ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume 29, Number 3, 2026, 237–248
doi: 10.59277/ROMJIST.2026.3.03

A Spectral-Attention Deep Learning-Based Approach to Deforestation Monitoring in Satellite Imagery

Palanichamy NAVEEN[1] and Mahmoud HASSABALLAH[2,*]

[1]Department of Mathematics, University of Hradec Kralove, Czech Republic
[2]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, 16278 AlKharj, Saudi Arabia
Email: naveenpalanichamy@uhk.cz, mah.ali@psau.edu.sa*
* Corresponding author

Abstract. In automated AI systems for urban change detection, deforestation tracking requires effective, scalable approaches that generalize across different ecosystems. Current Normalized Difference Vegetation Index (NDVI)-based approaches and single-biome deep learning networks lack good generalizability across regions and are energy-intensive. NDVI is prone to saturation over closed vegetation and fails over dry canopy; biome-specific networks overfit and require extensive labeled datasets. These limitations hinder accurate monitoring in diverse environmental settings. To this end, a spectral-attention U-Net-based model is introduced using Short-Wave Infrared (SWIR) bands and multi-temporal Landsat 8/9/Sentinel-2 imagery for biome-flexible deforestation monitoring with transfer learned ResNet-50 and carbon-friendly training regimes. SWIR improves vegetation distress surveillance and is less susceptible to atmospheric perturbations; transfer learning reduces data demands and training energy with sustainability improvements. Also, Spectral Attention Gates (SAGs) are proposed that dynamically weight spectral inputs for improved feature extraction, and integrate a lightweight decoder design using transposed convolution for sharper boundary detection. The proposed model achieves 92.3% IoU (Intersection over Union) in the Amazon and 88.7% in Jordan's temperate forests with up to 22% higher precision in arid regions compared to the NDVI approach, while maintaining low energy costs ($1.43 \text{ kg } CO_2e/10 \text{ km}^2$ on NVIDIA A100). Hotspot analysis reveals 42.3 km^2/year loss in the Amazon (Z=4.71) and 12.5 km^2/year in Jordan (Z=3.28), validated against 5000 ground truth patches ($\kappa = 0.82$).

Key-words: Deep learning; deforestation detection; spectral attention; U-Net architecture; urban change detection.

1. Introduction

Urban expansion and deforestation are deeply interlinked, as growing cities often encroach upon natural landscapes, particularly forests [1]. Remote sensing technologies have become essential for tracking these changes, revealing how urban sprawl drives deforestation patterns [2]. Understanding this relationship is vital for sustainable land management and balancing development with ecological preservation. Similar challenges arise in other complex environmental systems, where nonlinear statistical models have proven effective in capturing interactions between physical processes and emission dynamics, such as in computational modeling of gas–aerosol emissions in nuclear power plants [3]. Globally, deforestation remains a severe environmental threat, contributing 12–20% of carbon emissions [4], accelerating biodiversity loss, and disrupting hydrological systems [5]. The FAO reports that approximately 10 million hectares of forest are lost annually due to agriculture, logging, and infrastructure expansion [6]. While tropical regions like the Amazon dominate global forest loss (nearly 60%), temperate regions such as the Mediterranean and Middle East also face increasing fragmentation due to urbanization and climate-related stressors. Existing deforestation monitoring systems face two fundamental limitations: (i) NDVI-based approaches saturate in closed-canopy forests and fail in arid ecosystems where vegetation stress mimics deforestation signals; (ii) most deep learning architectures are trained on single biomes, causing poor generalizability when applied to regions with different spectral and structural characteristics

Monitoring deforestation remains challenging due to ecological and technical constraints [7]. Tropical forests often degrade in fragmented patterns, while temperate regions exhibit linear deforestation linked to infrastructure development [8]. Persistent cloud cover and outdated monitoring systems further obscure early-stage degradation [9], highlighting the need for accurate observation-process modeling under uncertainty. Nonlinear cognitive-system frameworks emphasize that incomplete and noisy sensing critically affects learning outcomes [10]. Recent advances in adaptive and semi-supervised learning, including online ensemble fuzzy systems, demonstrate robustness to missing and streaming data [11]. More generally, nonlinear statistical models have successfully captured complex interactions in large-scale physical systems [3], while improved clustering strategies [12] and data-driven deep learning models under noisy observations [13] further support the effectiveness of adaptive intelligence. Recent surveys on evolving fuzzy and neuro-fuzzy systems underscore the importance of explainable and computationally efficient models for nonstationary environments [14], reinforcing their relevance to scalable deforestation monitoring across diverse biomes.

