Published online on the page : https://ejournal.uinbukittinggi.ac.id/index.php/ijokid/

Knowbase : International Journal of Knowledge in Database

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

Application of Artificial Intelligence and Deep Learning in Remote Sensing Image Analysis for Natural Resource Monitoring and Management

Mohammad Eisa Sediqi[1], Musawer Hakimi[2]*

[1] *Kabul University, Kabul, Afghanistan*

[2] *Samangan University, Samangan, Afghanistan*

Article Information

Article History:

Submitted: March 13, 2026
Revision: April 30, 2026
Accepted: May 15, 2026
Published: June 28, 2026

Keywords

Deep Learning
Remote Sensing
Natural Resource Monitoring
Semantic Segmentation
Correspondence

Correspondence

E-mail: musawer@adc.edu.in

A B S T R A C T

Natural resource monitoring increasingly relies on satellite and airborne remote sensing to observe land cover change, forest loss, agricultural conditions, and water dynamics over large and often inaccessible areas. Conventional pixel-based and shallow machine-learning classifiers struggle to exploit the spatial, spectral, and temporal complexity of modern multi-sensor archives. This paper reviews and synthesizes recent developments in artificial intelligence (AI) and deep learning (DL) for remote sensing image analysis, focusing on four application domains central to natural resource management: land use/land cover classification, forest and deforestation monitoring, agricultural and crop monitoring, and water resource and flood mapping. Literature spanning convolutional neural networks, encoder-decoder segmentation architectures, and attention-based transformer models is compared in terms of methodology, sensor modality, and reported performance. A generalized processing workflow and an integrated cross-domain conceptual framework are proposed to guide the design of operational AI-based monitoring systems; unlike prior single-domain reviews, this synthesis is explicitly cross-domain, linking architecture choice to operational constraints shared across land cover, forest, agriculture, and water monitoring rather than treating each application silo in isolation. The review finds that transformer and hybrid CNN-transformer architectures tend to outperform purely convolutional baselines on scene-level classification tasks, although this advantage depends on application domain, dataset characteristics, and evaluation protocol and is less consistently observed for dense pixel-level segmentation, where U-Net-family encoder-decoders remain the dominant choice due to their balance of accuracy and computational cost. Persistent challenges include scarcity of high-quality labeled data, limited cross-region generalization, high computational demand, and limited interpretability of deep models for policy-relevant decision making. The paper concludes that multimodal data fusion, self-supervised pretraining, and explainable AI represent the most promising directions for advancing AI-driven natural resource monitoring toward operational, trustworthy deployment.

1. Introduction

Earth observation has entered an era of unprecedented data abundance. Constellations such as Landsat, Sentinel-1/2, Planet, and a growing number of commercial very-high-resolution and hyperspectral platforms now provide continuous, multi-decadal, multi-spectral coverage of the planet's surface. This growth in data volume and revisit frequency has transformed remote sensing from an episodic mapping tool into a candidate backbone for near-real-time natural resource monitoring, covering land cover and land use change, forest degradation and deforestation, crop condition and yield, and water resource dynamics including floods and droughts [1]. Timely and accurate information from these domains underpins climate policy, disaster response, food security planning, and sustainable land and water management.

Traditional approaches to remote sensing image analysis rely on pixel-based spectral classifiers or hand-engineered texture and shape features combined with shallow classifiers such as support vector machines or random forests. These methods are well understood and computationally inexpensive, but they generalize poorly across sensors, seasons, and geographic regions, and they cannot easily exploit the joint spatial-spectral-temporal structure present in modern imagery. Since the early 2010s, deep learning (DL) has offered an alternative: convolutional neural networks (CNNs) learn hierarchical spatial features directly from data, encoder-decoder architectures produce dense pixel-level predictions suited to mapping tasks, and, more recently, attention-based transformer models capture long-range spatial dependencies that convolutional receptive fields struggle to reach [1], [2], [8], [9]. Most recently, large-scale foundation models and vision-language models pretrained on multimodal Earth observation archives have begun to extend this trajectory beyond task-specific architectures, offering a single pretrained backbone that can be adapted to multiple remote sensing tasks with limited labeled data [41], [42].

Despite rapid progress, the literature on AI/DL for remote sensing remains fragmented across application silos: land cover classification, forest monitoring, precision agriculture, and hydrology are typically studied by separate research communities using domain-specific datasets and evaluation protocols, which makes it difficult to compare architectural choices across natural resource domains or to design an integrated monitoring system that serves multiple resource management functions simultaneously. Furthermore, most primary studies report accuracy figures under a single dataset and sensor configuration, leaving open questions about which architecture families transfer best across domains, and what operational constraints (labeled-data availability, computational budget, interpretability requirements) should drive architecture selection in practice. Prior reviews in this space remain confined to a single vantage point: broad surveys of land use/land cover classification [13], systematic reviews of hyperspectral-LiDAR fusion for land classification [14], and machine-learning-versus-deep-learning decision frameworks in land system science [15] each focus on one application domain, while general reviews of deep learning in remote sensing [1] and flood-mapping-specific surveys [32] stop at a single domain or a single methodological question. None of these works explicitly compares architecture performance, computational cost, and operational feasibility across land cover, forest, agriculture, and water monitoring within one cross-domain synthesis, which is the specific gap this review addresses. The unresolved problems motivating this review include the scarcity of high-quality labeled data for rare event classes such as active deforestation fronts or flash floods, the limited cross-region and cross-sensor transferability of trained models, the absence of standardized multi-domain benchmark datasets that would allow fair architecture comparison, and the limited explainability of deep models for policy-relevant decision making.

This paper addresses that gap with three objectives. First, it synthesizes recent DL architectures used in remote sensing image analysis into a coherent taxonomy spanning convolutional, encoder-decoder, and transformer-based families. Second, it reviews and compares representative applications across four natural resource domains: land cover/land use, forest and deforestation, agriculture, and water resources, extracting reported methodologies, sensor modalities, and outcomes. Third, it proposes a generalized processing workflow and an integrated cross-domain conceptual framework intended to guide the design of operational, multi-purpose AI-based natural resource monitoring systems. The novelty of this work lies not in a single new model, but in the cross-domain synthesis and framework that connects architecture choice to the practical constraints of natural resource management: whereas the existing review papers cited above synthesize evidence within one application domain or one architecture family, the present review and its accompanying conceptual framework (Section 3.6) integrate evidence and design guidance across all four domains simultaneously, a synthesis not presented in the reviewed literature.

2. Method

A systematic literature synthesis is employed in order to provide an overview of the main advancements in AI and deep learning used in remote sensing for the monitoring of natural resources. Not like studies performed using a single dataset, the study aims to offer a comparison of methods developed in different application domains, which include but are not limited to land cover classification, forest monitoring, agricultural monitoring, and mapping of water resources.

2.1 Literature Search Strategy

The literature search strategy was developed to systematically identify relevant studies related to artificial intelligence and deep learning applications in remote sensing-based natural resource monitoring. The search process considered multiple databases, keywords, publication periods, and methodological requirements to ensure comprehensive study coverage.

Table 1. Literature Search Strategy

Component	**Details**	**Purpose**	**Selection Criteria**	**Output**
Review Method	Structured literature synthesis approach	To compare AI and deep learning methodologies across different natural resource domains	Systematic analysis of existing studies	Comprehensive research synthesis
Research Scope	AI, deep learning, remote sensing, Earth observation, and environmental monitoring	To identify current technological developments	Studies related to natural resource monitoring	Domain-specific evidence base
Database Sources	IEEE Xplore, ScienceDirect, SpringerLink, MDPI (Remote Sensing and Sensors), preprint repositories	To retrieve high-quality scientific publications	Reputable academic databases	Candidate research articles
Publication Period	2015–2025	To capture historical development and recent advances	Preference given to 2021–2025 studies	Updated state-of-the-art review
Search Keywords	Deep learning, Artificial Intelligence, Remote sensing, Land cover, Deforestation, Crop mapping, Flood mapping, Transformer	To identify relevant research papers	Keyword combinations using AND/OR operators	Retrieved records
Language	English publications	To ensure accessibility and consistency	English-language articles only	Eligible studies

Component	Details	Purpose	Selection Criteria	Output
Initial Search Results	Approximately 1,000 records	To establish a broad research pool	All potentially relevant studies	Initial literature database

2.2 *Inclusion and Exclusion Criteria for Study Selection*

The inclusion and exclusion criteria were predefined to ensure that only relevant, high-quality, and methodologically sound studies were selected for this review. These criteria guided the screening process and improved the transparency and reproducibility of the literature selection procedure.

