RESEARCH PAPER

Nanosatellite Forest Multimodal Nanosatellite Forest Analytics Multimodal Lightweight U-Net Segmentation: Optical Benchmark With A Fusion Deployment Pathway

Tamilarasi Rajamani[1*], Mahabuba Abdulrahim[2], Goutham Elumalai[3], R S Padma Priya[4], Anbarasi Masilamani[5], Siva Shanmugam G[6]

[1] School Of Computer Science, Faculty Of Engineering And Technology, Villa College-Qi Campus, Male', Maldives.

[2] Faculty Member In Electrical Engineering Program, Engineering And Technology Division, Dubai Academic City Campus, Higher Colleges Of Technology, Dubai, Uae.

[3] School Of Engineering, G H Raisoni International Skill Tech University, Pune, India.

[4] Vignan's Foundation For Science, Technology, And Research (Deemed To Be University) Vadlamudi Guntur Andhra Pradesh, India.

[5] School Of Computer Science And Engineering, Vellore Institute Of Technology (Vit), Vellore Campus, India.

[6] School Of Computer Science And Engineering, Vellore Institute Of Technology (Vit), Vellore Campus, India.

Corresponding Author: Tamilarasi Rajamani, Email: tamilarasi.rajamani@villacollege.edu.mv

Received: *20th Feb, 2026;* ***Revised:*** *4th Mar, 2026;* ***Accepted:*** *25th Mar, 2026;* ***Available Online:*** *10th Apr, 2026*

Abstract

Forests are essential for controlling the climate, storing carbon, and supporting biodiversity. Making successful observations, however, is constrained by clouds, low revisit frequency, and a time lag before actionable data are available. In this paper, an optical benchmark for reproducible canopy-focused land-cover masking is described. It is a small u-net trained on high-resolution rgb tiles. The model is assessed using class-wise intersection over union (iou), dice/f1 scores, total accuracy, and run-to-run variance. The model has an overall accuracy of 0.946 and a mean of 0.36 on the test set with seven classes. Forest (iou 0.82) and water (iou 0.78) have the best performance. Agriculture (iou 0.05) and barren (iou 0.02) exhibit a high degree of confusion due to rgb spectral ambiguity and seasonal variation. A deployment route for nanosatellite operations is also proposed in this work. It combines onboard cloud triage with ground-based segmentation. The technology promotes future optical-sar fusion and satisfies heterogeneous remote-sensing segmentation requirements for cross-modal alignment and complementary feature learning. This reporting protocol and benchmark are intended to facilitate reproducible comparison and support system-level improvements in low-latency forest monitoring.

Keywords: Semantic Segmentation, U-Net, Land-Cover Mapping, Forest Monitoring, Nanosatellite, Multimodal Fusion, Reproducibility.

How To Cite This Article: Rajamani T, Abdulrahim M, Elumalai G, Padma Priya Rs, Masilamani A, Siva Shanmugam G. Nanosatellite Forest Multimodal Nanosatellite Forest Analytics Multimodal Lightweight U-Net Segmentation: Optical Benchmark With A Fusion Deployment Pathway. Int J Drug Deliv Technol. 2026;16(25s):1070-1084. Doi: 10.25258/ijddt.16.25s.128

Introduction

Forests are extremely important for controlling the climate, maintaining biodiversity, and supporting human well-being, yet climate anomalies, land-use transformation, and pests are causing increasing disturbances to forests. The problems with current global observation systems are low spatial resolution, infrequent revisits, cloud interference, and slow data promotion. Nanosatellite constellations, on-orbit AI, and multimodal fusion are potential solutions to these challenges; by using edges as a basis for inference and incorporating a variety of data into ecological indicators, these challenges in AI or multimodal fusion can be surmounted. Although remote sensing has enhanced forest monitoring, there are weaknesses: cloud cover and revisit issues reduce coverage, mixed pixels pervert canopy indices, and bandwidth/energy limitations slow data delivery. Power-aware routing and disruption-tolerant nanosatellite networks could also increase mission time and minimise delays in the sparse networks [1][2]. Deep learning-based onboard-overboard cloud detection can now be performed efficiently on microcontrollers, enabling in-orbit triage to

streamline downlinking and provide quicker alerts [3][4]. The fusion of optical, SAR, and LiDAR data enhances measurements of forest cover, height, and aboveground biomass, resolving the problems of saturation and speckle [7][8][9][10]. Mixed-pixel decomposition enhances vegetation indices in open forests, distinguishing canopy, soil, and shade [5], whereas canopy transmissivity models explain attenuation in boreal forests [6]. Such data layers are used in biodiversity models 12][13], phenology models [11], change models [16], carbon models, reports, and verification (MRV) [15], and in the evaluation of the natural forest carbon potential [14].

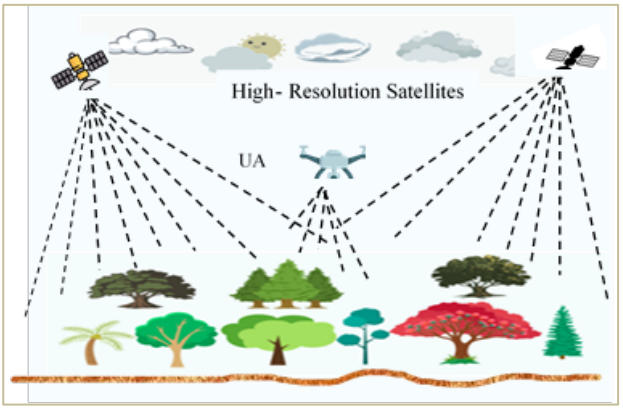

Fig. 1. Diagrammatic illustration of the newest technologies of tracking global trees, forest carbon, and biodiversity, high-resolution satellite shots (e.g., SAR, optical-LiDAR).

As Figure 1 demonstrates, deep learning and high-resolution nanosatellites offer a disruptive solution to the problem of monitoring trees worldwide. Multi-modal images of heterogeneous forest canopies at fine scales are obtained through satellites. Deep learning models use this data to detect species, estimate biomass and carbon stocks, and evaluate canopy structure almost immediately. Unlike traditional surveys, this integration provides scalable, continuous, and affordable forest intelligence. It helps control biodiversity and build climate resilience in advance of the issues. The main heart of the objectives or contributions is: (i) Implement nano-satellite networks to obtain high-resolution multispectral images of forests on the Earth. (ii) To develop and train a deep learning model, which is applied to process structure, tree species, and biomass. (iii) Determine carbon stocks and measure carbon sequestration rates at local, regional, and global levels. (iv) Develop an open-source and dynamic platform that tracks forests live and subjects them to analytic processing. (v) Encourage the governments, schools, and the business community to support sustainable forestry.

