch extraction from high-quality images, with vegetation imagery exhibiting extraordinary object density (519.5 objects per image) enabling effective training with fewer total images. However, this study has several limitations. Findings are geographically constrained to the Dir El Ksiba study area, and broader validation across diverse Mediterranean regions and seasonal variations is required before operational deployment. The dataset comprising 34 images from a single site limits spatial generalizability, and future work should incorporate cross-validation across multiple sites and multispectral imagery to improve phenological stage discrimination.

Acknowledgments

The authors would like to thank the Sultan Moulay Slimane University for providing resources and support for this research. Special thanks to the technical staff at the Geomatics, Georesources and Environment Laboratory for their assistance with data acquisition and processing.

References

[1] Shoshany, M., (2000). Satellite Remote Sensing of Natural Mediterranean Vegetation: A Review within an Ecological Context. *Progress in Physical Geography,* Vol. 24(2); 153-178. https://doi.org/10.1177/030913330002400201.

[2] Nieto Feliner, G., Cellinese, N., Crowl, A. A. and Frajman, B., (2023) Understanding Plant Diversity and Evolution in the Mediterranean Basin. *Frontiers in Plant Science,* Vol. 14. https://doi.org/10.3389/fpls.2023.1152340.

[3] Gartzia, M., Alados, C. L., Pérez-Cabello, F. and Bueno, C. G., (2013). Improving the Accuracy of Vegetation Classifications in Mountainous Areas. *Mountain Research and Development,* Vol. 33(1); 63-74. https://doi.org/10.1659/MRD-JOURNAL-D-12-00011.1.

[4] Elgadi, S., Zine, H., Dallahi, Y. and Ouhammou, A., (2024). Unveiling Floristic Diversity in the High Atlas: Insights from a Protected Reserve in a Global Mediterranean Biodiversity Hotspot. *Biosystems Diversity,* Vol. 32(4); 416-425. https://doi.org/10.15421/012445.

[5] Aboubaker, H., Raji, M., Belhadj-Khedija, C., Saurer, M., Khabba, S. and Er-Raki, S., (2023) Vegetation Cover Dynamics in the High Atlas Mountains of Morocco. *Remote Sensing,* Vol. 15(5). https://doi.org/10.3390/rs15051366.

[6] Anderson, K. and Gaston, K. J., (2013). Lightweight Unmanned Aerial Vehicles Will Revolutionize Spatial Ecology. *Frontiers in Ecology and the Environment,* Vol. 11(3); 138-146. https://doi.org/10.1890/120150.

[7] Lechner, A. M., Foody, G. M. and Boyd, D. S., (2020). Applications in Remote Sensing to Forest Ecology and Management. *One Earth,* Vol. 2(5); 405-412. https://doi.org/10.1016/j.oneear.2020.05.001.

[8] Thammaboribal, P., Tripathi, N., and Lipiloet, S. (2026). Comparative Study of UAV Mapping with and without Ground Control Points (GCPs). *International Journal of Geoinformatics*, Vol. 22(3), 36–50. https://doi.org/10.52939/ijg.v22i3.4861.

[9] Colomina, I. and Molina, P., (2014). Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 92; 79-97. https://doi.org/10.1016/j.isprsjprs.2014.02.013.

[10] Christin, S., Hervet, É. and Lecomte, N., (2019). Applications for Deep Learning in Ecology. *Methods in Ecology and Evolution,* Vol. 10(10); 1632-1644. https://doi.org/10.1111/2041-210X.13256.

[11] Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y. and Zhang, L., (2020) Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sensing of Environment,* Vol. 241. https://doi.org/10.1016/j.rse.2020.111716.

[12] Ronneberger, O., Fischer, P. and Brox, T., (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention,* Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.

[13] Badrouss, S., Daiaeddine, M. J., Bachaoui, E. M., Biniz, M., Mouncif, H., Ghmari, A. E., Harti, A. E. and Boulli, A., (2025). Modeling Plant Species Segmentation Using an Advanced U-Net and UAV Remote Sensing: A Case Study in the High Atlas Mountains of Morocco. *Modeling Earth Systems and Environment,* Vol. 11(1). https://doi.org/10.1007/s40808-024-02222-w.