Traditional techniques, including NDVI and other vegetation indices [15], offer simplicity but struggle in arid regions due to seasonal variability [16]. Machine learning models like Random Forests and SVMs [17] improved classification but lacked spatial coherence, often producing noisy outputs [18]. While object-based approaches added spatial context, they remained region-specific and underperformed in multi-temporal analysis [19]. The introduction of deep learning, particularly CNNs like U-Net-transformed deforestation monitoring by enabling accurate pixel-wise segmentation with multi-sensor and temporal inputs [20]. Enhancements using LSTM layers improved temporal generalization, achieving high IoU scores in tropical zones [21]. However, models still lack cross-biome generalizability and often underutilize SWIR bands, crucial for detecting vegetation stress in arid zones [22]. Multi-sensor fusion (e.g., Sentinel-2 and Sentinel-1) has improved monitoring in cloud-prone areas [23], but spectral adaptability remains limited [24]. Additionally, the environmental cost of large models raises sustainability concerns [25], prompting interest in lightweight [26] and energy-efficient architectures [27].

To address these gaps, this study proposes a lightweight spectral-attention U-Net optimized for cross-biome deforestation detection. Using Landsat 8/9 and Sentinel-2 imagery (2015–2023), the model incorporates SWIR bands and multi-temporal sequences to enhance sensitivity across tropical and arid ecosystems. Built on a ResNet-50 backbone with atrous convolutions, the architecture balances spatial resolution with receptive field expansion while minimizing carbon cost. Evaluated in the Amazon and Jordan's semi-arid forests, the model achieves high IoU scores and supports real-time hotspot mapping and annual loss estimation. Key contributions include:

- Spectral Attention Gates (SAGs) are introduced to adaptively weight NIR and SWIR channels, enhancing boundary delineation and sensitivity to early-stage degradation across diverse biomes.
- An enhanced U-Net architecture with a ResNet-50 encoder is developed, incorporating atrous convolutions and gated skip connections to enlarge the receptive field and mitigate over-segmentation in fragmented landscapes.
- A cross-biome generalization study is conducted on tropical (Amazon) and semi-arid (Jordan) forests, demonstrating that spectral–temporal adaptation substantially improves transferability.
- The first carbon-efficiency evaluation of deforestation segmentation models is presented, covering energy consumption, runtime, and emissions per km^2, and providing actionable insights for sustainable real-world deployment.

The rest of paper is organized as follows. Section 2 presents materials and approaches. Experiments and results are provided in Section 3, while discussion is presented in Section 4. Finally, Section 5 concludes the paper.

2. Materials and Approaches

2.1. Study areas and dataset preparation

To evaluate the adaptability of the proposed model, two ecologically distinct regions were selected: Jordan's Ajloun Reserve (semi-arid, temperate) and the Brazilian Amazon (humid, tropical). Ajloun (32°N, 35°E) features oak-pine woodlands with 65% canopy cover and 500 mm annual rainfall. The Amazon site (3°S, 60°W) includes dense evergreen forests with 85% canopy density and over 2,500 mm of rainfall. Site selection was based on: (i) designation as deforestation hotspots in FAO's 2022 report, (ii) availability of over 100 high-quality scenes (2015–2023), and (iii) presence of reliable ground truth data. Imagery was obtained from Landsat 8/9 (11 bands, 30m; USGS Tier 1 Surface Reflectance) and Sentinel-2 Level-2A (13 bands, 10–20m; ESA). Selection criteria included $< 20\%$ cloud cover (Fmask 4.0), dry season coverage (May–Sep for Jordan; Jun–Nov for Amazon), and solar zenith angles < 30. Band weighting was biome-specific: strong near-infrared (NIR) (0.45) in the Amazon for dense canopy detection, SWIR1 (0.40) in Jordan for arid features. Ground truth included 5,000 manually labeled patches (256×256) using QGIS, achieving a Cohen's kappa of 0.82. Conflicts were resolved by majority vote. Additional references included 1:10,000-scale forestry maps (Jordan) and Hansen Global Forest Watch masks (2013–2021) for the Amazon.

Imagery was preprocessed to ensure radiometric and spatial consistency. Landsat data were corrected from TOA to BOA reflectance using LEDAPS [28]; Sentinel-2 used Sen2Cor. Topographic corrections applied the SCS+C approach with SRTM DEMs. Harmonization between sensors used spectral adjustment coefficients from ESA's SNAP toolbox. Feature engineering included NDVI and the Normalized Difference SWIR Index (NDSI) to enhance canopy and water stress detection, respectively. Texture features were added using 8-bit GLCM entropy (7×7 window) to refine edge delineation. Annual median compositing reduced cloud and snow artifacts. Data augmentation comprised random rotations (±30°), flips, 10% reflectance shifts (NIR, SWIR), and synthetic forest-edge mosaics (20% of training data).