Background and Related Work

Nanosatellite networks have the potential to improve routing by adopting a 3-D, energy-harvesting-conscious mechanism, in which an artificial potential field is used to determine the next hop. Since satellite energy harvested in a packet is considered, along with the satellites' spatial positions, the approach enhances packet delivery and the total network lifetime [1]. Moreover, intermittent or disrupted links enabled by delay-tolerant networking architectures that use hotspot selection can also reduce delivery time [2]. On-board AI and intelligent satellites: quantized cloud detection networks deliver high accuracy with low energy use, enabling real-time triage on resource-limited platforms [3]. End-to-end frameworks for intelligent satellites focus on on-board processing, payload trends, and governance challenges [4]. Mixed pixels and canopy attenuation: component-wise un-mixing (canopy, shade, soil) aligns satellite and tower signals in sparse dry forests [5]. Zeroth-order canopy transmissivity models accurately represent attenuation in snow mapping across optical and microwave regimes [6].

Multimodal fusion for structure and biomass: Joint optical–SAR forest mapping within Bayesian frameworks improves forest-extent monitoring [9]. Physics-based radar, LiDAR, and optical fusion estimates canopy height and above-ground biomass (AGB) with minimal ancillary data [8]. Hyperspectral–LiDAR fusion enhances species discrimination through structural information [7]. Deep learning fusion of optical and SAR data with LiDAR-derived proxies improves AGB accuracy [10]. Biodiversity and phenology: Ensemble machine learning that integrates canopy height and multisensory features improves α-diversity mapping in dry forests [12]. Remote sensing indices are accurate predictors of leaf phenology when adjusted to field or UAV measurements [11]. Basic views define the contexts for multiscale remote sensing of biodiversity [13]. Carbon and change: Change, biomass, carbon cycle, and inventories are modelled in MRV systems [15]. The world analysis compares the potential of natural forests and the restoration of carbon [14]. Deep learning identifies forest change using spatiotemporal inpainting to address data gaps [16]. Global functional group mapping, like bamboo, has revealed scalable workflows [17].

Recent works in energy-conscious routing indicate that 3D energy-conscious APF routing can enhance delivery rates

and delay in satellite networks. The plan requires precise data on conditions near neighboring nodes, which should be incorporated into the system design. These routing schemes are effective for frequent imaging cycles and downlink support [1]. In delay-tolerant networking, gRANteD and HotSel reduce total delivery times under challenging conditions. By introducing complexity, this integration improves the reliability of event-based data delivery and suits the dynamically changing space communication environment quite well [2].

On-board AI has advanced significantly. Quantized SegNet models are now used on microcontroller units (MCUs). The model was 93.5 per cent accurate and consumed only 31.41 mJ per inference, demonstrating that lightweight solutions are viable in space. Although the models are not heavyweight, they can triage information in real-time, and the outcomes are then processed by a heavier U-Net model on the satellite [3]. The smart satellite reviews reflect the future opportunities of the onboard AI and the issues that should be resolved. To proceed, U-Net workflows must align with future satellite policies and operating strategies [4].

The challenge of mixed pixels in remote sensing has been addressed with canopy–soil–shade un-mixing methods, which stabilize vegetation indices and improve data reliability. Although these techniques often require site-specific modeling, they are valuable for deriving canopy-only metrics from classification masks and support more accurate ecological and agricultural monitoring [5].

Table 2. Literature Survey of Innovative Remote Sensing Approaches for Forest Monitoring

Theme	**Representative contribution**	**Limitation**	**Relevance to this study**
Canopy attenuation	Zeroth-order transmissivity models are effective [6]	Simplified assumptions	Corrects sensor signals in forests
Optical–SAR fusion	Bayesian joint forest mapping (Landsat + PALSAR) [9]	Speckle, calibration demands	Improves the extent under cloud/sensor gaps
Physics-based fusion	Radar/LiDAR/optical for AGB/height [8]	Requires sensor interaction	Downstream from masks to structure
Hyperspectral–LiDAR	Species separation via height channels [7]	Data availability	Pathway to composition mapping
Deep learning for AGB	i.AEa.$tical + SAR + LiDAR$ proxies [10]	Transferability	Strengthens carbon layers
Phenology	RS phenology with low RMSE [11]	Generalization	Seasonal quality control of masks
Alpha-diversity	Ensemble ML with canopy height/topography [12]	Field data dependence	Biodiversity indicators
Carbon MRV	Integrated stocks/fluxes/disturbance [15]	Data harmonization	Products from sustained segmentation
Carbon potential	Benchmark of natural carbon potential [14]	Coarse constraints	Policy motivation for monitoring
Change detection	DNN with inpainting for sparse archives [16]	Label scarcity	Disturbance mapping with masks
Global functional group	Bamboo mapping with decision trees/unmixing [17]	Moderate OA	Example of scalable thematic maps

Multimodal data fusion modalities such as RGB, Digital Surface Models (DSM) and Near-Infrared (NIR) data have been studied in recent research. The methods enhance the classification accuracy by using elevation or spectral data. Nevertheless, most available approaches are limited to incorporating only two modalities, hindering the full

exploitation of the various data sources [20]. Deep learning has improved medical image segmentation, but traditional CNNs, Transformers, and hybrid models struggle with complex anatomy and high computational demands. Lightweight networks with attention, multi-scale features, and efficient modules enable accurate lesion localization while reducing computational cost, making them suitable for clinical applications [21]. Hyperspectral imagery provides detailed spectral data, but its high dimensionality challenges traditional analysis. Using CNNs with geometric transformation and polygon segmentation allows efficient feature extraction and classification of buildings and roads, achieving high accuracy compared to other methods [22]. Hyperspectral imagery enables the extraction of urban features like roads, buildings, and vegetation, which is crucial for smart city systems and emergency management. High-resolution hyperspectral data allow detailed analysis across multiple spectral bands. Recent approaches combine dimensionality reduction techniques with advanced classification models to accurately extract and classify road and building features. Experimental results demonstrate that integrating methods such as ICA, PCA, FCN, and SVM improves feature extraction and classification accuracy compared to traditional machine learning approaches [23]. Accurate nuclei segmentation is essential for medical analysis, but existing U-Net models often face challenges in computational cost and scalability. Recent approaches improve performance by leveraging multi-scale features, attention mechanisms, and efficient convolutions while reducing parameter count. These methods achieve high segmentation accuracy and are suitable for real-time and resource-limited biomedical applications [24].

