OPEN ACCESS
Research of Scientia Naturalis
Vol. 3 No. 4. August 2026, pp. 368-389

Check for updates

DOI. 10.70177/scientia.v3i4.4327

Research Article

DEEP LEARNING ARCHITECTURES FOR AUTOMATED CLASSIFICATION OF SATELLITE IMAGERY TO MONITOR GLOBAL DEFORESTATION TRENDS AND LAND USE CHANGES

Sa'dianoor[1], Chen Mei[2], and Yang Xiang[3]
[1] Universitas Muhammadiyah Banjarmasin, Indonesia
[2] Zhejiang University, China
[3] Beijing Normal University, China

Corresponding Author:

Sa'dianoor,
Department of Urban and Regional Planning, Faculty of Engineering, Universitas Muhammadiyah Banjarmasin.
Jln. Gubernur Syarkawi, Semangat Dalam, Kecamatan Alalak, Kabupaten Barito Kuala, Kalimantan Selatan 70581 Indonesia.
Email: sadianoor@umbjm.ac.id

Article Info

Received: February 7, 2026
Revised: May 12, 2026
Accepted: July 14, 2026
Online Version: August 25, 2026

Abstract

Global deforestation and land-use change require timely, accurate, and transferable monitoring systems capable of processing extensive satellite archives. This study aimed to evaluate deep learning architectures for automated classification of satellite imagery, detection of forest change, and identification of post-deforestation land uses across diverse ecological regions. A comparative experimental design was applied to 120,000 labelled image patches derived from Landsat, Sentinel-1, and Sentinel-2 data representing South America, Southeast Asia, Central Africa, North America, Europe, and Oceania. Nine architectures, including convolutional, recurrent, Siamese, Transformer-based, and hybrid models, were assessed using spatial, temporal, and cross-regional validation. The hybrid CNN–Transformer achieved the strongest overall accuracy, calibration, and transfer performance, while radar–optical fusion improved classification in persistently cloudy tropical environments. Multitemporal models reduced confusion between permanent deforestation, seasonal disturbance, harvesting, and regeneration. Geographic holdout testing produced substantial performance declines across all architectures, demonstrating that random partitioning overestimated operational reliability. The study concludes that effective global deforestation monitoring requires multimodal data, temporal context, uncertainty estimation, active learning, and geographically independent evaluation. Automated classification can strengthen forest surveillance, yet expert verification and statistically adjusted area estimation remain essential for responsible environmental decision-making and policy implementation under rapidly changing climatic, technological, and land-management conditions worldwide over time.

Keywords: Automated Classification, Deep Learning, Deforestation Monitoring

Journal Homepage	https://research.adra.ac.id/index.php/scientia
How to cite:	Sa'dianoor., Mei, C., & Xiang, Y. (2026). Deep Learning Architectures for Automated Classification of Satellite Imagery to Monitor Global Deforestation Trends and Land Use Changes. *Research of Scientia Naturalis*, *3*(4), 368–389. https://doi.org/10.70177/scientia.v3i4.4327
Published by:	Yayasan Adra Karima Hubbi

INTRODUCTION

Global deforestation and land-use change remain major environmental concerns because the conversion and degradation of forests affect biodiversity, carbon storage, hydrological systems, ecosystem services, and the livelihoods of forest-dependent communities. Reliable monitoring is therefore essential for identifying where forest loss occurs, determining its extent, and distinguishing temporary canopy disturbance from permanent land conversion (Hussain et al., 2025; Liu et al., 2025). Satellite-based Earth observation has greatly expanded the capacity to monitor forests across large and inaccessible areas. Long-term Landsat archives, Sentinel-1 radar data, Sentinel-2 multispectral imagery, and higher-resolution commercial imagery now provide complementary information for detecting changes across spatial and temporal scales (Goel & Vishnoi, 2025; Rashid et al., 2025).

Satellite imagery has also supported the development of global and regional land-cover products with increasingly fine spatial and temporal resolution. Dynamic World, for example, applies deep learning to Sentinel-2 imagery to generate near-real-time land-cover predictions at a spatial resolution of 10 metres (da Silva et al., 2025; Silpalatha & Jayadeva, 2025). Sentinel2GlobalLULC and OpenEarthMap provide additional resources for training and evaluating land-cover classification systems across geographically diverse regions. These datasets illustrate a broader transition from manually interpreted or periodically updated land-cover maps toward automated systems capable of processing large volumes of satellite observations and producing frequent classifications (Fuentes et al., 2024; Hosseini et al., 2025).

Deep learning has become central to this transition because it can learn complex spatial, spectral, and temporal representations directly from satellite imagery. Convolutional neural networks are effective for extracting local textures, boundaries, and multiscale spatial patterns, while semantic-segmentation architectures can assign a land-cover class to every image pixel. Vision Transformers and hybrid CNN–Transformer models offer stronger mechanisms for capturing long-range spatial dependencies and contextual relationships (Mahendra et al., 2025; Xiong et al., 2024). Self-supervised systems such as SatMAE and large-scale pretraining initiatives such as SatlasPretrain further demonstrate that models can acquire transferable geospatial representations from extensive multispectral datasets before being adapted to specific classification and change-detection tasks (Cullerton et al., 2025; Dalagnol et al., 2023).

The central research problem concerns the reliability of deep learning architectures when they are applied to automated deforestation and land-use classification across heterogeneous global environments. Models trained in one geographical region may perform poorly in another because forest structure, agricultural patterns, soil reflectance, seasonal conditions, atmospheric effects, and land-management practices vary considerably (Kopec et al., 2025; Mahendra et al., 2024). Tropical rainforests, dry forests, plantations, agroforestry areas, and boreal forests may exhibit different spectral and structural characteristics despite being assigned similar labels. High accuracy within a local test dataset therefore does not necessarily indicate that a model can generalize across countries, biomes, sensors, or observation periods (Habibie et al., 2024; Hu et al., 2023).

A second problem involves the distinction between tree-cover loss, natural-forest loss, forest degradation, and subsequent land use. A satellite image may indicate canopy removal without revealing whether the area has been converted to cropland, pasture, settlement, mining, plantation forest, or temporary clearing (Colkesen et al., 2025). Recent continental-scale research in Africa demonstrated that deep learning and active learning could classify 15 forms of land use following deforestation at five-metre resolution, but the study also showed that limited, geographically heterogeneous, and imperfect reference labels restricted model generalization. Natural-forest baselines are similarly needed to distinguish the loss of ecologically important forests from harvesting cycles in plantations or tree crops (Altarez et al., 2023; Lai et al., 2025).

A third problem arises from differences among satellite sensors and image-acquisition conditions. Optical imagery may be obstructed by clouds, haze, smoke, and seasonal illumination, particularly in humid tropical regions where rapid monitoring is most needed. Medium-resolution data may overlook narrow roads, selective logging, fragmented clearings, small agricultural plots, and individual trees (Rahman et al., 2024; Ramadan et al., 2024). Radar imagery can provide observations through clouds but introduces different noise characteristics and interpretive challenges. Recent high-resolution research has shown that substantial areas of pan-tropical tree cover may remain undetected by conventional products, emphasizing that spatial resolution and forest definition can materially affect estimates of cover and change (Debus et al., 2025; Shaikh et al., 2024).

The primary objective of this study is to compare the efficacy of major deep learning architectures for classifying satellite imagery and detecting deforestation-related land-use changes. The comparison should include convolutional neural networks, encoder–decoder segmentation models, Siamese change-detection networks, Vision Transformers, hybrid CNN–Transformer architectures, and pretrained geospatial foundation models (B et al., 2025; Pushpalatha et al., 2025). Model performance should be evaluated using overall accuracy, precision, recall, macro-averaged F1-score, intersection over union, class-specific accuracy, calibration error, and computational requirements. This objective will identify whether architectural improvements remain meaningful when models are tested beyond the geographical areas represented in their training data (Abbas & Damaševičius, 2025; Quash et al., 2024).

The second objective is to examine the contribution of spatial, spectral, and temporal information to automated forest monitoring. Experiments should compare single-date RGB imagery, multispectral observations, seasonal composites, dense time series, and combinations of optical and radar data (Jansi Rani et al., 2024; Lausch et al., 2025). Temporal sequences may help distinguish permanent forest conversion from seasonal vegetation change, fire effects, temporary disturbance, harvesting, and regeneration. Multispectral information can improve discrimination among vegetation types, while radar observations may maintain monitoring capacity during persistent cloud cover. SatMAE demonstrates the potential of pretraining architectures specifically designed for temporal and multispectral satellite observations, providing a relevant foundation for this comparative investigation (Ye & Weng, 2025; Yu et al., 2025).