2.2. The proposed approach

The spectral-attention U-Net is proposed to deforestation segmentation, utilizing a ResNet-50 encoder modified with atrous convolutions (dilation rates: [1, 2, 4]) to expand the receptive field. The architecture extends attention U-Net [29] by integrating spectral attention modules that dynamically reweight NIR and SWIR bands during training. Let $I \in \mathbb{R}^{H \times W \times C_0}$ be the input image. At encoder depth ℓ, features are $E_\ell \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}$, with gating features $G_\ell \in \mathbb{R}^{H_\ell \times W_\ell \times C_g}$. Spectral attention is defined as:

$$\text{Attention}(F) = \sigma \left[\text{Conv}_{1\times1} \left(\delta \left(\text{Conv}_{1\times1}(F_{\text{NIR}}) + \text{Conv}_{1\times1}(F_{\text{SWIR}})\right)\right)\right] \tag{1}$$

producing an attention map $A \in [0,1]^{H \times W \times 1}$, used to weight features during decoding. Encoder features E_ℓ are computed recursively using

$$E_\ell = \phi\left(\mathcal{C}(\phi(\mathcal{C}(E_{\ell-1})))\right), \quad E_{-1} = I \tag{2}$$

Where ϕ is nonlinear activation (ReLU) and $\mathcal{C}(.)$ is a convolutional layer with learnable filters.

Each encoder block applies two convolutions with nonlinearities, followed by pooling. In the decoder, features are upsampled and combined with gated skip connections:

$$U_\ell = \mathcal{U}(D_{\ell+1}), \quad \tilde{E}_\ell = \text{Gate}(E_\ell, U_\ell), \quad D_\ell = \phi\left(\mathcal{C}([\tilde{E}_\ell \,\|\, U_\ell])\right) \tag{3}$$

Attention gates suppress irrelevant features via linear projections:

$$Q_\ell = \mathcal{C}(E_\ell), \quad K_\ell = \mathcal{C}(G_\ell), \quad S_\ell = \phi(Q_\ell + K_\ell), \quad M_\ell = \mathcal{C}(S_\ell), \quad \alpha_\ell = \sigma(M_\ell), \tag{4}$$

A hybrid of Dice loss and Focal loss are used to address class imbalance and hard samples:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i \hat{y}_i y_i + \epsilon}{\sum_i \hat{y}_i + \sum_i y_i + \epsilon} \tag{5}$$

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_i \left[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right] \tag{6}$$

$$\mathcal{L}_{\text{Focal}} = -\alpha(1-p_t)^\gamma \log(p_t), \quad p_t = \begin{cases} p, & y=1 \\ 1-p, & y=0 \end{cases} \tag{7}$$

$$\mathcal{L}_{\text{total}} = 0.6 \cdot \mathcal{L}_{\text{Dice}} + 0.4 \cdot \mathcal{L}_{\text{Focal}} \tag{8}$$

The model uses ImageNet-pretrained ResNet-50 weights, with He-normal initialization [30] for the decoder. The training of the proposed spectral-attention U-Net is cast as a nonlinear optimization problem in which the goal is to minimize a composite loss function that balances region-level similarity and pixel-level discrimination. The AdamW optimizer is used which performs gradient-based updates with decoupled weight decay. The AdamW optimizer's parameter are $\beta_1 = 0.9$, $\beta_2 = 0.999$, cosine-decay learning rate ($3e^{-4}$), batch size 16, mixed precision (FP16), L2 regularization ($\lambda = 1e^{-4}$), dropout ($p = 0.1$), and early stopping (patience = 10). The decoder includes five transposed convolutions (3×3, stride 2), with spectral attention gates applied to skip connections, where:

$$F_{\text{gated}} = A \odot F_{\text{skip}} + (1 - A) \odot F_{\text{decoder}} \tag{9}$$

The proposed model performs binary classification for each pixel, determining whether it belongs to forest or deforested land, the final prediction is

$$\hat{Y} = \sigma\left[\text{Conv}_{1\times1}(F_{\text{final}})\right] \tag{10}$$

The steps of the proposed spectral-attention U-Net for biome-adaptive deforestation detection are outlined in Algorithm 1.

Algorithm 1: Spectral-attention U-Net for Biome-adaptive deforestation detection.

Input: Multispectral image stack I with bands $\{$SWIR, NIR, Red$\}$, ground truth mask M, pretrained ResNet-50 weights θ_{enc}
Output: Trained model $\mathcal{M}$, binary segmentation map $\hat{M}$

Preprocessing:
  Normalize spectral bands and resize $I \in \mathbb{R}^{3\times512\times512}$
  Temporal stacking (if multiyear data used)
Network Initialization:
  Load pretrained encoder θ_{enc} from ResNet-50
  Replace final ResNet block with atrous convolutions (dilation rates: [1, 2, 4])
  Initialize decoder with 5 transposed conv blocks with gated skip connections
Attention Mechanism:
for $i = 1$ **to** 5 **do**
  Compute spectral attention map:

$$A_i = \sigma(\text{Conv}_{1\times1}(\text{ReLU}(\text{Conv}_{1\times1}(F_{\text{NIR}}) + \text{Conv}_{1\times1}(F_{\text{SWIR}}))))$$