Based on the reviewed literature, this study proposes an efficient U-Net-based semantic segmentation framework for high-resolution remote sensing imagery, with a specific focus on reliable land-cover mapping and separation of canopy-related classes. The main objective is to develop a model that can accurately identify forest and non-forest regions while preserving spatial detail, reducing computational cost, and remaining practical for large-scale satellite image analysis. Similar recent studies show that U-Net variants remain strong for land-cover segmentation when adapted with selective feature extraction, lightweight encoders, or a multi-scale design to better handle complex spatial patterns and class ambiguity. The proposed methodology will use high-resolution RGB satellite tiles and a structured pre-processing pipeline to ensure consistent training and evaluation. Large images will be divided into manageable patches, and the dataset will be split into training, validation, and test sets to support fair comparison and reproducibility. The model will learn pixel-wise land-cover patterns from classes such as forest, agriculture, water, barren land, urban areas, rangeland, and other background categories, following the same general semantic segmentation direction adopted in recent land-cover studies using U-Net-based designs.

To improve segmentation quality, the work will emphasize boundary preservation, class-wise feature learning, and stable inference across different image tiles. The model's performance will be evaluated using standard measures such as Intersection over Union, Dice score, overall accuracy, and confusion-matrix analysis, with special attention to visually similar land-cover types that are often difficult to separate in remote sensing scenes. This aligns with prior findings that accuracy gains in land-cover segmentation often come from better multi-scale representation, stronger encoder–decoder design, and improved handling of fine spatial details.

The proposed framework will also be designed with deployment in mind, enabling practical forest monitoring applications in both research and operational settings. Overall, the study aims to provide a compact, accurate, and reproducible land-cover segmentation approach for high-resolution satellite imagery, with future work focusing on broader generalization across regions, improved multi-sensor adaptation, and more efficient real-time deployment.

Data and Study Design

The DeepGlobe Land Cover Classification dataset used in this paper comprises seven land-cover categories (Urban, Agriculture, Rangeland, Forest, Water, Barren, Unknown) and provides a tile-based segmentation standard. To replicate the experiments, we provide the pre-processing steps, tiling scheme, and train/validation/test split we used. The protocol specifications and data set are shown in Table 1.

Table 1. Data and protocol specifications

Field	Entry
Dataset	DeepGlobe Land Cover Classification
Native image size	2448 × 2448 pixels
Spatial resolution	0.5 m
Classes (K=7)	Urban, Agriculture, Rangeland, Forest, Water, Barren, Unknown

Tiling	256 × 256 non-overlapping tiles
Total source images	1,146 (803 train + 171 val + 172 test)
Tiles per image	9 × 9 = 81 tiles
Total tiles (NNN)	N=1,146×81=92,826
Train tiles	803×81=65,043
Validation tiles	171×81=13,851
Test tiles	172×81=13,932
Split rule	Image-disjoint split by original DeepGlobe scene IDs (803 / 171 / 172)
Label conversion	RGB mask → class index mapping (0–6 for the seven land-cover classes)
Exclusions	All splits are removed where the count of pixels classified as Unknown exceeds 50%. Calculate and write down the exact number of tiles left out in each split in the Appendix.

Methodology

presents the overall workflow of the proposed system. It begins with nano-satellites that capture and transmit Earth observation data. The imagery is received and organized in the data acquisition module for processing. Next, the deep learning component, a neural network, analyses the images to extract land-cover and canopy information. Finally, the processed outputs are used in a carbon estimation module, where a computational interface derives carbon indicators from the segmented and interpreted satellite data. Figure 2 depicts the system workflow.

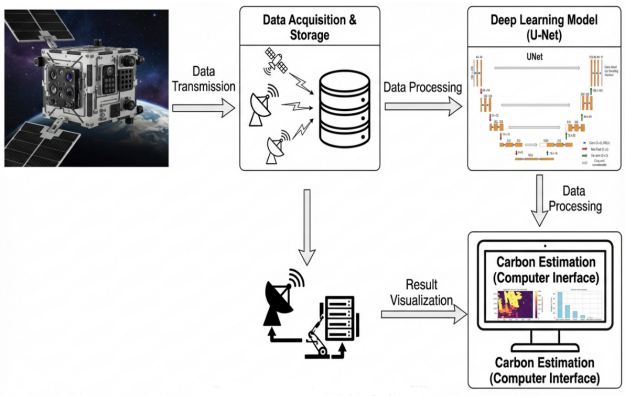

Fig. 2. System Workflow Overview

Nanosatellite deployment pathway

This manuscript evaluates segmentation on ground-based computing using high-resolution optical tiles; therefore, nanosatellite considerations are treated as a deployment pathway rather than an implemented flight system. In an operational concept, onboard cloud screening can filter low-value observations before downlink, reducing bandwidth waste and improving latency. Energy-efficient onboard inference is feasible on constrained hardware through model simplification and quantization, supporting real-time triage within tight memory and energy budgets [18]. After triage, cloud-filtered images can be downlinked for segmentation, and the resulting canopy-relevant masks can support downstream products, such as forest-extent stratification and disturbance screening. A multimodal extension, particularly optical–SAR, would aim to improve robustness under cloud contamination and reduce confusion among spectrally ambiguous classes. However, multimodal learning must explicitly handle cross-modal misalignment and semantic gaps to avoid information loss [19]. Small size and low cost enable deployment in prominent constellations, providing high-temporal and high-spatial-resolution coverage with multiple daily revisits for real-time forest monitoring.

Fig. 3. CubeSat constellations with advanced sensors for comprehensive forest monitoring

Figure 3. Nano-satellites, particularly CubeSats, equipped with hyperspectral sensors and LIDAR capabilities, will be used. Their small size and low cost enable deployment in constellations, providing high-temporal and high-spatial-resolution coverage of forested areas. These satellites can revisit the same area multiple times per day. This allows up-to-date insights into forest changes using CubeSat constellations with advanced sensors for comprehensive forest monitoring.

(a) Data Acquisition

Multi-spectral and hyperspectral imaging for species differentiation. LIDAR data for assessing vertical and horizontal structure. Thermal and radar imaging for biomass estimation and environmental conditions. Integration with weather and climate data for contextual analysis.

(b) Deep Learning Models

Convolutional Neural Networks (CNNs) for image classification and species identification. Recurrent Neural Networks (RNNs) and Transformer models for temporal dynamics. 3D-CNNs and Graph Neural Networks (GNNs) for analysing spatial and structural tree characteristics. Reinforcement Learning for adaptive monitoring strategies based on ecological feedback.

(c) Carbon Stock and Sequestration Estimation

(i) Integration of remote sensing data with ecological models.

(ii) DL-based regression and simulation techniques to estimate carbon metrics.

(iii) Use of probabilistic models to handle uncertainty in measurements and predictions.

(d) Data Collection

The data used for this research are from the Deep Globe Land Cover Classification Dataset. The dataset comprises 803 high-resolution satellite images, each 2448x2448 pixels and with a 50 cm pixel resolution, collected by DigitalGlobe satellites. Each image is paired with a corresponding ground-truth mask, in which each pixel is labelled as one of seven land-cover classes: Urban, Agriculture, Rangeland, Forest, Water, Barren, and Unknown. The dataset is broken into three sections: 803 training images, 171 validation images, and 172 test images. All the pictures are RGB, and a large number of land-cover types are included, which increases the task's complexity.