Table 2. Inclusion and Exclusion Criteria for Study Selection

Criterion Category	Inclusion Criteria	Exclusion Criteria	Selection Purpose
Research Topic	Studies focusing on artificial intelligence, deep learning, and remote sensing applications for natural resource monitoring.	Studies unrelated to AI, deep learning, remote sensing, or natural resource monitoring.	To maintain alignment with the research objectives.
Application Domain	Studies addressing land use/land cover classification, forest monitoring, deforestation detection, agricultural monitoring, crop mapping, water resource analysis, or flood mapping.	Studies focusing on unrelated domains such as medical imaging, industrial vision, or non-environmental applications.	To ensure consistency with the defined research scope.
Data Source	Studies using satellite, aerial, drone, or UAV-based remote sensing imagery.	Studies without remote sensing data or relying only on non-image datasets.	To focus on Earth observation-based monitoring approaches.
Artificial Intelligence Method	Studies applying deep learning models such as CNN, U-Net, ResNet, Transformer, Vision Transformer, GAN, or hybrid architectures.	Studies using only traditional statistical methods or machine learning approaches without deep learning components.	To analyze modern AI-driven approaches.
Publication Type	Peer-reviewed journal articles and recognized scientific preprints with available full text.	Editorials, opinion papers, technical notes, abstracts, and non-peer-reviewed documents.	To improve the reliability of reviewed evidence.
Publication Period	Studies published between 2015 and 2025, with emphasis on recent studies from 2021–2025.	Studies published outside the defined period unless	To capture recent developments while preserving important

Criterion Category	Inclusion Criteria	Exclusion Criteria	Selection Purpose
		considered foundational architecture papers.	theoretical foundations.
Language	Articles published in English.	Non-English publications without accessible translations.	To ensure consistent analysis and interpretation.
Methodological Information	Studies providing clear descriptions of datasets, preprocessing methods, AI architectures, and experimental procedures.	Studies lacking sufficient methodological details or unclear research procedures.	To support reproducibility and technical evaluation.
Performance Evaluation	Studies reporting measurable evaluation results using metrics such as Accuracy, Precision, Recall, F1-score, IoU, RMSE, or similar indicators.	Studies without experimental validation or performance assessment.	To enable quantitative comparison among approaches.
Full-Text Availability	Studies with accessible complete manuscripts for detailed analysis.	Studies where only titles, abstracts, or incomplete information were available.	To ensure comprehensive data extraction.
Duplicate Records	The first or most complete version of a study was retained.	Duplicate publications retrieved from multiple databases were removed.	To avoid repeated evidence and bias.
Foundational Studies	Highly influential deep learning architecture papers used as theoretical foundations were included.	Irrelevant foundational papers not connected to reviewed applications were excluded.	To establish background knowledge for model taxonomy.

2.3 Literature Screening and Selection Process

The selection process followed a systematic workflow consisting of identification, screening, eligibility assessment, and final inclusion. The procedure follows PRISMA principles to improve transparency and reproducibility.

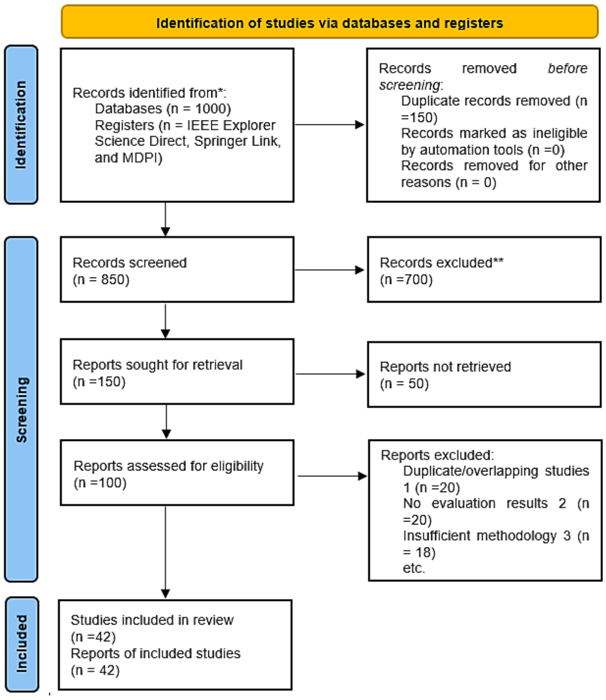

Figure 1: PRISMA flow Diagram for Literature Screening and Selection Process

2.4 *Quality Assessment Framework*

Each selected study was evaluated using predefined quality indicators to determine methodological reliability and scientific contribution.

Table 3. Quality Assessment Framework

Quality Dimension	Assessment Question	Evaluation Focus	Importance
Research Objective	Are research goals clearly defined?	Problem statement and objectives	Ensures research relevance
Dataset Information	Is dataset information provided?	Sensor, resolution, and source	Supports reproducibility
Method Description	Is the AI model sufficiently explained?	Architecture and training process	Enables technical comparison
Experimental Design	Are experiments appropriately conducted?	Validation and comparison methods	Ensures reliability
Evaluation Metrics	Are suitable metrics reported?	Accuracy, IoU, F1-score, RMSE	Enables performance comparison
Limitations	Are challenges discussed?	Data, computation, generalization	Identifies research gaps

2.5 *Data Extraction Framework*

Information extracted from selected studies was organized into multiple categories to support comparative analysis among AI-based remote sensing approaches.

Table 4. Data Extraction Framework

Extraction Category	Collected Information	Example Values	Purpose
Publication Information	Author, year, journal	2024, Remote Sensing	Literature classification
Application Domain	Target monitoring area	Forest, agriculture, flood	Domain comparison
Dataset	Data source and characteristics	Sentinel-2, Landsat, UAV	Data analysis
Sensor Type	Remote sensing modality	Optical, SAR, Hyperspectral	Sensor comparison
AI Architecture	Deep learning model	CNN, U-Net, Transformer	Model comparison
Learning Strategy	Training approach	Supervised, Self-supervised	Method analysis
Evaluation Metrics	Performance indicators	Accuracy, IoU, F1-score	Quantitative comparison
Limitations	Reported challenges	Data scarcity, computation	Research gap identification

2.6. Deep Learning Architecture Taxonomy

Deep learning models applied to remote sensing image analysis can be organized into four broad families, illustrated in Figure 1. Convolutional neural networks (CNNs), exemplified by ResNet and related backbones, learn local spatial-spectral features through stacked convolution and pooling operations and remain the default feature extractor for scene-level classification tasks [2]. Encoder-decoder segmentation networks, including U-Net, its nested variant UNet++, SegNet, DeepLab, and the remote-sensing-specific ResUNet-a, extend CNNs with a decoder path and skip connections that recover spatial resolution lost during downsampling, making them the dominant choice for pixel-level mapping tasks such as land cover and water body segmentation [3]-[7]. Although CNN- and encoder-decoder-based methods substantially improved spatial feature extraction and pixel-level mapping, their fixed, local receptive fields limit their ability to model long-range spatial dependencies across a scene, a limitation that motivated the emergence of attention-based transformer architectures in remote sensing. Attention-based transformer models, built on the self-attention mechanism introduced by Vaswani et al. [8] and adapted to images by the Vision Transformer (ViT) [9], model long-range dependencies across the entire image rather than a local receptive field; remote-sensing-specific variants such as ViT for scene classification, TRS, and Swin Transformer have shown competitive or superior performance to CNNs on benchmark scene-classification datasets [10]-[12]. A fourth, hybrid family combines convolutional and attention mechanisms, or incorporates generative adversarial networks (GANs) for data augmentation and sensor fusion, aiming to combine the local inductive bias of CNNs with the global context modeling of transformers.