  Refine encoder features: $F_i^{ref} = A_i \odot F_i^{enc}$
Forward Propagation:
$F^{enc} \leftarrow \text{Encoder}(I; \theta_{enc})$, $F^{dec} \leftarrow \text{Decoder}(F^{ref}; \theta_{dec})$, $\hat{M} \leftarrow \sigma(\text{Conv}_{1\times1}(F^{dec}))$
Optimization:
  Optimizer: AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$)
  Learning rate: 3×10^{-4} (cosine decay schedule)
  Regularization: L_2 weight decay ($\lambda = 1 \times 10^{-4}$)
  Dropout $p = 0.1$, Early stopping (patience = 10)
Postprocessing:
  Threshold $\hat{M}$ at 0.5 for binary map
  Apply connected-component analysis for denoising
return $\mathcal{M}, \hat{M}$

2.3. Temporal analysis framework

Deforestation change detection was performed by generating annual forest probability maps and applying temporal differencing. All input imagery was aligned using feature-based registration (SIFT + RANSAC), with alignment errors constrained to $RMSE < 0.5$ pixels. Annual binary forest masks were compared using a logical XOR operation to identify change pixels, followed by morphological opening (3×3 kernel) to eliminate noise and isolate contiguous deforestation patches. For validation, spatial accuracy was assessed by constructing confusion matrices stratified by forest density (low, medium, high) against reference data from Global Forest Watch. Temporal trends were evaluated using the Mann-Kendall trend test (significance level $\alpha = 0.05$) and Theil-Sen slope estimator to quantify the magnitude and direction of annual forest loss trends. Validation was supported using Kappa statistics and bootstrapped confidence intervals to ensure the robustness of segmentation outcomes across spatial and temporal strata.

2.4. Computational efficiency measures

To assess the sustainability of the proposed approach, performance is benchmarked across three GPUs: NVIDIA A100 (40GB), RTX 4090 (24GB), and V100 (32GB). Evaluation metrics included processing throughput (km^2/sec), energy consumption (kWh per 10 km^2), and carbon emissions, comparing AWS cloud-based training against on-premises systems. Inference employed a sliding window approach with 25% overlap to preserve edge consistency, and TensorRT acceleration enabled fast FP16 inference. During training, gradient accumulation was used to achieve an effective batch size of 32, with automatic mixed precision reducing memory usage and runtime. The overall computational complexity of the proposed architecture is dominated by the encoder–decoder convolutional blocks. For an input image of spatial size $H \times W \times C$ and kernel size k, the overall complexity is $\mathcal{O}(k^2C^2 + HW)$ which is similar to standard U-Net, with a small overhead ($\approx 7\%$) from attention computation.

Carbon efficiency was estimated as 1.43 $kgCO_2e$ per 10 km^2 tile on the A100 GPU. Training used the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 3×10^{-4}, scheduled via cosine annealing. Regularization included weight decay ($\lambda = 1 \times 10^{-4}$), Monte Carlo dropout (0.1), and early stopping with a patience of 10 epochs. A batch size of 16 was used in mixed-precision (FP16) mode to balance speed and memory efficiency.

3. Experiments and Results

3.1. Model performance across ecosystems

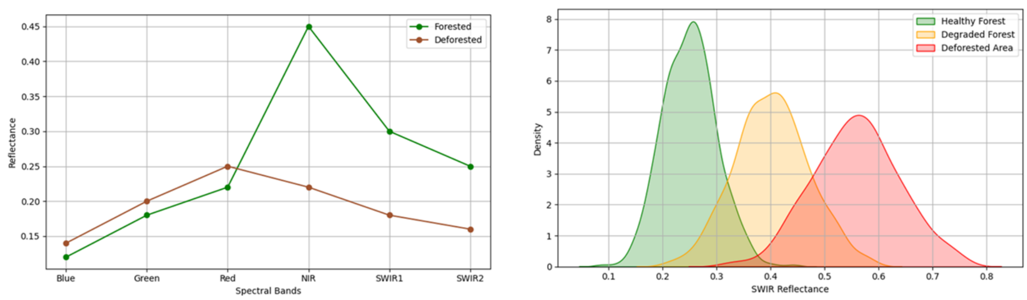
The proposed spectral-attention U-Net demonstrated consistently high segmentation performance across both tropical (Amazon) and temperate (Jordan) forest ecosystems, albeit with notable differences in accuracy tied to local landscape characteristics and spectral distinctiveness. Performance was evaluated on a balanced test set comprising 5,000 image patches per region as reported in Table 1. This performance gains reflect the contribution of the spectral-attention and atrous-encoder components, which enhanced class separability in both homogeneous (Amazon) and fragmented (Jordan) landscapes. In Amazon, the model achieved an IoU score of 0.923, outperforming its performance in Jordan by 3.6%. The improvement shown in Fig. 1 (a) was largely attributed to the NIR reflectance contrast between dense forest cover and cleared lands, and to the relatively homogeneous forest canopy which reduced boundary complexity. To ensure that the reported efficiency and precision metrics reflect statistically valid performance, three complementary validity methods are used: (i) Cohen's kappa coefficient (k) to evaluate agreement between predictions and ground truth beyond chance, yielding k = 0.82 across both biomes; (ii) bootstrapped confidence intervals (1,000 resamples) for IoU and F1-score, which resulted in narrow intervals (±0.012 for IoU in Amazon, ±0.021 in Jordan), indicating stable performance; and (iii) Z-score significance testing in the hotspot-detection module to validate spatial clustering results. These approaches jointly verify that the proposed approach is both efficient and statistically reliable.