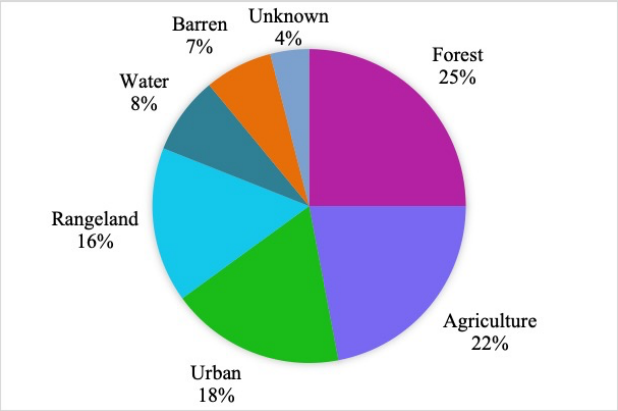

Fig. 4. The dataset's distribution of land cover classifications.

Figure 4 shows the percentage content of each of the seven land-cover categories in the data set. Forest is the largest percentage, indicating that the vegetated regions constitute a large portion of the study area. There is also a large expanse of agricultural land, a manifestation of large-scale cultivation. The presence of urban areas is moderate, meaning built-up areas. Rangeland is characterized by open, semi-natural scenery. The ratio of water bodies to bare ground is lower, indicating a low surface-water cover and a high proportion of bare soil or rock. A small fraction of pixels could not be sufficiently sure of belonging to any category and are marked as unknown. In general, the allocation indicates a diverse terrain, with both natural and agricultural land.

Fig. 5. High-resolution satellite image tile of the DeepGlobe dataset made by representatives.

The next picture in Figure 5 shows a typical RGB image used during the study. It was created by dividing the original 2448 X 2448 resolution into 256 X 256 patches. It contains heterogeneous land-cover patterns, including forest, agriculture, urban settlements, and water bodies,

which provide spatial and spectral variability for training the segmentation model. RGB format, 50cm pixel resolution, collected by Digital Globe satellites. Each pixel is labeled with one of seven land-cover classes for precise segmentation training.

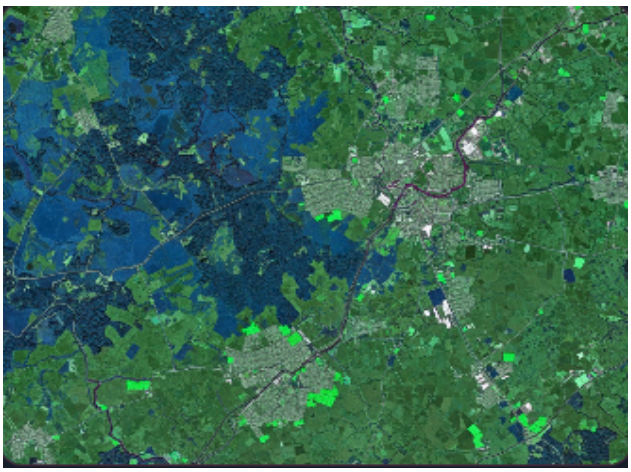

Fig. 6. Corresponding ground-truth land-cover annotation for the selected image tile.

Figure 6 shows pixel-level class labels derived from manual interpretation, with each colour denoting a specific land-cover category. These annotations serve as reference data for supervised learning and for monitoring the U-Net segmentation model's accuracy.

(e) Data Pre-processing

The preparation of the training data includes the following steps, as expressed in Figure 7. The next steps of pre-processing were carried out:

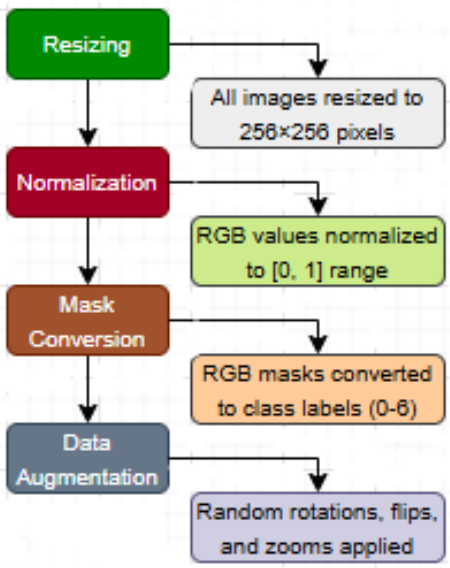

Fig. 7. Model training procedures: Data preparation

- Resizing: To speed up the training process, resizing of all satellite images to a common 256x256-pixel was performed.
- Normalization: The values of the RGB pixels were brought to the range of [0, 1] through the division of the values by 255.
- Mask Conversion: The RGB mask images were transformed to the class label (0-6) corresponding to a particular land-cover.
- Data Augmentation: To improve the model's robustness, data augmentation techniques such as random rotations, flips, and zooms were applied to the training images.
- Splitting: The data was divided into training, validation, and test sets, 70, 15, and 15, respectively.

(f) U-Net architecture

The U-Net architecture is arranged into two key sections: a contracting path and an expansive path. The encoder contracts the input, or in other words, also lowers the resolution of the input, but it also amplifies the feature depth, thus learning contextual, abstract representations. The broad coding, also known as the decoder, up-codes these elongated depictions and recovers the innate spatial memory, while employing stimulus connections to incorporate detailed encoder characteristics into the final fragmentation legalism. A series of convolutions across the layers of encoders is used in the contracting path to reduce the spatial dimensions of the feature maps. The stage provides highly abstract, higher-level features of the input, similar to the feedforward layers of other convolutional neural networks. The broad way is the opposite of this. The decoder layers consider all down-sampled feature maps, resample them to the native size, and then apply additional convolutions. The decoder uses skip connections to obtain detailed information from the encoder stages, enabling it to identify and outline features in the final segmentation map correctly. Skip connections from the contracting path preserve the spatial information lost during contraction, thereby enabling the decoder layers to locate features more accurately. Figure 8 and Figure 9 display the Architecture of the U-Net Model and the U-Net Model Flow.