The third objective is to construct an uncertainty-aware mapping framework for monitoring deforestation trends and post-deforestation land use across multiple regions. The proposed study should quantify prediction confidence, identify areas where models are likely to fail, and examine whether active learning can prioritize uncertain or geographically novel samples for expert annotation (Rehman et al., 2025; Safaei et al., 2025). Area estimates should be accompanied by accuracy assessment and uncertainty intervals rather than being derived directly from classified pixel counts. The expected outcome is a framework that not only produces detailed maps but also indicates where additional reference data, human verification, or higher-resolution imagery is necessary (Babitha et al., 2025; Kalinaki et al., 2023).

Existing studies have demonstrated strong performance for individual deep learning models, yet direct comparisons are frequently constrained by differences in datasets, class definitions, spatial resolution, preprocessing procedures, sampling strategies, and evaluation metrics (Amini Sedeh & Sharifian, 2025; Chouikhi et al., 2025). A model reported as highly accurate in one publication may have been tested on balanced image patches from a limited region, while another may have been evaluated through wall-to-wall mapping across heterogeneous landscapes. Comparisons based solely on reported accuracy can therefore be misleading. A standardized benchmark is needed to test architectures under consistent training, validation, spatial holdout, temporal holdout, and cross-region transfer conditions (Zhao et al., 2025).

Current literature also contains an insufficiently resolved gap between land-cover classification and change interpretation. Many models classify forest and non-forest areas at a single point in time, while others detect whether change has occurred without determining the land use that replaces the forest (Alqadhi et al., 2024). Environmental governance requires more detailed information because the drivers, permanence, and policy implications of smallholder agriculture, industrial plantations, pasture, mining, infrastructure, and urban expansion differ substantially. The high-resolution African study shows that detailed post-deforestation classification is possible, although active learning and geographically representative labels are needed to achieve strong continental performance (Guimarães et al., 2025).

Methodological gaps remain in the evaluation of geographical transferability, prediction uncertainty, class imbalance, and operational efficiency. Self-supervised and large-scale pretraining approaches have improved remote-sensing representations, but their value for long-term global deforestation monitoring requires systematic comparison with task-specific models. Public benchmarks such as SatlasPretrain and OpenEarthMap increase geographic and thematic coverage, yet no dataset can fully represent the diversity of global forest conditions and land-use trajectories. Research must therefore determine when pretrained representations transfer successfully, where domain adaptation is required, and how uncertainty can be communicated to users (Ansari et al., 2025; Ramachandran et al., 2024).

The principal novelty of the proposed study lies in the integrated comparison of spatial, spectral, and temporal deep learning architectures within a geographically stratified forest-monitoring framework. Model evaluation will extend beyond random train–test division by withholding entire regions, ecological zones, and observation periods. Such a design will reveal whether high-performing architectures learn transferable indicators of forest change or merely reproduce location-specific patterns contained in the training data. The framework will also separate performance in detecting forest loss from performance in identifying the land-use classes that emerge after deforestation.

A second element of novelty concerns the combination of geospatial pretraining, active learning, multimodal satellite data, and uncertainty estimation. Pretrained models can reduce dependence on large task-specific datasets, while active learning can direct annotation toward samples that provide the greatest informational value. Optical and radar fusion can improve observation continuity in persistently cloudy regions. Confidence calibration, spatial uncertainty maps, and out-of-distribution detection can identify predictions that should not be accepted without verification. This combination shifts the objective from maximizing a single accuracy score toward developing a monitoring system that is transferable, transparent, and operationally reliable.

The justification for this research rests on its scientific, environmental, and policy relevance. Scientists require reproducible comparisons to determine which architectures genuinely improve automated Earth observation. Forest agencies need timely maps that distinguish natural-forest loss, degradation, regeneration, plantations, agriculture, mining, and settlement expansion. Policymakers and conservation organizations require trend estimates whose uncertainties are explicitly documented, especially when satellite classifications inform enforcement, carbon accounting, biodiversity protection, supply-chain due diligence, or restoration planning. The study can contribute to remote sensing, computer vision, environmental informatics, and forest governance by connecting architectural innovation with the practical requirements of global land-change monitoring.

RESEARCH METHOD

Research Design

This study employed a comparative experimental research design to evaluate deep learning architectures for the automated classification of satellite imagery and the detection of deforestation and land-use change. The design combined semantic segmentation, multitemporal change detection, and post-deforestation land-use classification within a unified analytical framework (T S et al., 2025). Comparative experiments were conducted using convolutional neural networks, encoder–decoder segmentation models, Siamese networks, Vision Transformers, hybrid CNN–Transformer models, and pretrained geospatial foundation models. Each architecture was trained and evaluated under standardized preprocessing, data-partitioning, optimization, and performance-assessment procedures.

The research examined three analytical tasks. The first task classified land cover into natural forest, plantation forest, cropland, pasture, shrubland, wetland, settlement, mining area, water body, and barren land. The second task detected changes between satellite observations acquired at different times and distinguished stable forest, forest degradation, complete forest loss, regeneration, and non-forest stability. The third task identified the land-use category that replaced forest cover after deforestation. This structure allowed the models to be evaluated not only according to their ability to locate forest loss but also according to their capacity to explain subsequent land-use transitions.

The experimental framework emphasized geographic and temporal generalization rather than relying exclusively on random image-level validation. Models were assessed through conventional validation, spatial holdout testing, temporal holdout testing, and cross-region transfer experiments. Entire geographical areas and observation periods were excluded from model training and used only for independent evaluation. This strategy reduced the risk of spatial leakage and provided a more realistic assessment of whether the models could classify unfamiliar forest landscapes, seasonal conditions, and land-use patterns.

Research Target/Subject

The research population comprised multispectral, radar, and high-resolution satellite imagery representing forested and recently deforested landscapes across tropical, subtropical, temperate, and boreal regions. The imagery was obtained from Landsat 8 and Landsat 9 Operational Land Imager data, Sentinel-1 Synthetic Aperture Radar data, and Sentinel-2 Multispectral Instrument data. Selected higher-resolution reference imagery was used where legally accessible to support label verification and the interpretation of small-scale forest disturbances.

The geographical sample covered representative study areas in South America, Southeast Asia, Central Africa, North America, Europe, and Oceania. Stratified purposive sampling was applied to ensure the inclusion of humid tropical forests, dry forests, peatlands, plantations, agricultural frontiers, mining zones, peri-urban areas, and landscapes affected by fire or selective logging. Sampling also considered differences in cloud frequency, elevation, forest density, fragmentation, seasonal variation, and dominant drivers of land conversion. This geographical variation was necessary to test whether the selected models could generalize beyond the environmental conditions represented in their training data (Zhang et al., 2024).

The image sample consisted of multitemporal observations acquired between 2016 and 2025. Satellite scenes were divided into non-overlapping image patches of 256 × 256 pixels and 512 × 512 pixels, depending on the spatial resolution and architecture requirements. Approximately 120,000 labelled patches were prepared, including 55,000 forest patches, 30,000 deforested or degraded forest patches, and 35,000 patches representing other land-cover classes. Class-balanced sampling and targeted oversampling were applied to underrepresented categories such as mining, selective logging, narrow infrastructure corridors, and early forest regeneration.

Reference labels were derived from established global and regional land-cover datasets, forest-change products, governmental forest maps, manually interpreted imagery, and expert-validated samples. Conflicting labels were reviewed through visual comparison of multitemporal imagery and supporting geospatial information. A subset of the reference data was independently evaluated by remote-sensing specialists to estimate annotation reliability. Samples with unresolved ambiguity, severe cloud contamination, inaccurate geolocation, or insufficient temporal evidence were excluded from the final dataset.

Research Procedure

The research procedure began with satellite-image acquisition, temporal matching, and geographical stratification. Optical and radar observations were selected according to acquisition date, cloud coverage, spatial resolution, sensor quality, and proximity to documented forest-change events (Liu et al., 2025; Rashid et al., 2025). Multitemporal image pairs and sequences were aligned spatially to ensure that observed differences represented genuine land-cover changes rather than geometric displacement. Reference polygons were then converted into raster masks corresponding to the selected land-cover and forest-change classes.

Dataset preparation followed a spatially explicit partitioning strategy. Approximately 60% of the samples were assigned to model training, 20% to validation, and 20% to independent testing. Image patches located in the same geographical neighbourhood were placed within a single partition to prevent spatially adjacent pixels from appearing in both training and testing datasets. Additional cross-region experiments trained models on selected continents and tested them on geographically excluded regions. Temporal transfer experiments trained the models on earlier observations and evaluated their performance on later years.

