INTERNATIONAL JOURNAL OF REMOTE SENSING
https://doi.org/10.1080/01431161.2025.2520050

Taylor & Francis
Taylor & Francis Group

Deep learning approaches on Landsat observations: a review and meta-analysis

Zhixin Wang[a], Giorgos Mountrakis[a] and Neal J. Pastick[b]

[a]Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, NY, USA; [b]U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, US Geological Survey, Sioux Falls, SD, USA

ABSTRACT

Although multiple summary reviews of deep learning in remote sensing have been conducted, the field currently lacks a comprehensive analysis of deep learning utilizing Landsat observations in single time pixel-based classification, both qualitatively and quantitatively. This study addresses this gap by conducting a systematic review based on 352 case studies extracted from 279 peer-reviewed articles that supported direct comparisons across deep learning and traditional classifiers. Qualitatively, our findings are as follows: (i) Two dimensional Convolutional Neural Networks (CNNs) remain the dominant methods; however, exploring other architectures is becoming a growing trend. (ii) Most studies focused on single countries (e.g., China, USA, and India) and the last decade, (iii) Sentinel-2 are the most commonly fused sensors, and vegetation indices are the most commonly used spectral indices. The quantitative meta-analysis revealed that Deep Neural Networks (DNNs) offer an approximate 5% improvement for non-DNN accuracy of 80%, with potentially higher improvements for more challenging tasks. By contrasting between popular DNN methods, UNets achieve higher overall accuracy (OA) than CNNs (approximately 10% UNet improvement when CNN OA is at 80%). However, UNets are more challenging to implement because they require labels for large contiguous reference patches, while CNNs, at least at 30m Landsat pixel, pose lower spatial requirements. In another comparison, Transformers achieve higher OA than CNNs (approximately 5% Transformer improvement when CNN OA is at 80%), although the accuracy improvements are smaller compared to the increases associated with UNet benefits. With greater flexibility and more standardized architecture, Transformers can potentially simplify the design process of models. However, Transformers also face challenges having higher computational resource and memory requirements. Increasing availability of reference datasets, further algorithmic developments in the machine learning field, and improvements in cloud-based computing power and data storage, could increase the utility of deep learning methods on Landsat observations.

ARTICLE HISTORY
Received 27 February 2025
Accepted 8 June 2025

Keywords
Deep learning; landsat; convolutional neural network; pixel-based classification; overall accuracy improvement

CONTACT Giorgos Mountrakis gmountrakis@esf.edu Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, 1 Forestry Dr., Syracuse, NY 13210, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/01431161.2025.2520050

1. Introduction

In recent years, deep learning has outperformed traditional machine learning methods making a significant impact in various applications, including image/object classification (Krizhevsky, Sutskever, and Hinton 2012; Liu, Deng, and Yang 2019), semantic segmentation (Chen et al. 2017; Long, Shelhamer, and Darrell 2015), natural language processing (Collobert and Weston 2008; Sutskever, Vinyals, and Le 2014), and speech recognition (Deng and Yu 2014; Zhang et al. 2018). Deep neural networks are characterized by their depth, typically containing multiple hidden layers and several nodes. Although traditional machine learning (ML) methods have been widely used in remote sensing, they face several limitations compared to deep learning methods. ML methods often struggle with high-dimensional data due to the 'curse of dimensionality' (Li et al. 2019); Manual feature engineering of ML methods is time-consuming and may fail to capture complex spatial-spectral patterns (Cheng, Han, and Lu 2017); ML methods are also sensitive to noise and variations, reducing their robustness across diverse conditions (Boonprong et al. 2018); and ML methods lack the capacity to effectively detect and segment complex objects in heterogeneous environments due to their shallow learning structure (Zhang et al. 2023). Compared to traditional machine learning methods, deep learning methods can automatically extract abstract and composite features and deep learning algorithms can also be scaled to handle large datasets and complex problems by leveraging powerful Graphical Processing Units (GPUs) and Tensor Processing Units (TPUs). With increasing reference dataset sizes and types along with advances in algorithms and computing capabilities, efficiencies and capabilities of deep learning have increased. Recently, fast-growing deep learning has received substantial attention from remote sensing researchers, and it has been adopted in numerous tasks such as image pre-processing, classification, and target recognition (Gu, Wang, and Li 2019; Zhu et al. 2017). For example, Mountrakis and Heydari (2023) investigated the benefits of using deep learning for land cover land use classification compared with a traditional classifier of random forest (RF). Makantasis et al. (2015) trained a convolutional neural network (CNN) to extract spectral and spatial features for hyperspectral image classification. Cui et al. (2015) applied a deep belief network (DBN) to synthetic aperture radar (SAR) automatic target recognition with stacked Restricted Boltzmann Machines (RBM).