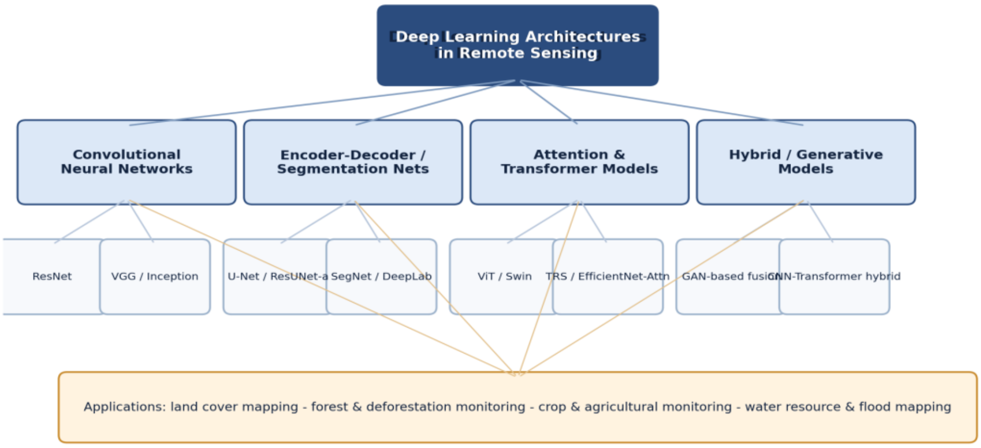

Figure 2. Taxonomy of Deep Learning Architectures Applied in Remote Sensing Image Analysis

The taxonomy in Figure 1 reflects a historical progression rather than a static classification. Figure 3 places the architectures reviewed in this paper on a timeline of key milestones, from the CNN breakthroughs of the early 2010s through the introduction of encoder-decoder segmentation networks, to the transformer-based models that have entered remote sensing since 2020. This progression is directly visible in the reviewed literature: forest and flood monitoring studies published before 2020 rely almost exclusively on CNN and U-Net-family backbones [19]-[25], [31]-[36], whereas post-2020 land-cover and scene-classification studies increasingly adopt attention and transformer variants [10]-[12], [17].

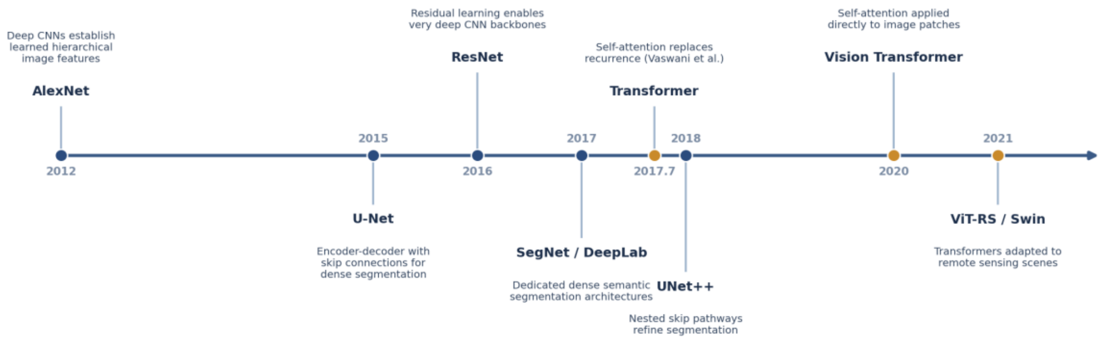

Figure 3. Timeline of Key Deep Learning Architecture Milestones Adopted in Remote Sensing

2.7. *Generalized Processing Workflow*

Across the reviewed domains, AI/DL-based remote sensing analysis follows a broadly consistent processing pipeline, summarized in Figure 3. Raw optical, synthetic aperture radar (SAR), hyperspectral, or UAV imagery is first acquired and radiometrically and geometrically pre-processed (atmospheric correction, orthorectification, normalization). A deep learning backbone then performs feature learning on the pre-processed imagery, followed by a task-specific head that performs classification (assigning a label to an image or region) or semantic segmentation (assigning a label to every pixel). Post-processing steps such as multi-temporal change detection, spatial filtering, and accuracy assessment against ground-truth or reference data refine the raw model output before it is delivered to decision-support systems used by resource managers, disaster response agencies, and policymakers.

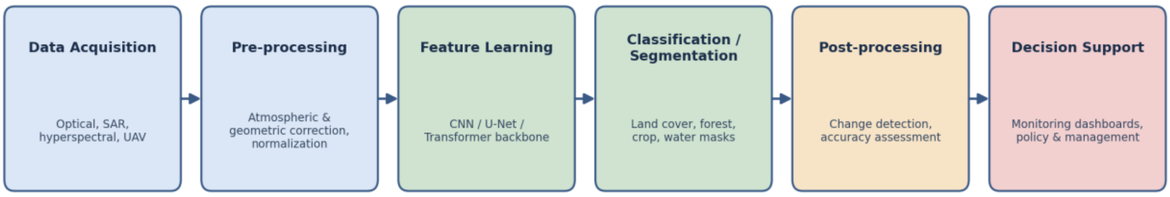

Figure 4. Generalized AI/Deep Learning Pipeline for Remote Sensing-Based Natural Resource Monitoring

A representative pixel-wise segmentation loss used to train the encoder-decoder and hybrid models reviewed in this paper is the weighted cross-entropy loss:

$$L = - \Sigma_i w_i \cdot y_i \cdot \log(\hat{y}_i) \qquad (1)$$

where y_i is the ground-truth class indicator for pixel i, $\hat{y}_i$ is the predicted class probability, and w_i is a class-balancing weight commonly introduced to counter the severe class imbalance typical of natural resource maps (e.g., water or deforested pixels forming a small minority of a scene) [7]. Model performance in the reviewed studies is most commonly reported using overall accuracy (OA), the Intersection-over-Union (IoU) metric,

$$IoU = TP / (TP + FP + FN) \qquad (2)$$

and the F1-score, where TP, FP, and FN denote true positive, false positive, and false negative pixel or scene counts, respectively. The weighted cross-entropy loss and the OA/IoU/F1 metrics are presented here, rather than a comprehensive list of loss functions and metrics used across all 40 reviewed studies, because they are the formulations most consistently adopted across the encoder-decoder and hybrid segmentation studies reviewed in Sections 3.1-3.4 and therefore serve as a representative, comparable baseline against which domain-specific variants (e.g., Dice or focal loss for extreme class imbalance) can be understood.

3 Results and Discussion

Figure 5 organizes the deep learning architecture families identified in the reviewed literature by primary task and representative models. The following subsections synthesize how these architectures have been applied within five major natural resource monitoring domains, followed by a discussion of cross-cutting methodological challenges.