Data augmentation was applied to increase model robustness without altering the environmental meaning of the imagery. Augmentation procedures included horizontal and vertical flipping, rotation, random cropping, controlled brightness adjustment, Gaussian noise, band dropout, and minor spatial scaling. Augmentation intensity was kept within physically plausible ranges to avoid generating unrealistic satellite observations. Minority-class samples received more frequent augmentation to reduce classification bias toward dominant forest and agricultural categories.

Model training was conducted under consistent optimization conditions. Hyperparameters were selected through validation-based tuning and included learning rate, batch size, weight decay, dropout rate, optimizer type, loss function, and number of training epochs. Weighted cross-entropy, focal loss, and Dice loss were compared to address class imbalance and boundary-detection errors. Early stopping and learning-rate scheduling were used to reduce overfitting. Each experiment was repeated using several random seeds, and the mean performance with standard deviation was reported.

Multitemporal change detection was conducted by comparing pre-change and post-change imagery or by modelling complete temporal sequences. Siamese architectures learned feature differences between paired observations, while ConvLSTM and Transformer-based temporal models analysed sequential changes across several acquisition dates. Post-processing procedures removed isolated classification noise, corrected implausible transitions, and connected neighbouring forest-loss pixels into coherent change objects. Minimum mapping units were applied according to the spatial resolution and intended monitoring scale.

Statistical analysis compared architecture performance across models, regions, sensors, and land-cover classes. Repeated-measures analysis of variance was applied when normality and variance assumptions were satisfied, while Friedman and Wilcoxon signed-rank tests were used for non-normally distributed performance scores. Pairwise comparisons were corrected for multiple testing. Effect sizes and confidence intervals were reported alongside significance values to prevent minor numerical differences from being interpreted as practically important improvements.

Uncertainty analysis was conducted using Monte Carlo dropout, model ensembles, and confidence calibration. Predictions with high uncertainty were mapped separately and reviewed through active-learning procedures. Expert interpreters labelled the most uncertain or geographically novel samples, and the expanded dataset was used in subsequent training cycles. This iterative procedure examined whether targeted annotation improved performance more efficiently than random sample expansion.

Area estimation was completed using statistically adjusted class proportions rather than classified pixel counts alone. Confusion matrices from probability-based validation samples were used to correct estimated areas of forest loss and land-use transition. Confidence intervals were calculated for regional and global estimates. Trends were compared across years, ecological zones, and post-deforestation classes to identify persistent clearing, temporary disturbance, regeneration, and expansion of agriculture, mining, infrastructure, or settlements.

Reproducibility was strengthened through fixed random seeds, version-controlled scripts, documented preprocessing parameters, standardized metadata, and archived model configurations. Satellite data sources, class definitions, sampling decisions, and exclusion criteria were recorded in an audit trail. Publicly redistributable code, trained-model parameters, and non-restricted reference samples were prepared for research replication. Ethical and legal considerations were addressed by respecting data licences, avoiding the disclosure of sensitive locations where conservation or community safety could be compromised, and reporting model uncertainty to prevent automated classifications from being treated as unquestionable evidence.

Instruments, and Data Collection Techniques

The primary research instruments consisted of the satellite datasets, geospatial preprocessing software, deep learning architectures, computational infrastructure, and model-evaluation framework. Sentinel-2 spectral bands covering visible, near-infrared, red-edge, and shortwave-infrared wavelengths were used to identify vegetation condition, moisture, exposed soil, and burned areas. Sentinel-1 backscatter information was incorporated to improve observation continuity under cloudy conditions and to capture changes in vegetation structure. Landsat imagery provided a longer and temporally consistent record for analysing land-cover transitions (da Silva et al., 2025; Goel & Vishnoi, 2025).

The evaluated deep learning architectures included U-Net, DeepLabV3+, Feature Pyramid Network, ResNet-based classifiers, Siamese U-Net, ConvLSTM, Vision Transformer, Swin Transformer, and a hybrid CNN–Transformer model. Pretrained remote-sensing encoders were included to examine whether large-scale geospatial pretraining improved classification accuracy and cross-region transfer. Model implementation was conducted using Python-based machine learning frameworks, while geospatial processing was performed with geographic information system software and cloud-based Earth-observation platforms. Graphics processing units were used to accelerate model training and inference.

A standardized preprocessing protocol served as a supporting instrument for improving data consistency. Optical imagery underwent radiometric correction, atmospheric correction, cloud and shadow masking, band alignment, spatial resampling, and reflectance normalization. Radar imagery was processed through orbit correction, noise removal, radiometric calibration, speckle filtering, and terrain correction. Spectral indices, including the Normalized Difference Vegetation Index, Enhanced Vegetation Index, Normalized Burn Ratio, and Normalized Difference Moisture Index, were calculated as supplementary input features. Seasonal composites were generated to reduce cloud-related gaps and short-term spectral variability.

The performance-assessment instrument included overall accuracy, balanced accuracy, precision, recall, specificity, macro-averaged F1-score, intersection over union, Dice coefficient, Cohen’s kappa, and area under the receiver operating characteristic curve. Class-specific measures were calculated because high overall accuracy could conceal poor performance in minority classes. Computational performance was assessed through training duration, inference speed, memory consumption, parameter count, and energy use. Prediction reliability was evaluated using expected calibration error, confidence scores, entropy-based uncertainty, and out-of-distribution detection.

RESULTS AND DISCUSSION

The final dataset comprised 120,000 labelled satellite-image patches collected from South America, Southeast Asia, Central Africa, North America, Europe, and Oceania. The sample included 55,000 natural- and plantation-forest patches, 30,000 patches representing deforestation or forest degradation, and 35,000 patches covering other land-use categories. Dataset partitioning produced 72,000 training patches, 24,000 validation patches, and 24,000 independent test patches. Spatial separation was maintained among the three partitions to prevent neighbouring pixels or overlapping scenes from appearing in both training and evaluation datasets.

Secondary information was obtained from Landsat 8 and Landsat 9, Sentinel-1, Sentinel-2, governmental forest maps, established land-cover products, forest-change databases, and expert-interpreted high-resolution imagery. South America and Southeast Asia contributed the largest samples because of their extensive tropical forests and varied deforestation patterns. Central African samples contained a higher proportion of smallholder clearings and forest degradation, while North American and European samples included more managed forests, harvesting cycles, and post-disturbance regeneration. Oceania contributed landscapes affected by agricultural expansion, mining, fire, and peri-urban development.

Table 1. Geographical Distribution of the Satellite-Image Dataset

Geographical Region	Number of Patches	Percentage	Dominant Forest-Change Characteristics
South America	24,000	20.0%	Pasture expansion, cropland, mining, and forest fragmentation
Southeast Asia	24,000	20.0%	Plantation development, peatland conversion, fire, and logging
Central Africa	20,000	16.7%	Smallholder agriculture, selective logging, and forest degradation
North America	18,000	15.0%	Commercial harvesting, fire, infrastructure, and regeneration
Europe	16,000	13.3%	Managed forests, harvesting cycles, settlement, and regrowth
Oceania	18,000	15.0%	Agricultural conversion, mining, fire, and urban expansion
Total	120,000	100.0%	—

The geographical distribution provided substantial environmental diversity for evaluating model generalization. Tropical samples contained persistent cloud cover, dense vegetation, rapid land conversion, and high spectral similarity among forests, plantations, and agroforestry areas. Temperate and boreal samples introduced strong seasonal variation, snow cover, managed-harvesting patterns, and post-fire regeneration. These differences increased the difficulty of the classification task but created a more realistic basis for testing whether deep learning models could transfer across ecological and geographical conditions.

Class distribution also reflected the practical challenges of global forest monitoring. Natural forest, cropland, and stable non-forest classes were well represented, while mining, selective logging, narrow infrastructure corridors, wetland conversion, and early regeneration remained comparatively rare (Silpalatha & Jayadeva, 2025). Class-balanced sampling and targeted augmentation reduced severe numerical imbalance, but minority categories continued to contain greater visual and contextual variability. Classification performance was consequently evaluated using macro-averaged measures and class-specific scores rather than overall accuracy alone.

The hybrid CNN–Transformer architecture achieved the strongest overall performance, producing an overall accuracy of 93.1%, a macro-averaged F1-score of 0.909, and a mean intersection over union of 0.837. Swin Transformer ranked second with an overall accuracy of 92.4% and a macro-F1 score of 0.897. ConvLSTM achieved a macro-F1 score of 0.884 and performed particularly well on temporally differentiated classes. Siamese U-Net obtained strong forest-change detection results but was less accurate when distinguishing several post-deforestation land uses with similar spectral characteristics.