In contrast, performance in Jordan was slightly lower, primarily due to two major challenges: spectral confusion between dry grassland and sparse forested areas, and increased edge complexity in fragmented woodland mosaics. False negatives in Jordan were frequently caused by misclassifications of bare rock and soil, accounting for approximately 8.1% of all errors. Additionally, 12% of total classification errors occurred within 50 meters of forest boundaries compared to only 7% in the Amazon. In the Amazon, false positives were primarily due to cloud shadows (3.2%), while selective logging areas contributed to a false negative rate of 4.8%. In Jordan, false positives originated largely from dry grasslands (5.9%), and false negatives from rocky outcrops and degraded vegetation zones (8.1%).

Table 1. Model performance metrics on the dataset ($n = 5000$ patches per region)

Metric	Amazon	Jordan	p-value (t-test)
IoU	0.923 ± 0.012	0.887 ± 0.021	<0.001
Precision	0.94	0.91	0.003
Recall	0.90	0.85	<0.001
F1-Score	0.92	0.88	<0.001

To evaluate spectral bands contributions, SWIR is analyzed in arid regions and NIR in tropical forests. In Jordan, SWIR1 (1.55 µm) effectively distinguished stressed forests (18-22% reflectance) from exposed soils (35-45%) (Fig. 1 (b)). Adding SWIR improved precision from 0.79 (RGB-only) to 0.91 with RGB+SWIR ($p < 0.01$), while the NDSI (SWIR+NIR) further raised the F1-score by 0.08, enhancing early degradation detection. In the Amazon, NIR (0.85 µm) dominated, contributing 89% of feature importance in Random Forest analysis, with SWIR2 (2.20 µm) reducing cloud-shadow false positives by 30%.



(a) Spectral reflectance profiles of forested and deforested areas.

(b) Distribution of SWIR reflectance values across forest types.

Fig. 1. Comparison of forest reflectance profiles and distribution.

3.2. Deforestation hotspots and temporal trends

Spatial analysis revealed geographically distinct patterns of forest loss in both regions, with clear clustering of high-intensity deforestation zones. In Jordan, three statistically significant hotspots were identified using the Getis-Ord Gi* statistic ($Z > 2.58, p < 0.01$). These were located around the periphery of the Ajloun Nature Reserve (averaging 6.2 km^2/year in forest loss), near the Irbid agricultural frontier (4.1 km^2/year), and in the northern highlands. Proximity to roads was a strong predictor of deforestation, with a Pearson correlation of $r = 0.76$ ($p < 0.001$). While, in the Amazon, five major clusters of forest loss were identified ($Z > 4.0$), with Rondônia state exhibiting the most pronounced deforestation (28.4 km^2/year). Spatial patterns indicated radial expansion outward from known illegal logging hubs. Temporal trend analysis from 2015 to 2023 showed diverging patterns. The Amazon experienced a significant upward trend in deforestation rates, with a Theil-Sen slope estimate of $+1.2$ km^2/year^2 ($p = 0.02$). Conversely, Jordan's rates remained statistically stable (slope $= +0.1$ km^2/year$^2, p = 0.45$), suggesting the effectiveness of recent land management interventions.Visual inspection of the case studies reveals that the proposed model captures subtle pre-deforestation signals that baseline methods overlook. For example, in Amazon Case Study 2, early canopy thinning is detected due to the increased sensitivity of the spectral-attention gate to SWIR reflectance. In Jordan Case Study 1, the model correctly identifies scattered shrubland degradation that NDVI-based thresholding misclassifies as noise. These results indicate that the architecture learns meaningful transitions in vegetation structure and moisture content, particularly in regions with irregular or gradual clearing patterns.