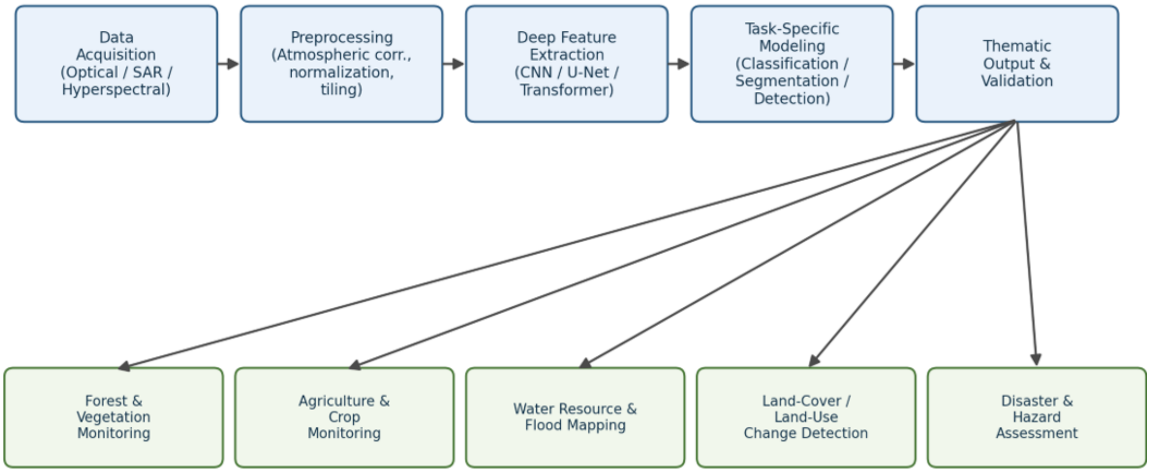

Figure 5. Deep Learning Architecture Families Organized by Primary Task and Representative Models Across the Reviewed Natural Resource Monitoring Domains

3.1. Land Use and Land Cover Classification

Land use/land cover (LULC) classification is the most extensively studied application of DL in remote sensing, reflecting its foundational role in urban planning, environmental protection, and resource management [13]. Early CNN-based LULC studies demonstrated clear gains over shallow classifiers on high-resolution multispectral imagery [16], while subsequent work extended this to hybrid MLP-CNN classifiers [39] and deep feature-fusion approaches for very-high-resolution scene classification [40]. More recent studies incorporate multi-sensor fusion: a transformer-based SegFormer model with spatial attention, applied to a stack of optical (Sentinel-2, PRISMA), SAR (Sentinel-1), elevation, and human-settlement layers, achieved an overall accuracy of 86.9%, with each additional data source contributing incremental accuracy gains [17]. Complementary reviews confirm that attention-augmented CNNs, such as an EfficientNet-B3 backbone with an attention module, and semi-supervised multi-CNN ensembles, further improve robustness when labeled training data are limited [37], [38]. Systematic reviews spanning hyperspectral-LiDAR fusion and decision-oriented comparisons of machine learning versus deep learning consistently report that DL narrows the gap between spectral similarity and true land-cover class boundaries, particularly in heterogeneous or urban landscapes [13]-[15].

Across these studies, the comparative picture is more nuanced than a simple claim that transformers outperform CNNs: attention-based fusion models such as the SegFormer pipeline [17] gain their advantage primarily when multiple sensor modalities are stacked and labeled training data are comparatively abundant, whereas plain CNN backbones remain competitive, and considerably cheaper to train and deploy, on single-sensor, moderately sized benchmarks [16], [40]. The attention-augmented and ensemble approaches that improve robustness under limited labels [37], [38] do so at the cost of additional architectural complexity and hyperparameter tuning, which smaller monitoring agencies may struggle to absorb. Taken together, the LULC literature shows a clear technological evolution from single-sensor CNN classifiers toward multi-sensor, attention-based fusion pipelines over the past decade, with the key finding that accuracy gains from added sensor modalities and attention mechanisms are real but incremental, typically a few accuracy points per additional data source [17], rather than transformative. Remaining challenges include continued reliance on high-resolution labeled reference data for heterogeneous urban landscapes and limited testing of fusion pipelines outside the specific sensor combinations reported in the original studies. For resource-management practice, these findings imply that agencies with access only to single-sensor optical archives can still obtain reliable LULC maps from mature CNN or encoder-decoder pipelines, while multi-sensor fusion and attention mechanisms are best reserved for applications, such as fine-grained urban land-cover discrimination, where the added accuracy justifies the additional data and computational cost.

3.2. Forest and Deforestation Monitoring

Forest monitoring applications place a premium on temporal consistency and sensitivity to subtle canopy disturbance. A U-Net model trained on monthly Planet NICFI composites produced a continuous tropical tree-cover time series for Mato Grosso, Brazil, from 2015 to 2021, demonstrating that encoder-decoder architectures can support operational, sub-annual deforestation tracking at 5 m resolution [19]. Forecasting, rather than retrospective detection, has also been explored: a three-dimensional CNN combining multispectral imagery, the Hansen global forest-change dataset, and an ALOS PALSAR digital surface model produced spatially explicit one-year-ahead deforestation risk forecasts from freely available global data layers [20]. For change detection specifically, superpixel-enhanced networks such as ESCNet [21] and unsupervised multimodal approaches based on Fourier-domain structural relationship analysis [22] address the practical difficulty of obtaining paired, cloud-free before/after imagery in tropical regions. Lightweight, high-resolution models such as RepDDNet [23] and multi-source fusion approaches [24] further target the computational cost of large-area monitoring, while aerial-imagery-based real-time monitoring systems [25] extend deforestation detection to finer local scales suited to forest-ranger operations.

Comparing these forest-monitoring studies critically, architecture choice here is shaped less by raw accuracy than by the temporal and computational demands of large-area monitoring: the Planet NICFI U-Net pipeline [19] excels at retrospective, sub-annual mapping but requires dense, cloud-free time series, whereas the 3D CNN forecasting approach [20] trades finer spatial resolution, 30 m versus 5 m, for the ability to anticipate deforestation risk ahead of the event, a capability retrospective segmentation models cannot offer. Lightweight models such as RepDDNet [23] and aerial-based real-time systems [25] sacrifice some accuracy relative to heavier fusion architectures [24] in exchange for the inference speed near-real-time ranger operations require, an accuracy-versus-latency trade-off that is largely absent from the LULC literature in Section 3.1. The domain has evolved from single-date change detection toward continuous, monthly-composite monitoring and, more recently, toward forecasting rather than purely retrospective detection; the key finding is that no single architecture dominates across the retrospective-detection, forecasting, and real-time-alerting use cases, and the appropriate choice depends on which of these operational modes a monitoring program prioritizes. Persistent challenges include the scarcity of paired, cloud-free before/after imagery in tropical regions and the difficulty of validating forecast-style models against ground truth that does not yet exist at prediction time. Practically, these results suggest that operational deforestation-monitoring systems may need to combine more than one architecture family, for example a lightweight detector for rapid alerting paired with a heavier encoder-decoder model for periodic, higher-accuracy verification, rather than relying on a single model for all monitoring functions.

3.3. Agricultural and Crop Resource Monitoring

Crop mapping and monitoring must contend with strong intra-class temporal variability driven by phenology and with frequent cloud contamination of optical imagery. Hyperspectral one-shot mapping approaches attempt to compensate for cloud gaps by extracting maximal information from single cloud-free acquisitions [26], while neural ordinary differential equation (Neural ODE) classifiers explicitly model continuous-time crop growth trajectories to remain robust to irregular, cloud-limited observation sequences [27]. Multi-scale label hierarchies embedded within recurrent and convolutional time-series models improve fine-grained crop-type discrimination beyond coarse crop/non-crop separation [28], and standardized multi-country benchmark datasets such as those built on Sentinel-2 time series now allow more direct comparison of crop-segmentation architectures across regions [29]. Long short-term memory (LSTM) recurrent networks applied to multi-year Landsat and Cropland Data Layer time series further demonstrate that temporal sequence models capture phenological patterns that single-date classifiers cannot [30]. Collectively, these studies indicate that, unlike static land cover mapping, crop monitoring benefits disproportionately from architectures that explicitly encode the time dimension, whether through recurrent units, temporal attention, or continuous-time formulations.