Since 1972, Landsat observations have been the 'workhorse' of remote sensing. The continuous Landsat archive (U.S. Geological Survey 2024) provides vital global land change information that is not otherwise available. Landsat observations provide a consistent, long-term archive, making it essential for historical LULC change analysis. In addition to its temporal depth, Landsat offers stable radiometric calibration and global coverage, supporting consistent, large-scale monitoring. Although Sentinel-2 data is also widely used due to its higher spatial and temporal resolution, the shorter temporal span (since 2015) limits its applicability for long-term analysis. In the last five decades, the continuous Landsat archive has supported various applications, such as environmental change monitoring, land use tracking, and forest management (Hansen and Loveland 2012; Woodcock et al. 2001). Recent studies have further demonstrated the utility of Landsat data in large-scale environmental applications. For example, Abdali et al. (2023) employed an ensemble of machine learning models within the Google Earth Engine (GEE) framework to classify crop types using multi-temporal Sentinel-1/2 and Landsat-8/9 imagery. Similarly, Zubkova et al. (2024) leveraged long-term Landsat data to analyse changes in fire regimes within protected areas in South Africa. Researchers have also applied deep learning methods to the Landsat archive. For example, Mohajerani and Saeedi (2019) proposed an end-to-end deep learning algorithm using Landsat 8 images for cloud detection. Mountrakis and Shah Heydari (2021) used deep neural networks with Landsat time series observations for continental U.S. land cover mapping. The number of studies using deep learning with Landsat observations is increasing, and a systematic approach to reviewing these studies systemically could help identify existing patterns and uses and to facilitate future research and applications.

Land use and land cover (LULC) mapping is essential for understanding environmental change, urbanization, and resource management. Recent studies have highlighted the diverse applications and implications of LULC analysis. For example, studies have shown that LULC changes, such as deforestation and agricultural expansion, have markedly altered hydrological processes in Ethiopia, including increased surface runoff and reduced infiltration (Birhanu et al. 2019; Tesfaw, Dzwairo, and Sahlu 2023). Likewise, urban-driven LULC changes have been linked to rising land surface temperatures (LST) in Bangladesh, intensifying urban heat island effects and posing challenges for climate resilience (Akter, Gazi, and Mia 2021; Saha et al. 2024). The increasing availability of earth observation data, such as Landsat, has further enabled large-scale and long-term LULC mapping, supporting sustainable planning and policy.

In recent years, several studies have reviewed deep learning in LULC classification using remote-sensing imagery. For instance, Kussul et al. (2017) focused on the classification of land cover and crop types using deep learning methods, while Alem and Kumar (2020) examined land cover and land use classification in remote sensing. Digra, Dhir, and Sharma (2022) provided a comprehensive review of deep learning-based land use and land cover classification in remote sensing. In addition to these reviews, numerous reviews have been conducted examining deep learning in remote sensing with several studies including but not explicitly focusing on Landsat observations. Some reviews surveyed both traditional machine learning and deep learning methods. Sun and Scanlon (2019) reviewed both methods in environmental and water management; Ouhami et al. (2021) summarized various data sources and fusion techniques for plant disease detection; Shirmard et al. (2022) reviewed different data sources in mineral applications. Other reviews focused on deep learning methods using a variety of remote sensing observations. For example, in classification tasks, Vali, Comai, and Matteucci (2020) explored deep learning approaches for land use and land cover classification, and Gu, Wang, and Li (2019) reviewed deep learning methods applied on scene classification and object detection. Neupane, Horanont, and Aryal (2021) reviewed deep learning methods for urban semantic segmentation. In image enhancement, Tsagkatakis et al. (2019) investigated typical deep learning neural network (DNN) architectures, and Wang, Bayram, and Sertel (2022) tested deep learning methods for single image super-resolution tasks based on a multi-sensor dataset. In general applications, Zhu et al. (2017) reviewed common deep learning networks with applications in remote sensing, and Ball, Anderson, and Chan (2017) reviewed different architectures, tasks and challenges of deep learning in remote sensing. Shi et al. (2020) summarized artificial intelligence-based change detection networks in urban studies, natural disasters, and astronomy, and Shafique et al. (2022) reviewed deep learning methods for change detection. Finally, Darwin et al. (2021) surveyed deep learning models in crop yield studies for smart farming.