[14] Kattenborn, T., Leitloff, J., Schiefer, F. and Hinz, S., (2021). Review on Convolutional Neural Networks (CNN) In Vegetation Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 173; 24-49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

[15] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. and Adam, H., (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint*. https://doi.org/10.48550/arXiv.1704.04861.

[16] Stoian, A., Poulain, V., Inglada, J., Poughon, V. and Derksen, D., (2019). Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sensing,* Vol. 11(17). https://doi.org/10.3390/rs11171986.

[17] El Afi, M., Alaoui, M., Hilali, A., Mosaid, H. and Barakat, A., (2023). Contribution to the Study of Landslides in Dir-El Ksiba (Atlas of Beni Mellal, Morocco) Using the Geological, Geotechnical, and Geochemical Approaches. *Journal of Sedimentary Environments,* Vol. 8(4); 605-615. https://doi.org/10.1007/s43217-023-00150-w.

[18] Intarat, K., Tuphimai, N., and Jangsawang, W. (2026). Integrating Smoothing Techniques with Convolutional Neural Networks for Rice Cropping Systems Classification in Suphan Buri, Thailand. *International Journal of Geoinformatics*, Vol. 22(5), 72–88. https://doi.org/10.52939/ijg.v22i5.4981.

[19] Peñuelas, J., Rutishauser, T. and Filella, I., (2013). Phenology Feedbacks on Climate Change. *Science,* Vol. 324(5929); 887-888. https://doi.org/10.1126/science.1173004.

[20] Hamwood, J., Alonso-Caneiro, D., Read, S. A., Vincent, S. J. and Collins, M. J., (2018). Effect of Patch Size and Network Architecture on a Convolutional Neural Network Approach for Automatic Segmentation of OCT Retinal Layers. *Biomedical Optics Express,* Vol. 9(7); 3049-3066. https://doi.org/10.1364/BOE.9.003049.

[21] Jennaoui, I., Bachaoui, E.M., Biniz, M., El Harti, A. and El Ghmari, A., (2026). Smart Integration of Sliding Window and Vote-Based Fusion: Advancing UAV-Based Instance Segmentation with YOLOv8 for High-Resolution Vegetation Mapping. *Remote Sensing Applications: Society and Environment,* Vol. 42. https://doi.org/10.1016/j.rsase.2026.101994.

[22] Supervisely, (2025). Supervisely: Web-Based Platform for Computer Vision Annotation and Machine Learning. [Online]. Available: https://supervisely.com [Accessed: May. 15, 2025]

[23] Guo, Y., Liu, Y., Georgiou, T. and Lew, M. S., (2022). A Review of Semantic Segmentation Using Deep Neural Networks. *International Journal of Multimedia Information Retrieval,* Vol. 7(2); 87-93. https://doi.org/10.1007/s13735-017-0141-z.

[24] Kattenborn, T., Eichel, J. and Fassnacht, F. E., (2019). Convolutional Neural Networks Enable Efficient, Accurate and Fine-Grained Segmentation of Plant Species and Communities from High-Resolution UAV Imagery. *Scientific Reports,* Vol. 9. https://doi.org/10.1038/s41598-019-53797-9.

[25] Wang, Y., Yang, L., Liu, X. and Yan, P., (2024). An Improved Semantic Segmentation Algorithm for High-Resolution Remote Sensing Images Based on DeepLabv3+. *Scientific Reports,* Vol. 14. https://doi.org/10.1038/s41598-024-60375-1.

[26] Popp, M. R. and Kalwij, J. M., (2023). Consumer-Grade UAV Imagery Facilitates Semantic Segmentation of Species-Rich Savanna Tree Layers. *Scientific Reports,* Vol. 13. https://doi.org/10.1038/s41598-023-40989-7.