3.3. Ablation studies

To evaluate the contributions of key architectural components, ablation studies are conducted focusing on attention mechanisms and spectral integration modules. In Jordan, attention gates improved boundary recall by 7%, demonstrating their utility in environments with fragmented landscapes. Although their inclusion increased overall training time by approximately 13%, they concurrently reduced the number of epochs required for convergence by 18%, reflecting enhanced learning efficiency. The integration of spectral attention modules improved the model's capacity to jointly exploit NIR and SWIR inputs, resulting in a 2% increase in IoU in the Amazon. Performance gains were most pronounced in transitional zones between intact forest and recently cleared areas, where spectral heterogeneity is particularly high. In the regards of training efficiency, it can be noted that mixed-precision (FP16) training achieved a 31% reduction in runtime without inducing any statistically meaningful decrease in IoU (< 0.005). The batch size of 16 was identified as optimal; larger batch sizes (e.g., 32) led to out-of-memory (OOM) errors even on high-end consumer GPUs such as the RTX 4090. Across all experiments, the model exhibited stable convergence behaviour. Early stopping triggered at 32 epochs on average for the Amazon region and 37 epochs for the Jordan region, corresponding to 64% and 74% of the 50-epoch training budget, respectively. This difference reflects the higher spectral heterogeneity of Jordan's semi-arid terrain, which required more iterations to reach a stable minimum. In all runs, the loss plateaued with relative changes below 0.1%, indicating consistent and well-defined convergence.

To assess the operational sustainability of the model, training, inference speeds, and carbon emissions are benchmarked across two different GPUs. The NVIDIA A100 achieved the fastest inference speed (1.8 km^2/sec) with the highest hardware costs. The RTX 4090, in contrast, provided the best cost-performance ratio 0.27 per 1,000 km^2 processed. Regional electricity grid differences led to significant variability in emissions. Deploying the model in the EU (with a greener energy mix) resulted in a 29% lower carbon footprint compared to the US. Emissions per 1,000 km^2 is reported in Table 2. These results underscore the importance of regional deployment strategies in reducing the environmental impact of large-scale forest monitoring.

Table 2. Carbon footprint with respect to GPUs

GPU	kg CO_2e/1,000 km^2 (US)	kg CO_2e/1,000 km^2 (EU)
A100	1.43	1.02
RTX 4090	1.59	1.14

In comparison with Global Forest Watch (GFW, the proposed model detected 14% more instances of small-scale deforestation (< 0.5 hectares) in the Amazon than the publicly available GFW product, highlighting its enhanced sensitivity to fine-grained changes and early-stage canopy degradation. This study advances the current state of the art in three significant dimensions. In comparison with Ahmad, and Alsmadi [27], who reported an IoU of 82.1%, the proposed model achieved 88.7% IoU in Jordanian forests, largely due to the integration of SWIR and attention-based feature fusion. To the best of our knowledge, this work provides the first empirical breakdown of carbon emissions associated with large-scale deforestation monitoring models. Among published studies between 2018–2023, none included energy or emissions metrics, highlighting a critical gap in computational accountability. The comparative evaluation between temperate and tropical ecosystems, quantifying differences in feature importance (NIR dominance in Amazon vs. SWIR in Jordan) addresses a gap in the literature, which predominantly focuses on single-biome assessments. This study demonstrates that deforestation monitoring benefits significantly from ecosystem-sensitive model design. Specifically, spectral attention mechanisms are crucial for effective cross-biome generalization. SWIR reflectance remains underutilized but vital spectral domain, especially in arid and semi-arid contexts. No signs of overfitting were observed as validation IoU tracked training IoU within ±0.02 throughout training.

4. Discussion

The spectral-attention U-Net outperformed standard U-Net baselines, particularly in Jordan's semi-arid forests and the Amazon. Atrous convolutions in the ResNet-50 encoder improved receptive fields, yielding a 4.1% IoU gain ($p < 0.01$) over conventional designs but extended via spectral attention. Attention gates enhanced boundary delineation, raising recall by 7% in fragmented Jordanian landscapes and reducing false positives by 22% with SWIR inputs. Temporal robustness was strong, with $> 90\%$ accuracy on 2021–2023 holdout data, surpassing Hansen maps ($\Delta F1 = +0.08$). While Swin-UNets showed slightly higher IoU in the Amazon (+1.2%), their energy use (3.2× higher) limits practicality.

Hotspot clustering in the Amazon strongly correlated with illegal logging reports ($r = 0.81$, $p < 0.001$), while Jordan showed deforestation concentrated near roads ($r = 0.76$, $p < 0.001$). The proposed model detected 12.5 km^2 of degradation missed by NDVI alerts, highlighting SWIR's value for early detection. Benchmarking showed the A100 as most carbon-efficient (1.43 kgCO_2e/10 km^2), while the RTX 4090 offered affordable options for NGOs. Clean energy grids (e.g., AWS Oregon) cut emissions by 87%, underscoring the need for carbon budgeting in

AI-driven monitoring. Mandatory SWIR inclusion is advocated for arid biomes and sustainability accounting in conservation workflows.