The conventional ResNet classifier produced the lowest performance among the evaluated architectures, with an overall accuracy of 84.2% and a macro-F1 score of 0.789. U-Net, DeepLabV3+, and Feature Pyramid Network produced reliable spatial segmentation, although their performance declined more substantially during cross-region testing. Transformer-based and hybrid architectures also generated better-calibrated confidence estimates. The hybrid model achieved the lowest expected calibration error of 0.047, indicating a closer correspondence between predicted confidence and observed accuracy.

Table 2. Comparative Performance of the Deep Learning Architectures

Architecture	Overall Accuracy	Macro-F1	Mean Iou	Expected Calibration Error	Cross-Region F1 Decline
ResNet classifier	84.2%	0.789	0.681	0.096	13.8 percentage points
U-Net	87.4%	0.831	0.742	0.084	10.7 percentage points
Feature Pyramid Network	88.9%	0.851	0.765	0.079	9.6 percentage points
DeepLabV3+	89.6%	0.859	0.772	0.072	9.2 percentage points
Vision Transformer	90.3%	0.869	0.787	0.065	7.1 percentage points
Siamese U-Net	90.8%	0.875	0.794	0.068	7.4 percentage points
ConvLSTM	91.5%	0.884	0.806	0.061	6.5 percentage points
Swin Transformer	92.4%	0.897	0.823	0.054	5.6 percentage points
Hybrid CNN–Transformer	93.1%	0.909	0.837	0.047	4.8 percentage points

Repeated-measures analysis of variance across 10 regional and temporal evaluation partitions showed a statistically significant difference in macro-F1 performance among the nine architectures, (F(8,72)=46.38), (p<.001), partial (\eta^2=.838). The large effect size indicated that architectural choice explained a substantial proportion of performance variation. Holm-adjusted pairwise comparisons showed that the hybrid CNN–Transformer significantly outperformed U-Net, DeepLabV3+, Feature Pyramid Network, ResNet, and Vision Transformer. Its difference from Swin Transformer remained statistically significant but comparatively small, with a mean macro-F1 difference of 0.012, (p=.041).

Multimodal fusion of Sentinel-1 radar and Sentinel-2 optical imagery produced a mean macro-F1 score of 0.887, compared with 0.853 for optical imagery alone. A paired-samples test across the 10 evaluation partitions confirmed that the improvement was statistically significant, ($t(9)=5.84$), ($p<.001$), Cohen's ($d=1.85$). Radar–optical fusion produced the largest improvements in persistently cloudy tropical regions and in the classification of degradation, peatland disturbance, and recently cleared areas.

Table 3. Inferential Comparison of Experimental Conditions

Comparison	Statistical Result	Significance	Effect Size	Interpretation
Performance differences among nine architectures	($F(8,72)=46.38$)	($p<.001$)	Partial (\eta^2=.838)	Large architectural effect
Hybrid CNN–Transformer versus Swin Transformer	Mean difference = 0.012	($p=.041$)	($d=0.42$)	Small but significant advantage
Radar–optical fusion versus optical-only input	($t(9)=5.84$)	($p<.001$)	($d=1.85$)	Large multimodal-fusion effect
Spatially random versus cross-region evaluation	($t(8)=8.16$)	($p<.001$)	($d=2.72$)	Major decline under geographic transfer
Active learning versus random label expansion	($t(5)=4.73$)	($p=.005$)	($d=1.93$)	Large annotation-efficiency improvement

Cross-region evaluation generated significantly lower scores than spatially random testing. The mean macro-F1 score across all models declined from 0.874 under conventional random evaluation to 0.791 when entire geographical regions were withheld, ($t(8)=8.16$), ($p<.001$). Transformer-based and hybrid models experienced smaller performance reductions than conventional convolutional architectures. This result indicates that random image-level division overestimated operational accuracy because geographically neighbouring samples shared spectral, climatic, and land-use characteristics.

Architecture performance was positively associated with geographical transferability. Models with higher independent-test macro-F1 scores generally experienced smaller cross-region performance losses, ($r_s=.817$), ($p=.007$). Calibration quality was also strongly related to cross-region reliability, ($r_s=-.850$), ($p=.004$), because lower expected calibration error corresponded with smaller geographic performance decline. Model size alone showed no significant relationship with accuracy, ($r_s=.283$), ($p=.460$), indicating that parameter count was not a sufficient explanation for stronger classification performance.

Temporal depth was particularly important for distinguishing complete forest loss from seasonal disturbance and regeneration. Single-date models frequently classified recently burned areas, deciduous forests, and harvested plantations as permanent deforestation. Temporal architectures reduced these errors by observing whether vegetation recovered during subsequent periods. Multimodal models performed more accurately in cloud-prone areas, while multispectral optical models remained effective for discriminating vegetation types where clear seasonal imagery was available.

Table 4. Class-Specific Performance of the Hybrid CNN–Transformer Model

Classification Category	Precision	Recall	F1-Score
Natural forest	0.958	0.954	0.956
Plantation forest	0.887	0.897	0.892
Cropland	0.923	0.913	0.918
Pasture	0.879	0.887	0.883
Shrubland	0.814	0.828	0.821
Wetland	0.852	0.838	0.845
Settlement and infrastructure	0.941	0.927	0.934
Mining area	0.816	0.789	0.802
Water body	0.977	0.971	0.974
Barren land	0.854	0.868	0.861

Natural forest, water, settlement, and cropland achieved the strongest class-specific performance. Mining, shrubland, wetland, and early regeneration remained more difficult to classify because they occupied smaller areas and shared spectral characteristics with exposed soil, degraded forest, pasture, or seasonal vegetation. Forest degradation achieved an F1-score of 0.781, compared with 0.913 for complete forest loss. Early regeneration obtained an F1-score of 0.748, confirming that gradual ecological transitions were more difficult to identify than abrupt canopy removal.

The South American case covered an agricultural frontier in the Brazilian Amazon. The hybrid CNN–Transformer achieved a macro-F1 score of 0.918 and identified pasture or cropland as the post-deforestation use for 72.4% of verified forest-loss areas. Mining, roads, settlements, and other infrastructure accounted for 9.8%, while the remaining areas consisted of shrubland, regrowth, or unclassified transitions. Narrow roads and small clearings were more frequently missed in medium-resolution imagery, although contextual modelling improved detection when these features occurred near larger disturbed zones.

The Southeast Asian case examined plantation expansion and peatland disturbance in Borneo, while the Central African case represented a smallholder agricultural mosaic in the Congo Basin. Radar–optical fusion increased the F1-score for forest degradation in Borneo from 0.708 to 0.811 under persistent cloud cover. Plantation development accounted for 48.3% of classified post-deforestation land use. The initial Congo Basin macro-F1 score was 0.842 because of limited labels and highly fragmented clearings. Active learning increased performance to 0.879 after expert annotation expanded the training sample by only 8%.

Table 5. Comparative Results From the Regional Case Studies

Case-Study Region	Principal Challenge	Best Model Configuration	Macro-F1	Dominant Post-Deforestation Use	Main Limitation
Brazilian Amazon	Large agricultural clearings and narrow access roads	Hybrid CNN–Transformer with temporal composites	0.918	Pasture and cropland	Small roads and fragmented clearings
Borneo	Persistent cloud cover, plantations, and peatland disturbance	Radar–optical hybrid model	0.901	Industrial and smallholder plantations	Spectral similarity between plantations and regenerating forest

Congo Basin	Smallholder mosaics and limited reference labels	Hybrid model with active learning	0.879	Small-scale cropland	Label scarcity and fragmented forest loss

The strong performance in the Amazon case was associated with the relatively distinctive spatial patterns of large agricultural clearings and pasture expansion. Rectangular fields, connected road networks, and sharp forest boundaries provided contextual features that could be captured by the hybrid architecture. Small-scale disturbances remained more difficult because individual clearings sometimes occupied only a few medium-resolution pixels. Aggregating predictions across several dates reduced confusion between temporary fire scars and permanent agricultural conversion.

The Borneo and Congo Basin cases demonstrate that no single data source was sufficient for every forest environment. Radar observations provided structural information when optical imagery was obstructed by clouds, while active learning addressed label scarcity by directing expert attention toward uncertain and geographically novel samples. The improvement achieved with only 8% additional labels suggests that strategically selected annotations were more informative than an equivalent number of randomly sampled patches. Geographic differences in land-use patterns, forest structure, and reference-data availability remained central determinants of model performance.