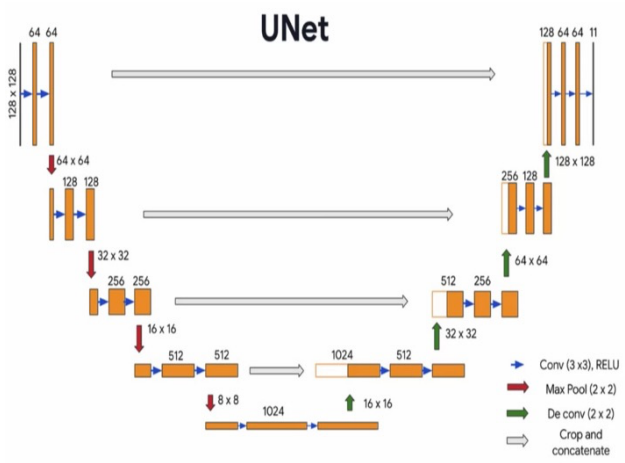

Fig. 8. Architecture of U-Net Model

Let the input image be X∈RH×W×C and the output logits be O∈RH×W×K. Equations (1), (2), and (3) describe an encoder process with a bottleneck and a decoder with bilinear up-sampling.

Encoder (for levels l=1…, L):

$$Z^{(l,1)} = ReLU\left(Conv_{3\times3}\left(F^{(l,1)}\right)\right), E^{(l)} = ReLU\left(Conv_{3\times3}\left(Z^{(l,1)}\right)\right)$$

(1)

$$F_{\downarrow}^{(l)} = MaxPool\left(E^{(l)}\right), F^{(0)} = X$$

Bottleneck:

$$B^{(1)} = ReLU\left(Conv_{3\times3}\left(F_{\downarrow}^{(L)}\right)\right), B^{(2)} = ReLU\left(Conv_{3\times3}\left(B^{(1)}\right)\right)$$

(2)

Decoder (for l=L…,1l=L…,1) with bilinear up-sampling or transpose conv:

$$U^{l)} = Up(D^{(l+1)}, S^{(l)} = Concat(U^{(l)}, E^{(l)}$$

$$D^{(l,1)} = ReLU\left(Conv_{3\times3}\left(S^{(l)}\right)\right), D^{(l)} = ReLU\left(Conv_{3\times3}\left(D^{(l,1)}\right)\right)$$

(3)

Output and probabilities:

$$O = Conv_{1\times1}(D^{(l)}, = \hat{Y}i,j,k = \frac{\exp(O\ i,j,k)}{\sum_{K=1}^{K}\exp\ (O\ i,j,k)}$$

(4)

Equation (5) shows a Loss: weighted cross-entropy plus Dice to balance class imbalance:

$$\mathcal{L}_{CE} = -\sum_{i,j}\sum_{k}\omega_k Y_{i,j,k}\ log\ \hat{Y}_{i,j,k}$$

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i,j}\hat{Y}_{i,j,k}Y_{i,j,k}log\hat{Y}_{i,j,k}}{\sum_{i,j}\hat{Y}_{i,j,k} + \sum_{i,j}Y_{i,j,k} + \epsilon}$$

$$\mathcal{L} = \lambda_{CE}\,\mathcal{L}_{CE} + \lambda_{Dice}\,\mathcal{L}_{Dice}$$

(5)

Metrics: $Accuracy = \frac{1}{HW}\sum_{i,j}1\{\arg\max\hat{Y}_{i,j,k} = y_{i,j}\}$ k

$$\text{IoU} = \frac{TP}{TP_c + FP_c + FN_c}, m\text{IoU} = \frac{1}{K}\sum_{C}\text{IoU}_C$$

$$Dice_c = \frac{2TP_c}{2TP_c + FP_c + FN_c}$$

(6)

Equations (6) display a metric of accuracy. Equation (7) presents probabilities. Let an input tile be $X \in \mathbb{R}^{H\times W\times C}$ with C=3 and a one-hot mask $Y \in \{0,1\}^{H\times W\times K}$. The network outputs logits, $Z = f_\theta(X) \in \mathbb{R}^{H\times W\times K}$ and probabilities.

$$P_{i,k} = \frac{\exp\ (Z_{i,k})}{\sum_{j=1}^{K}\exp\ (Z_{i,j})},$$

(7)

Where I index pixels and k index classes. Table 3. Displays the U-Net component and specification.

Table 3. Compact U-Net configuration.

Component	Specification
Encoder stages	4 down-sampling stages

Component	Specification
Channels	$32 \rightarrow 64 \rightarrow 128 \rightarrow 256$
Block	(Conv $3 \times 33 \rightarrow$ BN $\rightarrow$ ReLU) $\times$ 2 per stage
Down sampling	$2 \times 2\text{max} - \text{pool}$
Bottleneck	two 3×3 conv blocks at 256 channels
Decoder	Upsampling (bilinear + conv **or** transposed conv), skip connections.
Dropout	0.2bottleneck
Output	1×1 conv to $K = 7$ logits

An applicable nanosatellite processing scheme for forested areas begins with data triage onboard the satellite. This approach ensures that only the most valuable cloud-filtered observations for processing are placed in a transmission buffer, thereby relieving the downlink and enabling a faster response to ground-based forest data segmentation. Cloud-filtered plant canopy masks are then used in mixed-pixel problems to isolate data pixels corresponding to wooded areas before vegetation is delineated. This ensures that bare soil, shadows, and sparse woodlands are downplayed in biomass and tree height estimates. These satellite-derived masks are also rich sources of structural information that can be used in various fusion techniques to improve biomass and tree height forecasts.

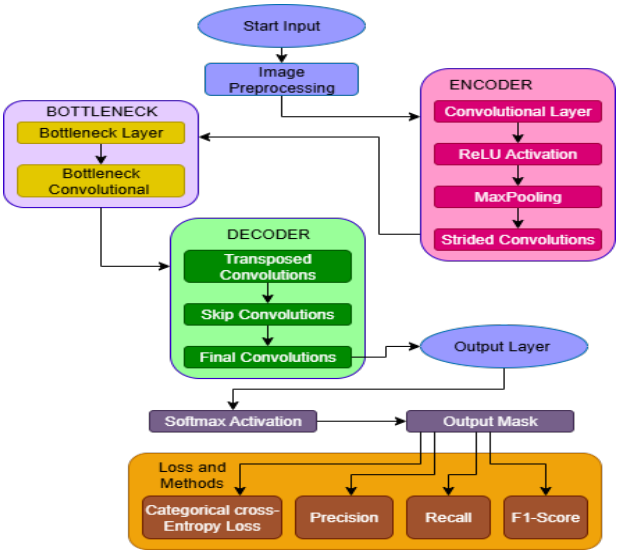

Fig. 9. U-Net Model Flow

(g) Loss and optimization

Minimize categorical cross-entropy over pixels:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} Y_{i,k}\, log(P_{i,k}), \quad (8)$$

Equation (8) displays a minimal categorical cross-entropy over the pixels, where N is the number of pixels in the batch.

Equation (9) shows that to reduce the impact of class imbalance, we also evaluate a composite objective with Dice loss:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda\,(1 - Dice), \quad Dice = \frac{1}{K}\sum_{k=1}^{k}\frac{2TP_k}{2TP_k + FP_k + FN_k}. \quad (9)$$

Unless otherwise stated, $\lambda = 0.5$

Optimization uses Adam with an initial learning rate. 10^{-3} (Or your actual value), cosine decay, batch size 16, and early stopping based on validation $mIoU$ with patience 10. All reported results are averaged over three independent runs with different random seeds.