A critical comparison across these crop-monitoring studies shows that, unlike the LULC and forest domains, architecture choice is driven primarily by how explicitly a model represents time rather than by the CNN-versus-transformer distinction: Neural ODE classifiers [27] and LSTM-based sequence models [30] outperform single-date classifiers specifically because they encode continuous or discrete phenological trajectories, while the multi-scale label-hierarchy approach [28] improves fine-grained crop-type discrimination by exploiting label structure rather than temporal modeling alone. This indicates that, for crop monitoring, the practical cost of cloud contamination and phenological variability outweighs any accuracy advantage a purely spatial architecture, CNN, transformer, or hybrid, might offer without an explicit temporal component. The domain has evolved from single-date, spatially focused classifiers toward time-series and multi-country benchmark-driven approaches such as the Sentinel-2 benchmark dataset [29], reflecting growing recognition that crop-type accuracy claims are meaningful only when compared on standardized, multi-region data. The key finding is that temporal architectures consistently and substantially outperform static classifiers under realistic, cloud-affected observation conditions, a more consistent pattern than the domain-dependent CNN-versus-transformer trade-offs seen in Sections 3.1 and 3.2. Remaining challenges include limited generalization of crop-type models across agro-climatic zones and growing seasons, and continued scarcity of densely labeled, multi-year ground-truth datasets outside a handful of benchmark regions. For practice, these findings suggest that agricultural monitoring agencies should prioritize investment in temporal data acquisition and sequence-aware architectures over investment in more complex spatial backbones alone.

3.4. Water Resource and Flood Monitoring

Because floods frequently coincide with cloud cover, synthetic aperture radar (SAR) is the dominant sensor modality in this domain, and nearly all reviewed studies pair SAR imagery with an encoder-decoder segmentation network. A modified DeepLabv3+ model with a lightweight MobileNetV2 encoder was applied to HISEA-1 SAR imagery covering 20,000 km2 of mixed rivers, reservoirs, and paddy fields, improving floodwater segmentation efficiency while addressing the speckle noise characteristic of SAR data [33]. U-Net-based classifiers trained on Sentinel-1 imagery have been applied operationally to tropical, cloud-prone flood events, including flooding in Tabasco, Mexico [31], [34], and standardized benchmark datasets such as Sen1Floods11 [35] and the more recent multi-sensor S1S2-Water dataset [36] are increasingly used to support reproducible comparison across flood-mapping architectures. Systematic reviews of the flood-mapping literature confirm a broad trend toward encoder-decoder and, increasingly, transformer-based ensembles that also report calibrated uncertainty estimates alongside inundation maps, reflecting the operational need for confidence bounds in disaster-response decision making [32].

Critically comparing the flood-monitoring literature, the near-universal reliance on SAR imagery and encoder-decoder segmentation, rather than transformer-based classification, reflects a domain-specific sensing constraint rather than an architectural preference: because floods coincide with cloud cover that defeats optical sensing, the choice of SAR narrows the useful architecture space to models such as DeepLabv3+ [33] and U-Net [31], [34] that handle SAR speckle noise and produce pixel-level inundation maps, largely independent of whether CNN, transformer, or hybrid feature extractors dominate elsewhere in the pipeline. The lightweight MobileNetV2 encoder used in [33] shows that, as in forest monitoring, computational efficiency is prioritized over marginal accuracy gains when disaster response requires near-real-time turnaround. The domain has evolved from single-event, study-specific flood maps toward standardized benchmark datasets such as Sen1Floods11 [35] and the multi-sensor S1S2-Water dataset [36], enabling more reproducible cross-study comparison than was previously possible. The key finding is that reported progress in this domain increasingly comes from calibrated uncertainty estimation alongside segmentation [32], rather than from segmentation accuracy alone, reflecting the operational reality that disaster-response decisions require confidence bounds, not just a binary flood or no-flood map. Remaining challenges include limited transferability of flood models trained in one river basin or climate regime to another, and the continued difficulty of acquiring timely SAR coverage in some regions during active flood events. Practically, these findings imply that flood-monitoring systems should be built around SAR-compatible encoder-decoder architectures with explicit uncertainty quantification, rather than around the CNN-versus-transformer question that dominates the LULC literature.

3.5. Comparative Synthesis

Table 5 summarizes representative studies across the four domains, and Table 6 compares the four architecture families identified in Section 2.1 in terms of their strengths and limitations for natural resource monitoring tasks.

Table 5. Comparative Synthesis of Representative Studies Across Natural Resource Domains

Domain	Representative Model	Sensor / Data Source	Key Reported Outcome	Reference
Land cover / land use mapping	SegFormer + spatial attention (Transformer)	Sentinel-1/2, PRISMA, Cartosat DEM, GHSL	Multi-sensor fusion raised overall classification accuracy to 86.9%	[17]
Land cover / land use mapping	CNN scene-classification benchmark	VHR optical imagery	Systematic benchmarking shows CNN backbones outperform shallow ML baselines	[13]
Forest / deforestation monitoring	U-Net (monthly composites)	Planet NICFI (5 m)	Enabled continuous monthly tropical tree-cover mapping 2015-2021	[19]
Forest / deforestation monitoring	3D CNN forecasting model	Multispectral + Hansen forest-change maps, ALOS PALSAR DSM	Forecast year-ahead deforestation risk at 30 m from limited global layers	[20]
Forest / deforestation monitoring	ESCNet (superpixel-enhanced)	Very-high-resolution optical imagery	End-to-end change-detection network improved boundary delineation	[21]
Agricultural / crop monitoring	Neural ODE crop classifier	Multi-temporal Sentinel-2	Maintained accuracy under variable cloud cover across a growing season	[27]
Agricultural / crop monitoring	Hierarchical multi-scale label CNN	Satellite image time series	Improved fine-grained crop-type mapping via label hierarchies	[28]
Agricultural / crop monitoring	LSTM recurrent network	Landsat + Cropland Data Layer time series	Captured temporal phenology patterns for multi-year land-cover labeling	[30]
Water resource / flood monitoring	Modified DeepLabv3+ (MobileNetV2)	HISEA-1 SAR imagery	Lightweight encoder improved floodwater segmentation efficiency over 20,000 km2	[33]

Domain	Representative Model	Sensor / Data Source	Key Reported Outcome	Reference
Water resource / flood monitoring	U-Net SAR classifier	Sentinel-1 SAR	Produced operational flood-extent maps for tropical, cloud-prone regions	[31],[34]

Table 6. Comparison of Deep Learning Architecture Families for Remote Sensing-Based Monitoring

Architecture Family	Strengths	Limitations	Refs
CNN (ResNet, VGG-type)	Strong local spectral-spatial feature extraction; mature, widely benchmarked	Limited long-range context; large labeled datasets typically required	[1],[2]
Encoder-decoder segmentation nets (U-Net, SegNet, DeepLab)	Pixel-level output well suited to mapping tasks; skip connections preserve fine boundaries	Sensitive to class imbalance; heavy computation at high resolution	[3]-[7]
Attention / Transformer (ViT, Swin, TRS)	Captures global spatial context; strong transfer performance on scene classification	Data-hungry; higher training cost without pretraining	[8]-[12]
Hybrid CNN-Transformer / GAN-based fusion	Combines local detail with global context; enables multi-sensor fusion and data augmentation	Increased architectural complexity; harder to interpret and deploy operationally	[9],[18]

Building on Tables 1 and 2, the reviewed literature suggests the following practical guidance for architecture selection. A plain CNN backbone is the appropriate default for small or moderately sized labeled datasets and local, texture-driven feature extraction, as in early LULC scene classification [16], [40]. U-Net-family encoder-decoders remain the preferred choice whenever the output is a pixel-level map, land cover, forest disturbance, or flood extent, because their skip connections preserve fine boundary detail at manageable computational cost [3]-[7], [19], [31], [33], [34]. Vision Transformer and Swin Transformer variants are best reserved for scene-level classification tasks with access to large, well-labeled, and ideally multi-sensor datasets, where their global context modeling yields a measurable accuracy advantage [10]-[12], [17]. Hybrid CNN-transformer and multi-sensor fusion architectures deliver the highest reported accuracy in this review but at the greatest computational and interpretability cost, and are therefore best suited to applications where accuracy is paramount and computational resources are not the binding constraint. Generative adversarial networks appear in the reviewed literature chiefly as a data-augmentation and sensor-fusion tool rather than as a stand-alone classifier, useful for expanding limited labeled training sets in data-scarce domains such as forest and flood monitoring. Autoencoder-based architectures, while not a primary focus of the studies reviewed here, are a natural complement for anomaly-style detection tasks, for example flagging atypical canopy-loss or flood-extent patterns that fall outside previously observed distributions, and represent a direction the reviewed literature has not yet fully explored.