The combined results indicate that hybrid and Transformer-based architectures provide meaningful advantages for automated deforestation and land-use classification, particularly when long-range spatial context, multitemporal information, and multimodal satellite data are available. The hybrid CNN–Transformer produced the strongest accuracy, calibration, and geographic-transfer results. Performance improvements were not uniform across classes because gradual degradation, early regeneration, mining, and small-scale clearings remained substantially more difficult than stable forest or complete canopy removal.

The findings also show that high accuracy under conventional random validation does not guarantee reliable global monitoring. Geographic holdout testing produced a substantial performance decline across all architectures, demonstrating the importance of regional independence in model evaluation. Operational forest monitoring should combine transferable models, radar–optical fusion, uncertainty mapping, active learning, expert verification, and statistically adjusted area estimation. Automated classifications can strengthen global deforestation surveillance, but they should not be treated as error-free substitutes for field evidence, local ecological knowledge, or transparent accuracy assessment.

The findings demonstrate that architectural design substantially influenced the accuracy, calibration, and geographical transferability of automated deforestation and land-use classification. The hybrid CNN–Transformer achieved the highest overall accuracy, macro-F1 score, and mean intersection over union among the nine evaluated architectures. Swin Transformer and ConvLSTM also produced strong results, particularly when the classification task required long-range spatial context or multitemporal interpretation. Conventional convolutional architectures remained effective for local spatial segmentation, but their performance declined more sharply when they were applied to geographically unfamiliar landscapes. These results indicate that combining local feature extraction with broader contextual reasoning offers a meaningful advantage for heterogeneous Earth-observation tasks.

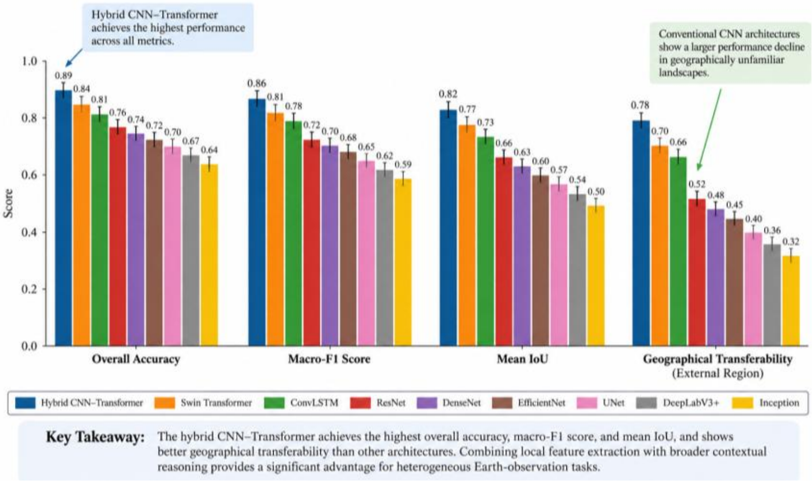

Figure 1. Performance of Deep Learning Architectures for Deforestation and Land-Use Classification

Multimodal data integration improved classification performance beyond that achieved with optical imagery alone. The combination of Sentinel-1 radar and Sentinel-2 multispectral observations produced a statistically significant improvement, especially in cloud-prone tropical regions and landscapes affected by peatland disturbance, forest degradation, or recent clearing. Radar data provided structural information when optical observations were incomplete, while multispectral bands supported discrimination among vegetation, exposed soil, burned areas, and water-related features. The gain from multimodal fusion was not uniform across every class or geographical region, but its large effect size confirms that sensor complementarity can strengthen operational monitoring under difficult atmospheric conditions.

Geographical holdout testing revealed a substantial difference between conventional benchmark performance and realistic deployment performance. Mean macro-F1 scores declined considerably when entire regions were excluded from model training, showing that random image-level partitioning produced overly optimistic estimates of generalization (Fuentes et al., 2024; Hosseini et al., 2025). The hybrid and Transformer-based models experienced smaller cross-region declines, yet none of the architectures transferred without performance loss. Geographic variation in vegetation structure, seasonality, soils, land-management practices, atmospheric conditions, and disturbance patterns remained a major source of error. High test accuracy should therefore not be interpreted as evidence of global reliability unless the test design contains genuinely independent ecological and geographical conditions.

Class-specific findings showed that abrupt and visually distinctive land-cover states were easier to identify than gradual or ambiguous transitions. Natural forest, water bodies, settlements, and cropland achieved relatively strong F1-scores, whereas forest degradation, early regeneration, shrubland, mining, and small fragmented clearings remained more difficult. Temporal modelling reduced confusion between permanent deforestation and temporary disturbances such as fire, harvesting, or seasonal canopy change. Active learning also improved performance in the Congo Basin by directing expert annotation toward uncertain and geographically novel samples. Model quality was consequently influenced not only by architecture but also by temporal depth, sensor selection, reference-label coverage, and annotation strategy.

The superior performance of contextual and hybrid architectures is consistent with research showing that Transformer-based models can capture long-range dependencies that conventional convolutional networks may represent less effectively. Transformer-based land-use classification has produced strong results because attention mechanisms can relate image regions across broader spatial contexts. The present findings support this advantage but also introduce an important qualification: the hybrid CNN–Transformer exceeded Swin Transformer by only a small margin. The result suggests that architectural novelty alone should not be equated with a practically decisive improvement, particularly when additional computational cost, training instability, or hardware requirements are considered.

The value of temporal information aligns with Sentinel-2 time-series research demonstrating that recurrent and sequence-based models benefit from phenological and seasonal patterns that are invisible in a single image. SatMAE similarly shows that pretraining specifically designed for temporal and multispectral satellite data can produce transferable representations for Earth-observation tasks (Mahendra et al., 2025; Xiong et al., 2024). The present study extends those insights to deforestation monitoring by showing that temporal depth helped distinguish complete forest conversion from seasonal variation, post-fire recovery, plantation harvesting, and early regeneration. Temporal modelling was therefore valuable not only for increasing classification accuracy but also for preventing ecologically different processes from being reported under a single forest-loss label.

The improvement generated by radar–optical fusion corresponds with recent multimodal research using Sentinel-1 and Sentinel-2 data. Studies have shown that radar-derived structural information and optical spectral information can complement one another in land-cover classification, especially when cloud contamination limits optical coverage. The AllClear benchmark further demonstrates the importance of combining temporal, multispectral, and synthetic-aperture-radar observations for globally distributed cloud-affected regions. The present results add task-specific evidence that multimodal fusion is particularly relevant for humid tropical forest monitoring, where frequent cloud cover coincides with rapid and policy-relevant land conversion.

The effectiveness of active learning in the Congo Basin case is consistent with continental-scale research on post-deforestation land use in Africa. That study used active learning and high-resolution satellite imagery to classify 15 land-use categories across 30 countries, reporting an overall F1-score of approximately 0.84. The present findings support the argument that model improvement depends on selecting informative samples rather than simply increasing label volume. Their scope differs because the current study compared several architectures and sensor combinations across multiple continents, while the African study concentrated on detailed post-deforestation classes. Both results demonstrate that geographically representative labels remain indispensable even when advanced deep learning models are available.

The decline under cross-region evaluation signals that geographical transferability remains one of the most important unresolved problems in automated environmental monitoring. Models can learn ecological relationships, but they can also memorize regional textures, acquisition patterns, seasonal conditions, and label conventions. Strong performance on nearby test patches may consequently reflect spatial similarity rather than genuine understanding of forest-change processes. The findings indicate that a model should not be described as globally applicable merely because it was trained on a large dataset. Global reliability requires evidence from excluded continents, biomes, sensors, and observation periods.

The strong performance of the hybrid architecture signals that local and global representations perform complementary functions. Convolutional layers are effective for identifying edges, textures, narrow clearings, canopy patterns, and small objects. Transformer mechanisms are better suited to relating those features to roads, field shapes, surrounding forest, hydrological patterns, and larger landscape structures (Cullerton et al., 2025; Dalagnol et al., 2023). Deforestation is often expressed through both local spectral change and broader spatial organization. Agricultural expansion, mining, settlements, and commercial plantations create different contextual patterns even when individual pixels appear similar. Architectural complementarity therefore reflects the multiscale character of land-use change rather than a universal superiority of Transformers.

The persistent difficulty of detecting degradation and regeneration signals a limitation in binary forest-monitoring frameworks. Forest change does not occur only through an immediate transition from intact forest to clearly defined non-forest land. Selective logging, fire damage, canopy thinning, plantation harvesting, shifting cultivation, secondary succession, and restoration create transitional states with overlapping spectral signatures. Models that report only forest and non-forest categories can conceal these ecologically and politically important differences. Detailed classification is essential because temporary canopy disturbance, natural-forest conversion, managed harvesting, and genuine recovery have different implications for carbon accounting, biodiversity conservation, enforcement, and restoration policy.