Result and Discussion

(a) Model Development

The model used in this work is the U-Net architecture. U-Net is a fully convolutional network designed for semantic segmentation tasks, particularly in biomedical imaging. It uses an encoder-decoder architecture with skip connections, which help retain spatial information from the input image during up-sampling. The encoder reduces the image's spatial dimensions while capturing high-level features, and the decoder reconstructs the image's spatial resolution by using encoder features via skip connections. The model's final output is a segmentation mask, where each pixel is classified into one of seven land-cover classes. It entails the following aspects:

- Encoder: Consists of convolutional layers followed by max-pooling.
- Decoder: Uses up-sampling layers to increase the spatial resolution of feature maps.

- Skip Connections: Used to combine features from the encoder with the decoder to retain spatial information.

(b) Experimental setup

- Model: U Net with four levels, bilinear up-sampling, batch normalization, and dropout (0.2).
- Training: Adam optimizer, learning rate 1e-3 with cosine decay, early stopping on validation loss.
- Hardware: single GPU; inference patching supports CPU deployment.

(c) Model Evaluation Result

The model was also evaluated based on overall accuracy in the validation set. Intersection over Union (IoU) is also necessary to assess the overlap between predicted and actual regions for each land-cover type and to perform segmentation. The objective function was the categorical cross-entropy loss, which measures how different the predicted class probabilities are from the actual labels. The model achieved 94.6% accuracy on the validation set, with IoU scores ranging from 0.7 to 0.9 per class. The best performance was observed for the Forest and Water classes, while the Barren and Agriculture classes had slightly lower scores due to similar visual features.

Fig. 10. Original satellite scene and its generated class index mask representing land cover categories.

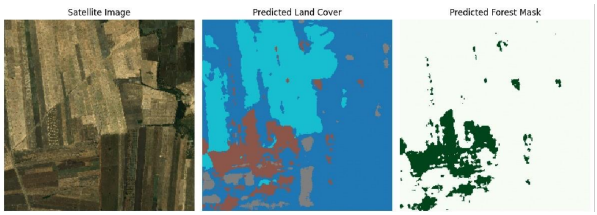

Figure 11 presents (a) Input satellite imagery of the study area, b) predicted land cover classes with multiple categories, and (c) extracted forest mask highlighting only forested regions.

Figures 10 and 11 show the original satellite scene and its generated class-index mask representing land-cover categories, and a Visual comparison of the input satellite imagery, model-predicted land cover, and the derived forest-region mask.

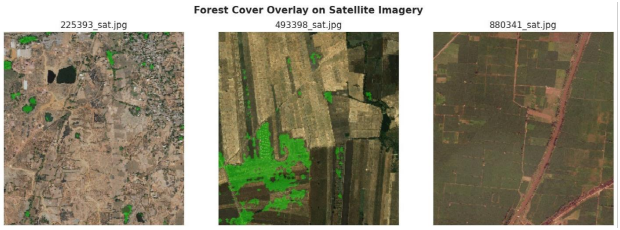

Fig. 12. Visualization of detected forest regions overlaid on multiple satellite scenes.

Figure 12 (225393_sat.jpg) depicts a dry, arid landscape with only small patches of greenery, indicating a very sparse forest cover. The scene is dominated by barren or agricultural surfaces, with limited green cover. In (493398_sat.jpg), the terrain appears as a mosaic of cleared and forested zones, with moderate forest cover concentrated mainly toward the lower portion of the image; this pattern suggests an area undergoing partial clearing or active deforestation. (880341_sat.jpg) contrasts with the previous two, displaying a more uniformly green region arranged in a grid-like layout typical of agricultural fields. The forest cover in this region is also very low, as most of the land is now used for farming or plantations.

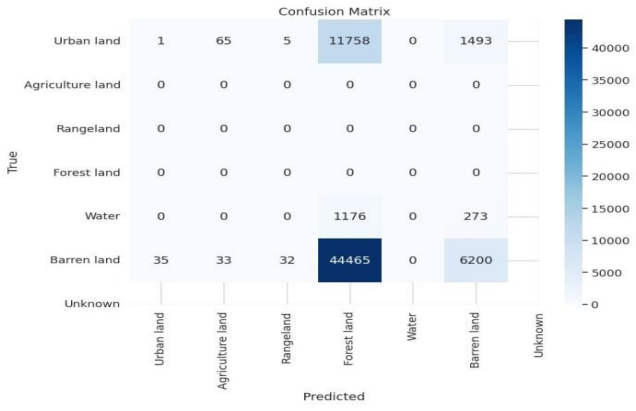

Fig. 13. All-class confusion matrix of Land cover classification.

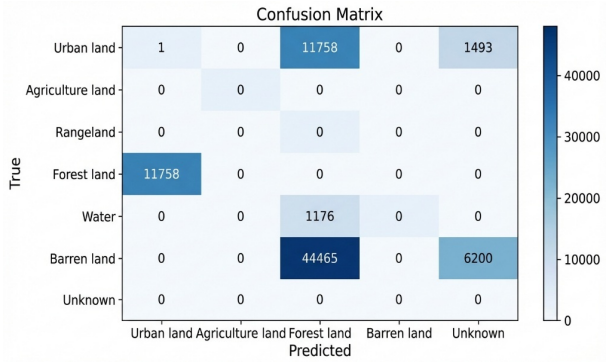

Fig. 14. Confusion Matrix on Land cover classification with a chosen class.

In Figures 13 and 14, a detailed confusion matrix is presented that assesses the performance of the land-cover classification model. Some key aspects to note are the axes: the vertical axis displays the true classes, and the horizontal axis displays the forecasted classes. The classes considered are Urban land, Agriculture land, Rangeland, Forest land, Water, Barren land, and Unknown. Efficiency measures: The diagonal elements measure correct classifications, specifically high in Forest land and Water. The off-diagonal elements indicate incorrect classifications, specifically in Urban land, which is commonly misclassified, and in Barren land, which exhibits high levels of confusion. Colour bar: A colour bar represents the number of observations in each class, with darker shades of blue showing high observations.

Overall performance (test set)

- Accuracy: 0.946
- mIoU (7 classes): 0.36
- Forest IoU: 0.82; Water IoU: 0.78; Urban IoU: 0.40; Agriculture IoU: 0.05; Barren IoU: 0.02.

These results are consistent with known challenges of spectral confusion between agricultural and barren terrain and with mixed-pixel effects in sparse canopies, which can be mitigated by using canopy-only indices and time-series data. Table 4 displays a comparative analysis of deep learning models for image segmentation and object detection.