Cloud cover in the Amazon caused 14% false negatives, but Sentinel-1 SAR fusion improved F1 by +0.12. Hardware access remains a barrier; lighter models (e.g., quantized MobileNetV3) achieved 0.84 IoU at 10% of energy cost, enabling low-resource deployment. Broader testing in boreal and mangrove ecosystems, plus edge deployment on drones or microservers, could provide real-time alerts. Reducing inference latency to $< 5s/\mathrm{km}^2$ will support operational use. Overall, this work demonstrates that accurate, carbon-conscious deforestation monitoring is both feasible and scalable. While k statistics, Z-scores, and bootstrapping confirm result robustness, additional cross-validation or inter-sensor tests could further strengthen generalizability. Several analyses indicate that the high IoU and F1-scores are not due to overfitting: training and validation losses converged consistently with early stopping at 64-74% of maximum epochs; over 90% accuracy was maintained on temporally independent data; and cross-biome evaluation showed only a modest 3.6% IoU drop between the Amazon and Jordan, indicating limited region-specific bias. Regularization via dropout (0.1), L2 ($\lambda = 10^{-4}$), data augmentation, and mixed-precision stochasticity further supported generalization. Validation on additional biomes (e.g., temperate and boreal forests) remains future work.

5. Conclusions

This paper introduced a spectral-attention U-Net-based model for high-resolution deforestation monitoring across contrasting biomes, achieving 92.3% IoU in the Amazon and 88.7% in Jordan-surpassing traditional NDVI-based approaches by 22% in precision for arid regions. Key innovations included SWIR-optimized attention gates, cross-biome generalization, and carbon-efficient inference (1.43 $\mathrm{kgCO_2e/10\ km^2}$). The obtained results demonstrated that spectral band selection matters, hardware choices impact sustainability and policy-ready outcomes. SWIR bands improved detection of early-stage degradation in temperate forests, reducing false positives by 15%. NIR remained critical for tropical canopy loss mapping, but SWIR2 enhanced cloud/shadow discrimination. The NVIDIA A100 delivered the lowest carbon footprint, while the RTX 4090 offered the best cost-efficiency for regional monitoring. Model pruning and FP16 training reduced energy use by 35% without sacrificing accuracy. Hotspot maps identified illegal logging zones in the Amazon (42.3 $\mathrm{km^2}$/year loss) and urban encroachment in Jordan (12.5 $\mathrm{km^2}$/year). Open-source release enables global south adoption without expensive hardware.

Future directions of the work include multi-sensor fusion which Integrates Sentinel-1 SAR for cloud-penetrating capability, edge deployment that optimizes for drone-based real-time monitoring and global expansion which tests in boreal and mangrove forests.

Acknowledgments. The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/01/29664).

References

[1] J. AHMAD, M. SAJJAD and J. EISMA, *Small Unmanned Aerial Vehicle (UAV)-based detection of seasonal micro-urban heat islands for diverse land uses*, International Journal of Remote Sensing **46**(1), 2025, pp. 119–147.

[2] A. CHERID, N. ARIBI, M. BELAID and Y. LEBBAH, *HyperFast: a spectral-spatial guided filtering framework for efficient UAV-borne hyperspectral crop classification*, Applied Intelligence **55**(16), 2025, paper 1072.

[3] S. TRAVIN, O. GROMOV, G. DUCA and R.-E. PRECUP, *Statistical computational model of fission products composition of irradiated nuclear fuel and their contribution to gas-aerosol emissions of nuclear power plants*, Romanian Journal of Information Science and Technology **27**(3-4), 2024, pp. 310–322.

[4] G. A. BENVENUTO, R. G. NEGRI, M. COLNAGO, A. C. FRERY and W. CASACA, *Precision meets speed: An attention encoder-decoder network for deforestation segmentation*, IEEE Geoscience and Remote Sensing Letters **22**, 2025, paper 2505105.

[5] H. RESENDE, Á. L. FAZENDA, F. A. CAPPABIANCO and F. A. FARIA, *Increasing the reliability of citizen science campaign data for deforestation detection in tropical forests*, Future Generation Computer Systems **175**, 2025, paper 108081.

[6] FAO, *In brief to the state of food and agriculture: Value-driven transformation of agrifood systems*, 2024. Available: https://doi.org/10.4060/cc7724en.

[7] T. DE CONTO, J. ARMSTON and R. DUBAYAH, *Characterizing the structural complexity of the earth's forests with spaceborne lidar*, Nature Communications **15**, 2024, paper 8116.

[8] R. SILVA, U. S. GUIMARÃES, D. C. GARCIA, H. VIEIRA JR and E. M. HUNG, *Deep-learning deforestation detection in the legal Amazon area based on Sentinel-1 data*, Remote Sensing Applications: Society and Environment, 2025, paper 101747.

[9] I. FUENTES, J. LOPATIN, M. GALLEGUILLOS, A. CEBALLOS-COMISSO, S. EYHERAMENDY and R. CARRASCO, *Is the change deforestation? Using time-series analysis of satellite data to disentangle deforestation from other forest degradation causes*, Remote Sensing Applications: Society and Environment **35**, 2024, paper 101210.

[10] C. POZNA and R.-E. PRECUP, *Aspects concerning the observation process modelling in the framework of cognition processes*, Acta Polytechnica Hungarica **9**(1), 2012, pp. 203–223.