Three cross-domain patterns emerge from this synthesis. First, encoder-decoder architectures (U-Net and its variants) remain the practical default for pixel-level mapping tasks across all four domains, owing to their favorable balance between segmentation accuracy and computational cost, and their ability to be trained from comparatively modest labeled datasets relative to transformer models. Second, transformer and attention-augmented architectures increasingly outperform purely convolutional baselines specifically on scene-level classification tasks (Section 3.1), but their advantage on dense pixel-level segmentation tasks central to forest, agriculture, and flood monitoring is less consistently established, likely reflecting both the smaller labeled datasets typical of these domains and the higher data requirements of self-attention. Third, sensor choice is strongly domain-dependent: optical time series dominate agricultural monitoring where phenology is informative, whereas SAR dominates flood and, increasingly, forest disturbance monitoring where persistent cloud cover would otherwise interrupt optical observation.

Figure 6 visualizes the qualitative trade-offs summarized in Table 2 across five practical selection criteria. The plot illustrates why no single architecture family dominates across all criteria: encoder-decoder networks offer the most balanced profile for operational mapping, transformer-based models trade data and compute efficiency for stronger global context modeling, and hybrid CNN-transformer designs push segmentation accuracy further at the cost of interpretability, reinforcing that architecture selection should be driven by the operational constraints of the specific natural resource monitoring task rather than by benchmark accuracy alone.

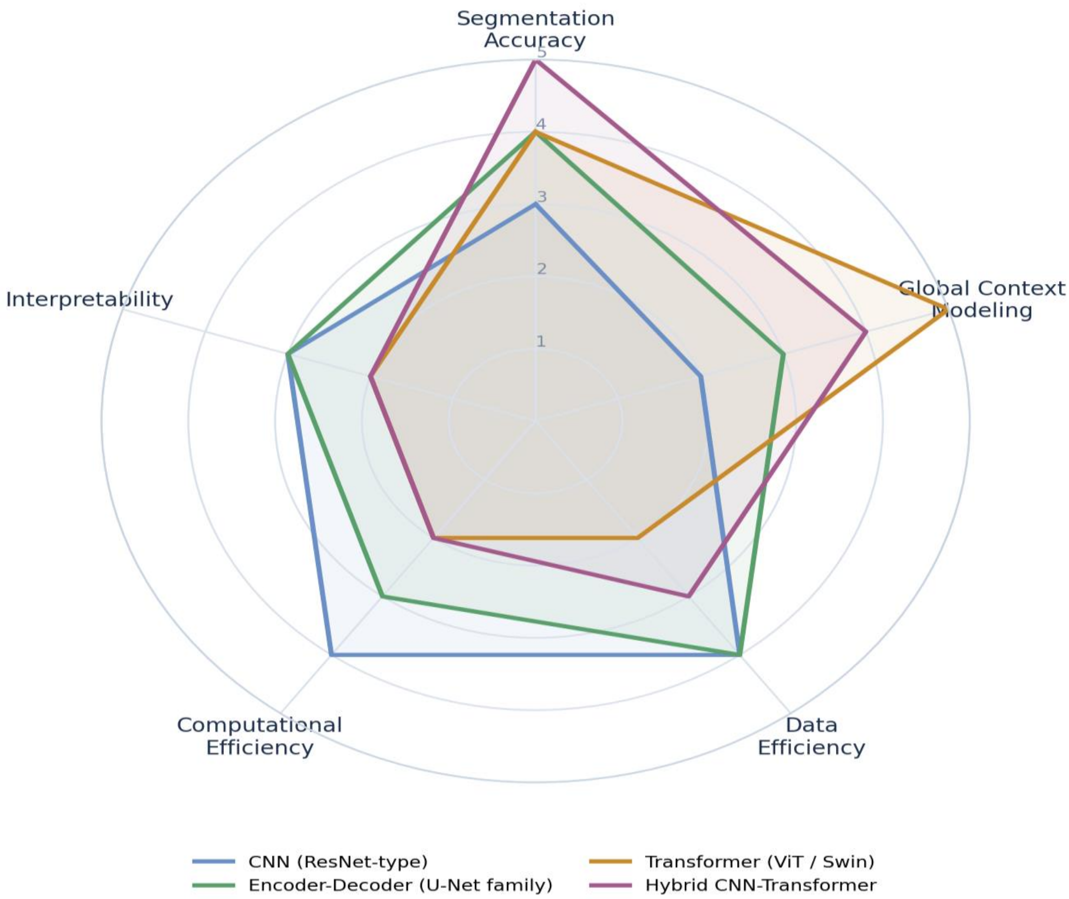

Figure 6. Qualitative Comparison of Deep Learning Architecture Families Across Practical Selection Criteria

3.6. Emerging AI Technologies and Future Trajectories

Beyond the architecture families synthesized in Sections 2.1 and 3.1-3.5, the period 2024-2025 has seen rapid growth in large-scale foundation models pretrained on broad, often multimodal, Earth observation archives rather than trained from scratch for a single task. Vision-language models (VLMs) adapted to remote sensing pair image encoders with natural-language understanding, enabling free-text querying of satellite imagery and shifting the emphasis of remote sensing AI from purely descriptive classification toward decision-support interaction with domain experts [41]. Parallel work on remote sensing foundation models more broadly demonstrates that a single pretrained backbone, fine-tuned or adapted with comparatively little labeled data, can be applied across several of the tasks reviewed in this paper, including land cover classification, change detection, and segmentation, directly addressing the cross-domain fragmentation identified in Section 1 [42]. Large-scale geospatial pretraining, in which models are self-supervised on unlabeled multi-temporal, multi-sensor archives before task-specific fine-tuning, is particularly relevant to the labeled-data scarcity challenge noted throughout Sections 3.1-3.4, since it reduces the amount of domain-specific labeled data required to reach competitive accuracy. Multimodal remote sensing models that jointly ingest optical, SAR, and elevation data extend the multi-sensor fusion trend already visible in the LULC literature [17] toward a more general-purpose fusion capability not tied to a single downstream task. For natural resource monitoring specifically, these emerging technologies suggest a plausible future trajectory in which the domain-specific encoder-decoder and transformer models reviewed in this paper are progressively complemented, or in some cases replaced, by fine-tuned foundation-model adapters; however, this transition remains largely prospective in the literature reviewed here and has not yet been demonstrated at the operational scale that Sections 3.1-3.4 describe for current architectures.

3.7. Challenges and an Integrated Framework

Despite consistent accuracy gains reported across the reviewed literature, several challenges limit the transition from research prototypes to operational natural resource management systems. Labeled training data remain scarce and expensive to produce at the spatial and temporal scale required for global or national monitoring, particularly for rare event classes such as active deforestation fronts or flash floods. Cross-region and cross-sensor generalization remains limited; models trained on one geography or sensor configuration frequently degrade when applied elsewhere, motivating growing interest in domain adaptation and self-supervised pretraining on large unlabeled archives. Computational cost, especially for transformer-based and multi-sensor fusion models, constrains deployment in resource-limited settings where monitoring is often most needed. Finally, the limited interpretability of deep models complicates their use in policy and legal contexts, where resource managers must be able to justify decisions such as flagging an area for enforcement action.