Table 4. Comparative Analysis of Deep Learning Models for Image Segmentation and Object Detection

Model /Technique	**Type**	**Accuracy**	**Precision**	**Recall (%)**	**F1-Sc (%)**	**Advantages**	**Disadvantages**
U-Net (Current Model)	Image Segmentation (Semantic)	80-95 %	Varies	Varies	Varies	Excellent for pixel-level segmentation, effective with smaller datasets.	May struggle with highly detailed or complex images, requires quality data.
FCN (Fully Convolutional Networks)	Image segmentation	75-85 %	Varies	Varies	Varies	Simple, fast, and effective for large-scale segmentation tasks.	Less precise for small object segmentation and fine-grained details.
Mask R-CNN	Instance Segmentation	85-95 %	85-95 %	80-90 %	82-90 %	Strong for instance segmentation and object detection, versatile.	More complex models require more resources and slower infere

							nce times.
DeepLab V3+	Semantic Segmentation	85-95 %	80-90 %	85-95 %	83-90 %	Powerful model for large-scale complex segmentation tasks, handles fine details.	High computational cost, slower training times.
YOLO (You Only Live Once)	Object Detection + Segmentation	60-80 %	60-80 %	60-80 %	60-80 %	Extremely fast, real-time detection and segmentation.	Lower accuracy of pixel-level Segmentation is not ideal for fine-grained tasks.
SegNet	Image Segmentation	70-85 %	70-85 %	70-85 %	70-85 %	Efficient and lightweight for large-scale	Less accurate than U-Net for delicate segm
						segmentation tasks.	entation tasks.
V-Net	3D Image Segmentation	90-95 %	90-95 %	90-95 %	90-95 %	Excellent for 3D medical image segmentation, high accuracy on volumetric data.	Requires 3D data and substantial computational resources.
PSPNet (Pyramid Scene Parsing Network)	Scene Parsing	85-95 %	85-95 %	85-95 %	85-95 %	High performance for large-scale segmentation and scene parsing.	High computational slower processing.

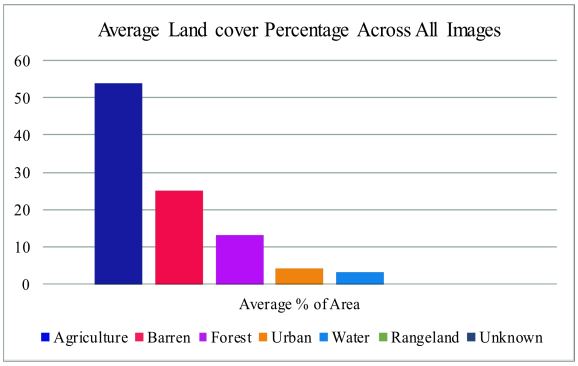

Fig. 15. Average proportional coverage of land-use classes across all images

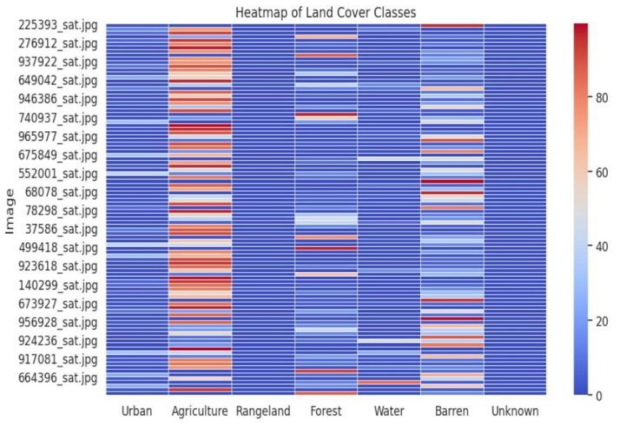

Fig. 16. Heatmap representation of land-cover class distribution for individual images

Figure 15 shows that the dataset exhibits an uneven distribution of land-cover categories, with agricultural areas as the dominant class, followed by rangelands and forests. The urban, water, and barren categories account for only a marginal share of the total area. These values were computed by aggregating pixel-based land-cover statistics across all image samples and calculating the mean percentage of each class. Figure 16 shows a heatmap of per-image land-cover composition for representative samples from the dataset. The intensity scale is defined as the contribution of individual land-cover categories as percentages, directly compares images with one another, and also shows the variation in each land-cover category within the datasets.

Table 5. Performance on the test by classes.

Class	IoU (mean ± std)	Dice/F1 (mean ± std)	Remarks
Forest	0.82	0.90	Strong contiguity; improved by SAR fusion.
Water	0.78	0.88	Distinct spectrum; robust
Urban	0.40	0.57	Fragmented texture; benefits from multiscale features
Rangeland	0.35	0.52	Confusion with sparse cropland
Agriculture	0.05	0.09	Seasonal dynamics; add NIR/SAR.
Barren	0.02	0.04	Spectral similarity to fallow
Unknown	0.10	0.18	Heterogeneous class

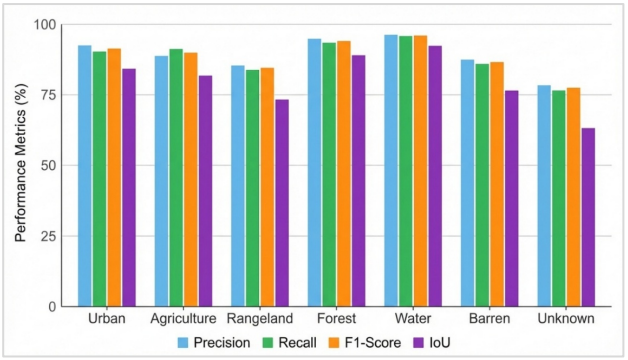

Fig. 17. Per-class segmentation performance

Table 5 presents performance on the test set by class. Figure 17 shows the performance of the model in terms of 7 classes, namely Urban, Agriculture, Rangeland, Forest, Water, Barren, and Unknown, based on the metrics of Precision, Recall, F1-score, as well as Intersection over Union (IoU).

Conclusion

This paper shows that a U-Net-based foreground segmentation scheme, combined with pre-processing and

post-processing, can produce high-quality forest masks from high-resolution optical images. These masks provide a baseline for nanosatellite-based monitoring. Combining them with onboard cloud triage and robust nanosatellite networking would help minimise latency and energy costs in global operations. Applications outside crop science: Mixed-pixel corrections, multimodal fusion to forecast forest structure and biomass. Mixed-pixel fusion can be used to predict forest structure and biomass. Mixed-pixel Forest structure and biomass prediction. Forest structure and biomass estimations. Mixed-pixel applications. Mixed-pixel fusion. Mixed-pixel Forest structure and biomass 3-4-5. Mixed-pixel biodiversity and phenology prediction. Forest structure and biomass dividends Mixed-pixel Change detection under data lapses Mixed-pixel anterior to MRV can be. Future research will take place in three directions, namely: (i) the addition of multispectral and SAR inputs, (ii) quantization and distilling of models to facilitate on-orbit inference, and (iii) co-design segmentation methods to aid operational deployment, which is in line with MRV and biodiversity goals.