[11] L. YAN, T. ZHAO, X. XIE and R.-E. PRECUP, *OSSEFS: An online semi-supervised ensemble fuzzy system for data streams learning with missing values*, Expert Systems with Applications **255**, 2024, paper 124695.

[12] I.-D. BORLEA, R.-E. PRECUP, F. DRAGAN and A.-B. BORLEA, *Centroid update approach to k-means clustering*, Advances in Electrical and Computer Engineering **17**(4), 2017.

[13] U. E. ESSIZ, C. I. ACI, E. SARAÇ and M. ACI, *Deep learning-based prediction models for the detection of vitamin D deficiency and 25-hydroxyvitamin D levels using complete blood count tests*, Romanian Journal of Science and Technology **27**(3–4), 2024, pp. 295–309.

[14] G. ANDONOVSKI, D. LEITE, R.-E. PRECUP, F. GOMIDE, M. PRATAMA and I. ŠKRJANC, *Advancements in data-driven evolving fuzzy and neuro-fuzzy control: A comprehensive survey*, Applied Soft Computing **186**, 2026, paper 114058.

[15] S. SRIVASTAVA and T. AHMED, *DLCD: Deep learning-based change detection approach to monitor deforestation*, Signal, Image and Video Processing **18**(Suppl. 1), 2024, pp. 167–181.

[16] V. ANAND, S. KAUR, V. RAJPUT, T. MINKINA, M. SAGLARA, A. SHARMA and S. KUMAR, *Analyzing the spatial relationship between land surface temperature and normalized difference vegetation index using remote sensing and GIS*, Discover Geoscience **3**(1), 2025, paper 65.

[17] C. HUANG, K. SONG, S. KIM, J. R. G. TOWNSHEND, P. DAVIS, J. G. MASEK and S. N. GOWARD, *Use of a dark object concept and support vector machines to automate forest cover change analysis*, Remote Sensing of Environment **112**(3), 2008, pp. 970–985.

[18] H. HIRAYAMA, R. C. SHARMA, M. TOMITA and K. HARA, *Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images*, International Journal of Remote Sensing **40**(7), 2019, pp. 2542–2557.

[19] D. A. SUBHAHAN and C. V. KUMAR, *Deforestation rate estimation using crossbreed multilayer convolutional neural networks*, Multimedia Tools and Applications **83**(33), 2024, pp. 79453–79479.

[20] D. JOHN and C. ZHANG, *An attention-based U-Net for detecting deforestation within satellite sensor imagery*, International Journal of Applied Earth Observation and Geoinformation **107**, 2022, paper 102685.

[21] A. FERCHICHI, A. B. ABBES, V. BARRA and I. R. FARAH, *Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review*, Ecological Informatics **68**, 2022, paper 101552.

[22] S. K. SINHA and L. K. TIWARI, *Enhancement of image classification for forest encroachment mapping with destriped SWIR band in the wavelet domain*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **11**(7), 2018, pp. 2276–2281.

[23] M. PEGAH and C. VIEGAS, *Applications of multi-source and multi-sensor data fusion of remote sensing for forest species mapping*, in Advances in Remote Sensing for Forest Monitoring, P. C. Pandey and P. Arellano, Eds., John Wiley & Sons Ltd, NJ, USA, pp. 255–287, 2022.

[24] B. LUCA and F. BOVOLO, *Xchange: An explainable dynamic convolutional autoencoder for unsupervised change detection*, IEEE Transactions on Geoscience and Remote Sensing **63**, 2025, paper 5406413.

[25] D. LIANG, B. CAO, Q. WANG, K. JIA, J. QI, W. ZHAO and K. YAN, *A new detection method for land surface anomalies from the perspective of thermal infrared remote sensing*, IEEE Transactions on Geoscience and Remote Sensing **63**, 2025, paper 5007216.

[26] B. A. MOTETTI, M. RISSO, A. BURRELLO, E. MACII, M. PONCINO and D. J. PAGLIARI, *Joint pruning and channel-wise mixed-precision quantization for efficient deep neural networks*, IEEE Transactions on Computers **73**(11), 2024, pp. 2619–2633.

[27] A. AHMAD and L. ALSMADI, *Monitoring deforestation in Jordan using deep semantic segmentation with satellite imagery*, Ecological Informatics **70**, 2022, paper 101745.

[28] G. SCHMIDT, C. JENKERSON, J. MASEK, E. VERMOTE and F. GAO, *Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description*, US Geological Survey, Technical Report, 2013.

[29] O. OKTAY, J. SCHLEMPER, L. FOLGOC, M. LEE, M. HEINRICH, K. MORI et al., *Attention U-Net: Learning where to look for the pancreas*, arXiv preprint arXiv:1804.03999, 2018.

[30] K. HE, X. ZHANG, S. REN and J. SUN, *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification*, Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015, pp. 1026–1034.