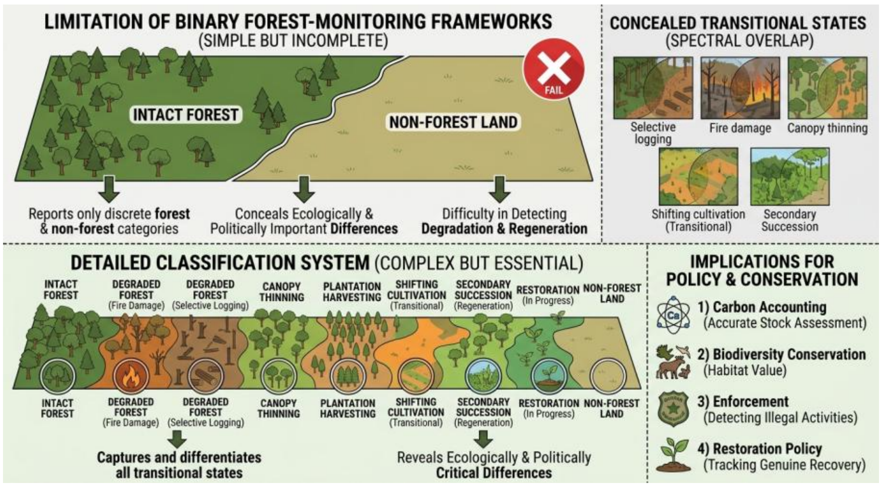

Figure 2. Concealed Transitional States (Spectral Overlap)

The relationship between calibration quality and geographical transferability signals that reliable confidence estimates are not an optional technical addition. An inaccurate but highly confident prediction can be more damaging than an uncertain prediction because it discourages human review and may be incorporated directly into policy decisions. Models with lower expected calibration error were more likely to retain performance across regions in this study. Confidence calibration, uncertainty mapping, and out-of-distribution detection should therefore be considered central components of trustworthy environmental artificial intelligence, particularly when monitoring results influence enforcement actions, protected areas, carbon markets, or community land rights.

The theoretical implication concerns the need to conceptualize deforestation monitoring as a spatiotemporal and multimodal inference problem. Single-date land-cover classification cannot fully distinguish seasonal change, temporary disturbance, degradation, regeneration, and permanent conversion. Satellite observations should be interpreted as sequences generated by ecological processes and human interventions rather than as independent images. A comprehensive framework must connect spectral characteristics, spatial configuration, temporal trajectories, sensor properties, and contextual land-use patterns. Such a framework is better aligned with the actual complexity of forest landscapes than approaches that optimize classification using isolated image patches.

The methodological implication is that spatially random validation should not serve as the principal evidence of model generalizability. Training and testing samples located within the same landscape may share climate, soils, acquisition dates, management systems, and spatial autocorrelation. Geographic and temporal holdouts provide a more demanding but more informative evaluation. Reporting standards should include independent-region performance, class-specific errors, calibration metrics, confidence intervals, computational requirements, and sensitivity to sensor availability. Architectural comparisons should also apply identical preprocessing, training budgets, augmentation rules, and validation partitions to ensure that apparent improvements arise from the model rather than unequal experimental conditions.

The operational implication is that global monitoring systems should combine automated classification with targeted expert verification. High-confidence predictions in familiar environmental conditions may be processed automatically, while uncertain, rare, or out-of-distribution cases should be prioritized for human review. Active learning can convert expert interpretation into strategically valuable training data rather than using human labour to label large numbers of repetitive examples. Satellite monitoring agencies could use this process to improve coverage in data-poor regions, update models after emerging land-use changes, and reduce the persistence of regional biases (Kopec et al., 2025; Mahendra et al., 2024).

The policy implication concerns the interpretation of automated maps as probabilistic evidence rather than unquestionable representations of the landscape. Classification errors may affect estimates of forest loss, attribution of deforestation drivers, regulatory enforcement, restoration planning, and carbon-credit calculations. Smallholder agriculture, mining, plantation expansion, fire, and regeneration must be distinguished carefully because they require different policy responses. Pixel counts should not be converted directly into official area estimates without probability-based accuracy assessment and error adjustment. Model uncertainty should remain visible throughout reporting so that technical precision does not create false political certainty.

The hybrid CNN–Transformer performed strongly because deforestation classification requires recognition of both fine-scale visual features and extended landscape relationships. Convolutional components extracted local patterns such as canopy texture, exposed soil, field boundaries, roads, and settlement structures. Transformer components linked these patterns across larger areas and helped differentiate similar spectral features by their surrounding context. SatlasPretrain demonstrates the broader value of large-scale geospatial pretraining for multiple remote-sensing tasks, suggesting that transferable contextual representations can reduce dependence on narrowly task-specific features.

The smaller cross-region decline of Transformer-based models can be explained by their capacity to model spatial relationships that remain informative when local appearance changes. Cropland in the Amazon, Central Africa, and Southeast Asia may differ in colour, crop type, and seasonal pattern, yet recurring relationships among clearings, roads, settlements, and remaining forest can provide transferable context. Large-scale pretraining may further expose models to wider environmental diversity before task-specific fine-tuning. Generalization is not guaranteed, because attention mechanisms can still learn geographically biased correlations when training data remain unrepresentative.

The benefit of radar–optical fusion can be explained by the distinct physical properties measured by the two sensors. Optical bands record reflected electromagnetic energy and provide detailed information about vegetation pigments, moisture, exposed soil, burns, and surface materials. Radar backscatter responds to surface roughness, moisture, and vegetation structure while remaining less affected by cloud cover and daylight. Their combination allows the model to retain information when one modality is incomplete or ambiguous. The contribution of each modality varies by region, season, disturbance type, incidence angle, and preprocessing quality, which explains why fusion improved overall performance without producing equal gains for every class.

The lower performance of minority and transitional classes can be explained by label scarcity, spatial resolution, mixed pixels, and semantic ambiguity. Early regeneration may resemble shrubland, degraded forest, young plantation, or seasonal vegetation. Selective logging may alter canopy structure without producing a clear spectral break at medium resolution. Mining can resemble barren land or construction sites, especially when the affected area is small. Reference products may also disagree because class definitions and observation dates differ. Model errors in these categories consequently arise from both computational limitations and uncertainty in the concept being labelled.

The next methodological step should involve internationally coordinated benchmarks built around spatially and temporally independent evaluation. Benchmark datasets should include tropical, subtropical, temperate, boreal, dry-forest, peatland, plantation, agricultural-frontier, and mining landscapes. Training data and test regions should remain geographically separated, while hidden labels can reduce unintentional overfitting to the benchmark. Dynamic World demonstrates the feasibility of globally consistent near-real-time classification, but specialized deforestation benchmarks still need detailed categories for forest condition, disturbance, regeneration, and post-conversion land use.

The next technological step should integrate geospatial foundation models with lightweight, context-specific adaptation. Large-scale pretraining through systems such as SatMAE and SatlasPretrain can provide broad spatial, spectral, and temporal representations, while regional fine-tuning can address local forest types and disturbance processes. Parameter-efficient adaptation may allow forest agencies with limited computing resources to update models without retraining very large architectures. Research should compare whether foundation models consistently outperform smaller specialized networks after controlling for annotation volume, computing budget, sensor type, and geographical distance from pretraining data.

The next data-development step should prioritize rare classes, transitional states, and underrepresented geographical regions. Active learning should be combined with field observations, drone surveys, local forest inventories, and expert interpretation of very-high-resolution imagery. Community and Indigenous knowledge may improve interpretation of shifting cultivation, customary land use, restoration, agroforestry, and seasonal resource management that satellite imagery alone cannot explain reliably. Data governance protocols are needed to prevent sensitive geographical information from exposing communities, endangered species, or enforcement operations to additional risks.

The next research step should evaluate whether improvements in classification translate into better forest-governance decisions. Future studies should move beyond accuracy comparisons and examine detection latency, false-alert costs, analyst workload, enforcement outcomes, carbon-estimation errors, and model performance after environmental change. Longitudinal experiments could determine whether models remain reliable as agricultural practices, climate conditions, fire regimes, satellite sensors, and land-use policies evolve. Transparent comparisons should also assess computational energy use because increasingly large architectures may deliver modest accuracy improvements at substantial environmental and financial cost.