To address the fragmentation noted in Section 1, Figure 6 proposes an integrated conceptual framework in which a shared AI/DL analytics core, built on the architecture taxonomy of Section 2.1, is fed by a common pool of multi-sensor data sources (optical, SAR, hyperspectral, UAV, and multi-temporal archives) and serves the four natural resource domains reviewed in this paper simultaneously, rather than as isolated single-purpose systems. This differs from existing AI-based remote sensing frameworks in scope rather than in its individual components: single-domain frameworks proposed in prior LULC [13], [17] and flood-mapping [32] reviews optimize a pipeline for one application, whereas the framework proposed here treats the four domains as sharing a common analytics core and data pool, with domain-specific heads attached only at the task-specific output stage. Such shared infrastructure would, in principle, allow, for example, a change-detection backbone trained for deforestation monitoring to be explored as a starting point for fine-tuning toward flood-extent mapping with comparatively little additional labeled data; this cross-domain transfer is a potential application of the proposed framework rather than a capability demonstrated by the studies reviewed here, none of which reports transferring a trained backbone across the four domains, and would need to be validated empirically before being relied upon operationally. The framework would additionally allow resource managers to access land cover, forest, crop, and water information from a single monitoring platform rather than disparate domain-specific tools.

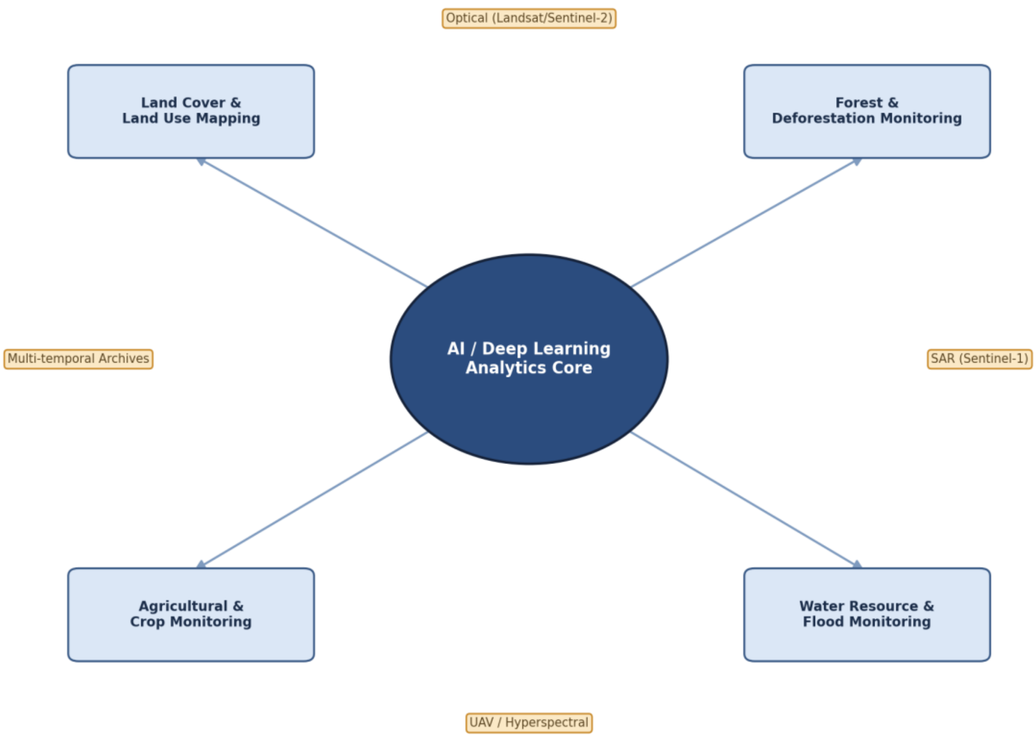

Figure 7. Integrated Conceptual Framework for AI-Driven Multi-Domain Natural Resource Monitoring and Management

3.8. Practical Implementation Considerations

Realizing the shared architecture proposed in Figure 7 requires resolving several practical issues that are largely outside the scope of the algorithmic comparisons in Sections 3.1-3.6 but determine whether the framework can be deployed by resource-management agencies rather than remaining a research prototype. Data governance is a first-order concern, since a shared multi-domain platform pools imagery and derived products across agencies and jurisdictions that may operate under different data-sharing agreements, licensing terms, and national data-sovereignty rules. Interoperability follows closely: the sensor, format, and coordinate-reference heterogeneity across optical, SAR, hyperspectral, and UAV sources noted throughout this review must be resolved through common data standards before a shared analytics core can ingest them reliably. Computational infrastructure and cloud computing are practical enablers rather than optional extras, since the transformer and hybrid architectures found in Section 3.5 to deliver the highest accuracy also carry the greatest computational cost, making cloud-based or shared high-performance computing access a precondition for many monitoring agencies rather than a convenience. Scalability considerations follow: a framework validated on the study areas reported in Sections 3.1-3.4 must be demonstrated at national or continental scale, with the attendant data-volume and processing-time implications, before it can be considered operational. Policy adoption depends on the interpretability limitations discussed above, since resource managers and courts require auditable justification for enforcement decisions, meaning explainability is a deployment precondition rather than a desirable extra. Privacy considerations arise where very-high-resolution imagery approaches a scale at which individual property or activity becomes identifiable, requiring governance safeguards additional to those covering the imagery itself. Finally, the cost of implementation, spanning data licensing, computational infrastructure, and the specialized personnel needed to operate and maintain AI-based monitoring systems, will determine which agencies can adopt the proposed framework without external donor or research-project support, and should be assessed explicitly in any future operational pilot of the framework proposed in Figure 6.

4 Conclusion

This paper reviewed and synthesized recent applications of artificial intelligence and deep learning in remote sensing image analysis for natural resource monitoring, spanning land use/land cover classification, forest and deforestation monitoring, agricultural and crop monitoring, and water resource and flood mapping. The synthesis shows that encoder-decoder segmentation architectures remain the practical backbone for pixel-level mapping tasks, while attention-based transformer models are increasingly competitive or superior for scene-level classification, and that sensor choice, optical for phenology-sensitive agricultural monitoring, SAR for cloud-prone forest and flood monitoring, is strongly shaped by domain requirements rather than by architecture alone. A generalized processing workflow and an integrated cross-domain conceptual framework were proposed to connect these domain-specific findings and to support the design of shared, multi-purpose AI-based monitoring infrastructure. Persistent barriers to operational adoption include scarce labeled data, limited cross-region generalization, high computational cost for transformer and fusion models, and limited interpretability for policy-relevant decisions. Future research should prioritize self-supervised and foundation-model pretraining on large unlabeled Earth observation archives to reduce labeled-data dependence, lightweight architectures suited to deployment in resource-constrained monitoring agencies, standardized multi-domain benchmark datasets that allow fair cross-domain architecture comparison, and explainable AI methods that make deep learning outputs auditable for the legal and policy contexts in which natural resource management decisions are ultimately made.

Acknowledgment

The corresponding author sincerely acknowledges Kabul University and Samangan University for providing academic support and a suitable research environment for completing this study. The first author also appreciates the valuable contributions of previous researchers and publicly available scientific resources that supported the development of this manuscript. AI-based tools were used only for translation from Dari to English and language refinement, and the corresponding author takes full responsibility for the accuracy, originality, and integrity of the entire manuscript.

References

[1] W. Han, X. Zhang, Y. Wang, L. Wang, X. Huang, J. Li, S. Wang, W. Chen, X. Li, R. Feng, R. Fan, X. Zhang, and Y. Wang, "A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 202, pp. 87–113, Aug. 2023, doi: 10.1016/j.isprsjprs.2023.05.032.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770-778.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI), 2015, pp. 234-241.