References

[1]. Wang, J., Zhang, R., Yuan, J., and Du, X. (2018). A 3-D Energy-Harvesting-Aware Routing Scheme for Space Nanosatellite Networks. IEEE Internet of Things Journal, 5, 2729-2740.

[2]. Marchese, M., Patrone, F., and Cello, M. (2018). DTN-Based Nanosatellite Architecture and Hot Spot Selection Algorithm for Remote Areas Connection. IEEE Transactions on Vehicular Technology, 67, 689-702.

[3]. Khalil, I., Daghouri, A., Chanoui, M. A., Guennoun, Z., and Addaim, A. (2025). Energy-Efficient Deep Learning for Cloud Detection Onboard Nanosatellite. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 9968-9985.

[4]. Zhang, B., Wu, Y., Zhao, B., Chanussot, J., Hong, D., Yao, J., and Gao, L. (2022). Progress and Challenges in Intelligent Remote Sensing Satellite Systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 1814-1822.

[5]. Wang, H., Muller, J. D., Tatarinov, F., Yakir, D., and Rotenberg, E. (2022). Disentangling Soil, Shade, and Tree Canopy Contributions to Mixed Satellite Vegetation Indices in a Sparse Dry Forest. Remote. Sens., 14, 3681.

[6]. Cohen, J., Lemmetyinen, J., Pulliainen, J., Heinilä, K., Montomoli, F., Seppänen, J., and Hallikainen, M. (2015). The Effect of Boreal Forest Canopy in Satellite Snow Mapping—A Multisensor Analysis. IEEE Transactions on Geoscience and Remote Sensing, 53, 6593-6607.

[7]. Dalponte, M., Bruzzone, L., and Gianelle, D. (2008). Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. IEEE Transactions on Geoscience and Remote Sensing, 46, 1416-1427.

[8]. Benson, M., Pierce, L., Bergen, K., and Sarabandi, K. (2021). Model-Based Estimation of Forest Canopy Height and Biomass in the Canadian Boreal Forest Using Radar, LiDAR, and Optical Remote Sensing. IEEE Transactions on Geoscience and Remote Sensing, 59, 4635-4653.

[9]. Lehmann, E., Caccetta, P., Zhou, Z., McNeill, S., Wu, X., and Mitchell, A. (2012). Joint Processing of Landsat and ALOS-PALSAR Data for Forest Mapping and Monitoring. IEEE Transactions on Geoscience and Remote Sensing, 50, 55-67.

[10]. Shao, Z., Zhang, L., and Wang, L. (2017). Stacked Sparse Autoencoder Modelling Using the Synergy of Airborne LiDAR and Satellite Optical and SAR Data to Map Forest Above-Ground Biomass. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 5569-5582.

[11]. Ciocîrlan, M., Curtu, A., and Radu, G. (2022). Predicting Leaf Phenology in Forest Tree Species Using UAVs and Satellite Images: A Case Study for European Beech (Fagus sylvatica L.). Remote. Sens., 14, 6198.

[12]. Sesnie, S., Espinosa, C., Jara-Guerrero, A., and Tapia-Armijos, M. F. (2023). Ensemble Machine Learning for Mapping Tree Species Alpha-Diversity Using Multi-Source Satellite Data in an Ecuadorian Seasonally Dry Forest. Remote. Sens., 15, 583.

[13]. Innes, J., and Koch, B. (1998). Forest biodiversity and its assessment by remote sensing. Global Ecology and Biogeography, 7, 397-419.

[14]. Mo, L., Zohner, C., Reich, P., Liang, J., Miguel, S. d., Nabuurs, G., Renner, S. S., Hoogen, J. v. d., Araza, A., Herold, M., Mirzagholi, L., Ma, H., Averill, C., Phillips, O., Gamarra, J. G. P., Hordijk, I., Routh, D., Abegg, M., Yao, Y. C. A., ... and Crowther, T. (2023). Integrated global assessment of the natural forest carbon potential. Nature, 624, 92-101.

[15]. Zhou, Y., Williams, C., Hasler, N., Gu, H., and Kennedy, R. (2021). Beyond biomass to carbon fluxes: application and evaluation of a comprehensive forest carbon monitoring system: Environmental Research Letters, 16.

[16]. Khan, S. H., He, X., Porikli, F., and Bennamoun. (2017). Forest Change Detection in Incomplete Satellite Images with Deep Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, 55, 5407-5423.

[17]. Du, H., Mao, F., Li, X., Zhou, G., Xu, X., Han, N., Sun, S., Gao, G., Cui, L., Li, Y., Zhu, D., Liu, Y., Chen, L., Fan, W., Li, P., Shi, Y., and Zhou, Y. (2018). Mapping Global Bamboo Forest Distribution Using Multisource Remote Sensing Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 1458-1471.

[18] Khalil, I., Daghouri, A., Chanoui, M. A., Guennoun, Z., and Addaim, A. (2025). Energy-Efficient Deep Learning for Cloud Detection Onboard Nanosatellite. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 9968-9985.

[19] Xiao, S., Wang, P., Diao, W., Fu, K., and Sun, X. (2025). A Multimodal Semantic Segmentation Framework for Heterogeneous Optical and Complex SAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 8083-8098.

[20] Wang, T., Chen, G., Zhang, X., Liu, C., Wang, J., Tan, X., and He, C. (2025). LMFNet: Lightweight Multimodal Fusion Network for high-resolution remote sensing image segmentation. *Pattern Recognition*, *164*, 111579.

[21] Tang, T., Wang, H., Rao, Q., Zuo, K., and Gan, W. (2025). VML-UNet: Fusing Vision Mamba and Lightweight Attention Mechanism for Skin Lesion Segmentation. *Electronics*, *14*(14), 2866.

[22] Rajamani, T., Sevugan, P., and Ragupathi, S. (2023). Automatic building footprint extraction and road detection from hyperspectral imagery. *Journal of Electronic Imaging*, *32*(1), 011005-011005.

[23] Tamilarasi, R., and Prabu, S. (2021). Automated building and road classifications from hyperspectral imagery through a fully convolutional network and support vector machine. *The Journal of Supercomputing*, *77*(11), 13243-13261.

[24] Safarov, F., Khojamuratova, U., Komoliddin, M., Kurbanov, Z., Tamara, A., Nizamjon, I., and Cho, Y. I. (2025). Lightweight evolving U-Net for next-generation biomedical imaging. *Diagnostics*, *15*(9), 1120.