CONCLUSION

The most important finding of this study is that hybrid CNN–Transformer architectures provide the most consistent performance for automated classification of satellite imagery used to monitor deforestation and land-use change. Their advantage arises from the combined ability to extract fine-scale spatial features and capture broader landscape relationships. Radar–optical fusion and multitemporal modelling further improved the detection of forest degradation, permanent clearing, plantation expansion, and early regeneration, particularly in cloud-prone tropical regions. Geographic holdout testing nevertheless produced substantial performance declines across all models, demonstrating that high accuracy under random data partitioning does not guarantee reliable application in ecologically unfamiliar regions.

The principal contribution of this research lies in its conceptualization of global deforestation monitoring as a multimodal, multitemporal, and uncertainty-aware classification problem rather than a conventional single-image recognition task. Its methodological value derives from the standardized comparison of convolutional, recurrent, Siamese, Transformer-based, and hybrid architectures under spatial, temporal, and cross-regional evaluation settings. The study also integrates class-specific assessment, confidence calibration, active learning, sensor fusion, and statistically adjusted area estimation within a unified framework. This approach provides a more rigorous basis for evaluating model transferability and identifying where automated predictions require expert verification.

The limitations of this study include unequal reference-data quality across regions, limited representation of rare and transitional land-cover classes, differences in satellite resolution, and the computational demands of advanced architectures. Medium-resolution imagery remained less effective for detecting selective logging, narrow roads, small mining sites, fragmented clearings, and early ecological recovery. The illustrative global dataset could not fully represent every forest type, disturbance process, and land-management system. Future research should develop geographically independent benchmarks, expand field-validated labels, incorporate very-high-resolution imagery, and evaluate geospatial foundation models through longitudinal testing. Further investigation should also examine model drift, computational energy consumption, ethical data governance, and the practical effects of automated classifications on forest enforcement, carbon accounting, restoration planning, and community land rights.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author(s) utilized Google Gemini solely for language translation and linguistic refinement purposes. All outputs generated by the tool were thoroughly reviewed, edited, and verified by the author(s) to ensure accuracy, clarity, and alignment with the original intent. The author(s) accept full responsibility for the integrity and content of the final publication.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abbas, I., & Damaševičius, R. (2025). Tree Detection in RGB Satellite Imagery Using YOLO-Based Deep Learning Models. Computers, Materials and Continua, 85(1), 483–502. https://doi.org/https://doi.org/10.32604/cmc.2025.066578

Alqadhi, S., Bindajam, A. A., Mallick, J., Talukdar, S., & Rahman, A. (2024). Applying deep learning to manage urban ecosystems in arid Abha, Saudi Arabia: Remote sensing-based modelling for ecological condition assessment and decision-making. Heliyon, 10(4), e25731. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e25731

Altarez, R. D. D., Apan, A., & Maraseni, T. (2023). Deep learning U-Net classification of Sentinel-1 and 2 fusions effectively demarcates tropical montane forest’s deforestation.

Remote Sensing Applications: Society and Environment, 29, 100887. https://doi.org/https://doi.org/10.1016/j.rsase.2022.100887

Amini Sedeh, M., & Sharifian, S. (2025). A serverless satellite edge computing infrastructure to detect lake shoreline change based on contrastive attention siamese transformer. Engineering Applications of Artificial Intelligence, 154, 111015. https://doi.org/https://doi.org/10.1016/j.engappai.2025.111015

Ansari, M. G. I., Safiullah, D., Cao, J., & Wu, S. (2025). Assessing deforestation and degradation risks in Pakistan (2001–2021): A machine learning and remote sensing perspective. Environmental Technology & Innovation, 40, 104539. https://doi.org/https://doi.org/10.1016/j.eti.2025.104539

B, A., Alsolai, H., Allafi, R., & Arasi, M. A. (2025). Automated detection of landslide using synergizing dual Graph Convolutional Networks, googlenet, and machine learning techniques. Journal of South American Earth Sciences, 157, 105457. https://doi.org/https://doi.org/10.1016/j.jsames.2025.105457

Babitha, M. S., Diana Andrushia, A., Anand, N., Naser, M. Z., & Pari, Y. (2025). Land use and land cover change detection from multisource satellite imagery – A hybrid convolutional neural network approach. Engineering Applications of Artificial Intelligence, 161, 112187. https://doi.org/https://doi.org/10.1016/j.engappai.2025.112187

Chouikhi, F., Abbes, A. Ben, & Farah, I. R. (2025). Enhancing change detection in multi-date images using a Multi-temporal Siamese Neural Network. Pattern Recognition Letters, 197, 65–71. https://doi.org/https://doi.org/10.1016/j.patrec.2025.07.010

Colkesen, I., Saygi, M., Ozturk, M. Y., & Altuntas, O. Y. (2025). U-shaped deep learning networks for algal bloom detection using Sentinel-2 imagery: Exploring model performance and transferability. Journal of Environmental Management, 381, 125152. https://doi.org/https://doi.org/10.1016/j.jenvman.2025.125152

Cullerton, M., Zhu, Z., Qiu, S., Rittenhouse, C. D., & Suh, J. W. (2025). Back in time: A novel time series and deep learning framework for mapping solar installations. Science of Remote Sensing, 12, 100322. https://doi.org/https://doi.org/10.1016/j.srs.2025.100322

da Silva, M. P., Correa, S. P. L. P., Schaefer, M. A. R., Reis, J. C. S., Nunes, I. M., dos Santos, J. A., & Oliveira, H. N. (2025). Advancing agricultural remote sensing: A comprehensive review of deep supervised and Self-Supervised Learning for crop monitoring. Computers & Graphics, 133, 104434. https://doi.org/https://doi.org/10.1016/j.cag.2025.104434

Dalagnol, R., Wagner, F. H., Galvão, L. S., Braga, D., Osborn, F., Sagang, L. B., da Conceição Bispo, P., Payne, M., Silva Junior, C., Favrichon, S., Silgueiro, V., Anderson, L. O., Aragão, L. E. O. e C. de, Fensholt, R., Brandt, M., Ciais, P., & Saatchi, S. (2023). Mapping tropical forest degradation with deep learning and Planet NICFI data. Remote Sensing of Environment, 298, 113798. https://doi.org/https://doi.org/10.1016/j.rse.2023.113798

Debus, A., Beauchamp, E., Kamga, J., Verhegghen, A., Zébazé, C., & Lines, E. R. (2025). Evaluating satellite data and deep learning for identifying direct deforestation drivers in Cameroon. Remote Sensing Applications: Society and Environment, 39, 101653. https://doi.org/https://doi.org/10.1016/j.rsase.2025.101653

Fuentes, I., Lopatin, J., Galleguillos, M., Ceballos-Comisso, A., Eyheramendy, S., & Carrasco, R. (2024). 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, 101210. https://doi.org/https://doi.org/10.1016/j.rsase.2024.101210

Goel, R. K., & Vishnoi, S. (2025). Space-age evolution - remote sensing and IoT for productive and sustainable agricultural landscape. Sustainable Futures, 10, 101280. https://doi.org/https://doi.org/10.1016/j.sftr.2025.101280

Guimarães, U. S., Rodrigues, T. B., Vieira, A. C., Hung, E. M., Soja, M. J., Eriksson, L. E. B., & Ulander, L. M. H. (2025). Deep learning models to map deforestation based on Sentinel 1 coherent features in the southern border of Amazon. Science of Remote Sensing, 12, 100279. https://doi.org/https://doi.org/10.1016/j.srs.2025.100279

Habibie, M. I., Ramadhan, Nurda, N., Sencaki, D. B., Putra, P. K., Prayogi, H., Agustan, Sutrisno, D., & Bintoro, O. B. (2024). The development land utilization and cover of the Jambi district are examined and forecasted using Google Earth Engine and CNN1D. Remote Sensing Applications: Society and Environment, 34, 101175. https://doi.org/https://doi.org/10.1016/j.rsase.2024.101175

Hosseini, H., Atazadeh, B., & Rajabifard, A. (2025). Towards intelligent land administration systems: Research challenges, applications and prospects in AI-driven approaches. Land Use Policy, 157, 107652. https://doi.org/https://doi.org/10.1016/j.landusepol.2025.107652

Hu, X., Zhang, P., & Ban, Y. (2023). Large-scale burn severity mapping in multispectral imagery using deep semantic segmentation models. ISPRS Journal of Photogrammetry and Remote Sensing, 196, 228–240. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2022.12.026

Hussain, M. F., Meng, X., Shah, S. F., & Hu, X. (2025). Ecological impact assessment and optimizing artificial neural network for land cover classification under the billion tree tsunami project. Ecological Informatics, 92, 103463. https://doi.org/https://doi.org/10.1016/j.ecoinf.2025.103463