[4] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis (DLMIA), 2018, pp. 3-11.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481-2495, 2017.

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834-848, 2018.

[7] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data," ISPRS J. Photogramm. Remote Sens., vol. 162, pp. 94-114, 2020.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017.

[9] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.11929, 2020.

[10] Y. Bazi, L. Bashmal, M. M. Al Rahhal, R. Al Dayil, and N. Al Ajlan, "Vision Transformers for Remote Sensing Image Classification," Remote Sens., vol. 13, no. 3, p. 516, 2021.

[11] J. He et al., "TRS: Transformers for Remote Sensing Scene Classification," Remote Sens., vol. 13, no. 20, p. 4143, 2021.

[12] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 10012-10022.

[13] L. Wang et al., "Land Use and Land Cover Classification Meets Deep Learning: A Review," Sensors, vol. 23, no. 21, p. 8966, 2023.

[14] A. Alem and S. Kumar, "Deep Learning for Land Use Classification: A Systematic Review of HS-LiDAR Imagery," Artif. Intell. Rev., vol. 58, 2025.

[15] Frontiers Editorial, "Machine Learning versus Deep Learning in Land System Science: A Decision-Making Framework for Effective Land Classification," Front. Remote Sens., vol. 5, art. 1374862, 2024.

[16] B. Huang, B. Zhao, and Y. Song, "Urban Land-Use Mapping Using a Deep Convolutional Neural Network with High Spatial Resolution Multispectral Remote Sensing Imagery," Remote Sens. Environ., vol. 214, pp. 73-86, 2018.

[17] C. A. Advaith, S. Agrawal, V. Kumar, S. Kumar, and P. S. Tiwari, "Deep Learning Based Land Use Land Cover Classification on Multi-Sensor Remote Sensing Data," ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., vol. X-5/W2-2025, pp. 7-14, 2025.

[18] Y. Quan, R. Zhang, J. Li, S. Ji, H. Guo, and A. Yu, "Learning SAR-Optical Cross-Modal Features for Land Cover Classification," Remote Sens., vol. 16, no. 3, p. 431, 2024.

[19] F. H. Wagner, R. Dalagnol, C. H. L. Silva-Junior, G. Carter, A. L. Ritz, M. C. M. Hirye, J. P. H. B. Ometto, and S. Saatchi, "Mapping Tropical Forest Cover and Deforestation with Planet NICFI Satellite Images and Deep Learning in Mato Grosso State (Brazil) from 2015 to 2021," Remote Sens., vol. 15, no. 2, p. 521, 2023.

[20] J. G. C. Ball et al., "Using Deep Convolutional Neural Networks to Forecast Spatial Patterns of Amazonian Deforestation," Methods Ecol. Evol., vol. 13, no. 11, pp. 2622-2634, 2022.

[21] H. Zhang, M. Lin, G. Yang, and L. Zhang, "ESCNet: An End-to-End Superpixel-Enhanced Change Detection Network for Very-High-Resolution Remote Sensing Images," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 1, pp. 28-42, 2023.

[22] H. Chen, N. Yokoya, and M. Chini, "Fourier Domain Structural Relationship Analysis for Unsupervised Multimodal Change Detection," ISPRS J. Photogramm. Remote Sens., vol. 198, pp. 99-114, 2023.

[23] Anonymous, "RepDDNet: A Fast and Accurate Deforestation Detection Model with High-Resolution Remote Sensing Image," Int. J. Digit. Earth, vol. 16, no. 1, pp. 2013-2033, 2023.

[24] M. Tamaazousti, Y. Zhang, and F. Falchi, "Deep Learning for Deforestation Detection in Multi-Source Remote Sensing Imagery," Remote Sens., vol. 12, no. 7, p. 1085, 2020.

[25] H. Zhi, J. Li, and Y. Zhang, "Real-Time Forest Monitoring Using Deep Learning from Aerial Imagery," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 3705-3715, 2022.

[26] S. Meng, X. Wang, X. Hu, C. Luo, and Y. Zhong, "Deep Learning-Based Crop Mapping in the Cloudy Season Using One-Shot Hyperspectral Satellite Imagery," Comput. Electron. Agric., vol. 186, art. 106188, 2021.

[27] N. Metzger, M. O. Turkoglu, S. D'Aronco, J. D. Wegner, and K. Schindler, "Crop Classification Under Varying Cloud Cover with Neural Ordinary Differential Equations," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1-12, 2022.

[28] M. O. Turkoglu, S. D'Aronco, G. Perich, F. Liebisch, C. Streit, K. Schindler, and J. D. Wegner, "Crop Mapping from Image Time Series: Deep Learning with Multi-Scale Label Hierarchies," Remote Sens. Environ., vol. 264, art. 112603, 2021.

[29] D. Sykas, M. Sdraka, D. Zografakis, and I. Papoutsis, "A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation with Deep Learning," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 15, pp. 3323-3339, 2022.

[30] Z. Sun, L. Di, and H. Fang, "Using Long Short-Term Memory Recurrent Neural Network in Land Cover Classification on Landsat and Cropland Data Layer Time Series," Int. J. Remote Sens., vol. 40, no. 2, pp. 593-614, 2019.

[31] G. Konapala, S. V. Kumar, and S. K. Ahmad, "Exploring Sentinel-1 and Sentinel-2 Diversity for Flood Inundation Mapping Using Deep Learning," ISPRS J. Photogramm. Remote Sens., vol. 180, pp. 163-173, 2021.

[32] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina, "Deep Learning Methods for Flood Mapping: A Review of Existing Applications and Future Research Directions," Hydrol. Earth Syst. Sci., vol. 26, no. 16, pp. 4345-4378, 2022.

[33] X. Zhang et al., "High-Performance Segmentation for Flood Mapping of HISEA-1 SAR Remote Sensing Images," Remote Sens., vol. 14, no. 21, p. 5504, 2022.

[34] F. Pech-May, J. V. Sanchez-Hernandez, L. A. Lopez-Gomez, J. Magana-Govea, and E. M. Mil-Chontal, "Flooded Areas Detection through SAR Images and U-Net Deep Learning Model," Computacion y Sistemas, vol. 27, no. 2, pp. 449-458, 2023.

[35] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2020, pp. 210-211.

[36] M. Wieland, F. Fichtner, S. Martinis, S. Groth, C. Krullikowski, S. Plank, and M. Motagh, "S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies from Sentinel-1 and Sentinel-2 Satellite Images," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 16, pp. 6062-6081, 2023.

[37] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model with Attention," IEEE Access, vol. 9, pp. 14078-14094, 2021.

[38] R. Fan, R. Feng, L. Wang, J. Yan, and X. Zhang, "Semi-MCNN: A Semisupervised Multi-CNN Ensemble Learning Method for Urban Land Cover Classification Using Submeter HRRS Images," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, pp. 4973-4987, 2020.

[39] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, and P. M. Atkinson, "A Hybrid MLP-CNN Classifier for Very Fine Resolution Remotely Sensed Image Classification," ISPRS J. Photogramm. Remote Sens., vol. 140, pp. 133-144, 2018.

[40] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep Feature Fusion for VHR Remote Sensing Scene Classification," IEEE Trans. Geosci. Remote Sens., vol. 55, no. 8, pp. 4775-4784, 2017.

[41] X. Li, C. Wen, Y. Hu, Z. Yuan, and X. X. Zhu, "Vision-Language Models in Remote Sensing: Current Progress and Future Trends," IEEE Geosci. Remote Sens. Mag., vol. 12, no. 2, pp. 32-66, 2024.

[42] C. Huo, K. Chen, S. Zhang, Z. Wang, H. Yan, J. Shen, Y. Hong, G. Qi, H. Fang, and Z. Wang, "When Remote Sensing Meets Foundation Model: A Survey and Beyond," Remote Sens., vol. 17, no. 2, art. 179, 2025.