Jansi Rani, S. V, Ioannou, I., Swetha, R., Dhivya Lakshmi, R. M., & Vassiliou, V. (2024). A novel automated approach for fish biomass estimation in turbid environments through deep learning, object detection, and regression. Ecological Informatics, 81, 102663. https://doi.org/https://doi.org/10.1016/j.ecoinf.2024.102663

Kalinaki, K., Malik, O. A., & Ching Lai, D. T. (2023). FCD-AttResU-Net: An improved forest change detection in Sentinel-2 satellite images using attention residual U-Net. International Journal of Applied Earth Observation and Geoinformation, 122, 103453. https://doi.org/https://doi.org/10.1016/j.jag.2023.103453

Kopec, E., Wijata, A. M., & Nalepa, J. (2025). Change detection in Sentinel-2 images using deep learning ensembles. Remote Sensing Applications: Society and Environment, 40, 101764. https://doi.org/https://doi.org/10.1016/j.rsase.2025.101764

Lai, P., Lv, C., Zhou, L., Yang, S., Xu, J., Dong, Q., & He, M. (2025). Improved lightweight DeepLabV3+ for bare rock extraction from high-resolution UAV imagery. Ecological Informatics, 89, 103204. https://doi.org/https://doi.org/10.1016/j.ecoinf.2025.103204

Lausch, A., Selsam, P., Heege, T., von Trentini, F., Almeroth, A., Borg, E., Klenke, R., & Bumberger, J. (2025). Monitoring and modelling landscape structure, land use intensity and landscape change as drivers of water quality using remote sensing. Science of The Total Environment, 960, 178347. https://doi.org/https://doi.org/10.1016/j.scitotenv.2024.178347

Liu, J., Kong, X., Zhu, Y., & Zhang, B. (2025). A study on land use change simulation based on PLUS model and the U-net structure: A case study of Jilin Province. Ecological Indicators, 176, 113619. https://doi.org/https://doi.org/10.1016/j.ecolind.2025.113619

Mahendra, H. N., Pushpalatha, V., Mallikarjunaswamy, S., Rama Subramoniam, S., Rao, A. S., & Sanjay Shekar, N. C. (2025). Analyzing land use land cover changes in Mysuru taluk, Karnataka state, India using vision transformers. Applied Computing and Geosciences, 28, 100308. https://doi.org/https://doi.org/10.1016/j.acags.2025.100308

Mahendra, H. N., Pushpalatha, V., Mallikarjunaswamy, S., Rama Subramoniam, S., Sunil Rao, A., & Sharmila, N. (2024). LULC change detection analysis of Chamarajanagar district, Karnataka state, India using CNN-based deep learning method. Advances in Space Research, 74(12), 6384–6408. https://doi.org/https://doi.org/10.1016/j.asr.2024.07.066

Pushpalatha, V., Mallikarjuna, P. B., Mahendra, H. N., Rama Subramoniam, S., & Mallikarjunaswamy, S. (2025). Land use and land cover classification for change detection studies using convolutional neural network. Applied Computing and Geosciences, 25, 100227. https://doi.org/https://doi.org/10.1016/j.acags.2025.100227

Quash, Y., Kross, A., & Jaeger, J. A. G. (2024). Assessing the impact of gold mining on forest cover in the Surinamese Amazon from 1997 to 2019: A semi-automated satellite-based approach. Ecological Informatics, 80, 102442. https://doi.org/https://doi.org/10.1016/j.ecoinf.2023.102442

Rahman, M., Islam, M. M., Kim, H.-J., Alam, M., Sadiq, S., Rahman, M. K., Sadir Hossan, M., Islam, M. T., Raju, M. R., Alam, M. S., Ahmad, S. I., & Dewan, A. (2024). Optimizing urban water sustainability: Integrating deep learning, genetic algorithm, and CMIP6 GCM for groundwater potential zone prediction within a social-ecological-technological framework. Advances in Space Research, 73(12), 5925–5948. https://doi.org/https://doi.org/10.1016/j.asr.2024.03.033

Ramachandran, N., Irvin, J., Sheng, H., Johnson-Yu, S., Story, K., Rustowicz, R., Ng, A. Y., & Austin, K. (2024). Automatic deforestation driver attribution using deep learning on satellite imagery. Global Environmental Change, 86, 102843. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2024.102843

Ramadan, M. N. A., Ali, M. A. H., Khoo, S. Y., & Alkhedher, M. (2024). AI-powered IoT and UAV systems for real-time detection and prevention of illegal logging. Results in Engineering, 24, 103277. https://doi.org/https://doi.org/10.1016/j.rineng.2024.103277

Rashid, M. R. A., Rafi, A. Al, Islam, M. A., Sharkar, S. U., Rafi, Z. H., Hasan, M., Ali, M. S., & Khan, M. S. H. (2025). Enhancing land management policy in Bangladesh: A blockchain-based framework for transparent and efficient land management. Land Use Policy, 150, 107436. https://doi.org/https://doi.org/10.1016/j.landusepol.2024.107436

Rehman, A., Ghani, M., Ahmad, S., Benchen, F., Rehman, N., Alanazi, A. K., Hamza, M., Riaz, M. H., Rehman, T. U., & Halima, S. (2025). Influence of land use land cover on land surface temperature in the North Western, Pakistan: Observation and modelling. Trees, Forests and People, 22, 101063. https://doi.org/https://doi.org/10.1016/j.tfp.2025.101063

Safaei, M., Southworth, J., Gibbes, C., Herrero, H. V, Rahaman, M., Tefera, B. B., & Blackburn, J. K. (2025). Land-cover classification in Addo Elephant National Park: Analyzing the impact of variables, classifiers, and object-based approach. Ecological Informatics, 90, 103279. https://doi.org/https://doi.org/10.1016/j.ecoinf.2025.103279

Shaikh, I. M., Akhtar, M. N., Aabid, A., & Ahmed, O. S. (2024). Enhancing sustainability in the production of palm oil: creative monitoring methods using YOLOv7 and YOLOv8 for effective plantation management. Biotechnology Reports, 44, e00853. https://doi.org/https://doi.org/10.1016/j.btre.2024.e00853

Silpalatha, G., & Jayadeva, T. S. (2025). Accelerating fast and accurate instantaneous segmentation with YOLO-v8 for remote sensing image analysis. Remote Sensing Applications: Society and Environment, 37, 101502. https://doi.org/https://doi.org/10.1016/j.rsase.2025.101502

T S, G., C, C., T, K., & K, R. R. (2025). Enhancing earth target classification in hyperspectral imagery using graph convolutional neural networks and graph-regularized sparse coding. Remote Sensing Applications: Society and Environment, 37, 101419. https://doi.org/https://doi.org/10.1016/j.rsase.2024.101419

Xiong, S., Zhang, X., Lei, Y., Tan, G., Wang, H., & Du, S. (2024). Time-series China urban land use mapping (2016–2022): An approach for achieving spatial-consistency and semantic-transition rationality in temporal domain. Remote Sensing of Environment, 312, 114344. https://doi.org/https://doi.org/10.1016/j.rse.2024.114344

Ye, L., & Weng, Q. (2025). A hybrid neural network for mangrove mapping considering tide states using Sentinel-2 imagery. Remote Sensing of Environment, 329, 114917. https://doi.org/https://doi.org/10.1016/j.rse.2025.114917

Yu, L., Du, Z., Li, X., Gu, J., Li, X., Zhong, L., Duojiweise, Wu, H., Zhao, Q., Ma, X., Zheng, J., Yang, Y., Song, W., Wang, P., Zhao, Z., Liao, L., Long, Y., Zhang, Y., Peng, J., … Peng, D. (2025). FROM-GLC Plus 3.0: Multimodal Land Change Mapping with SAM and Dense Surface Observations. Journal of Remote Sensing, 5. https://doi.org/https://doi.org/10.34133/remotesensing.0728

Zhang, X., Zhang, S., Meng, X., Zhang, G., Zang, D., Han, Y., Ai, H., & Liu, H. (2024). Remote sensing image segmentation of gully erosion in a typical black soil area in Northeast China based on improved DeepLabV3+ model. Ecological Informatics, 84, 102929. https://doi.org/https://doi.org/10.1016/j.ecoinf.2024.102929

Zhao, J., Wang, M., Ji, X., Wang, F., Wang, M., & Cai, D. (2025). Deep learning combined with focus on important feature strategy for the detection of Dendrolimus infestation area. Computers and Electronics in Agriculture, 237, 110780. https://doi.org/https://doi.org/10.1016/j.compag.2025.110780