In some aforementioned reviews, some mentioned Landsat with other remote sensing sensors (i.e. MODIS, SAR, Lidar, UAV) when descripting datasets or data sources (Darwin et al. 2021; Gu, Wang, and Li 2019; Vali, Comai, and Matteucci 2020; Zhu et al. 2017); some reviews included studies using Landsat data, but with traditional machine learning methods (Ouhami et al. 2021; Shirmard et al. 2022; A. Y. Sun and Scanlon 2019). The rest of the reviews included only a limited number of deep learning studies using Landsat observations. For example, Shi et al. (2020) mentioned two studies using deep learning with Landsat for forest change detection. Tsagkatakis et al. (2019) discussed a study with a CNN method using Landsat 7 data (U.S. Geological Survey 2024) for pan-sharpening and another study with DNN to fuse Landsat and MODIS data (NASA 2024). In addition, some of these reviews provided related summary tables. Neupane, Horanont, and Aryal (2021) included one related study for urban LULC with patch size, method, and metric. Wang, Bayram, and Sertel (2022) have one study with method and accuracy information for single image super resolution. In Ball, Anderson, and Chan (2017), four related studies were covered with sensors, data types, methods, and tasks. However, none of the reviews comprehensively and extensively surveyed studies of deep learning using Landsat observations in single-time pixel-based classification tasks. While semantic segmentation approaches have recently gained ground, pixel-based classifications remain prevalent due to the medium spatial resolution of Landsat observations at the approximately 30 m cell size. More importantly, there is currently no comprehensive meta-analysis that quantitatively compares deep learning methods.

This review makes several unique contributions. Our work provides the first comprehensive and extensive survey of deep learning specifically using Landsat data for single-time pixel-based classification tasks. Moreover, this review goes beyond qualitative analysis by conducting a meta-analysis to quantitatively compare the accuracies of deep learning methods to correctly classify land use and land cover. By evaluating the accuracy of recent deep learning architectures, such as Transformers, this study could help guide future research and development in this rapidly evolving field. Here, our objective was to address the following research questions in the context of deep learning using Landsat observations:

- What are the qualitative characteristics of related case studies including time span, study area, input/output data characteristics, methods, and classes of interest?
- Do deep learning methods provide increased accuracy when compared to traditional classifiers and if so, can these be quantified?
- Can incorporating more recent deep learning developments such as Transformer increase accuracies of land cover estimates compared to traditional DNN methods such as CNNs?

2. Methods

2.1. Peer-reviewed article selection

Peer-reviewed article selection Peer-reviewed article identification used the Scopus database (Elsevier 2024), which provides comprehensive citations and abstracts of peer-reviewed literature in various science fields including remote sensing. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method was implemented for peer-reviewed article selection (Moher et al. 2009). The peer-reviewed article search included peer-reviewed articles published up to July 2024. To ensure a comprehensive search, numerous deep learning architectures were also included in the detailed search criteria applied on Scopus: ('Landsat' AND ('deep learning' OR 'deep neural network' OR 'convolutional neural network' OR 'recurrent neural network' OR 'long short-term memory network' OR 'generative adversarial network' OR 'autoencoder network' OR 'fully convolutional network' OR 'restricted boltzmann machine' OR 'deep belief network' OR 'multilayer perceptron')). Initially, a total of 805 documents were returned. These peer-reviewed articles were manually scanned to remove unsuitable articles using the following rules:

- Publications not in English as automated translation would not provide sufficient confidence to understand the particularities of each non-English scientific article for proper inclusion.
- Publications not in the remote sensing field.
- Publications that did not include Landsat data.
- Publications that used Landsat data but did not incorporate deep learning methods.
- Publications that used Landsat data but did not focus on single-pixel classification tasks.

The detailed selection steps are provided in Figure 1. At the end of the selection process, 279 peer-reviewed articles were identified and analysed, comprising 261 comparison peer-reviewed articles (having deep neural networks (DNNs) that are contrasted to other DNN and traditional classifiers). It is important to emphasize that to facilitate proper comparisons between DNN and non-DNN methods we only included studies here that have tested various methods on the exact same dataset. Each of these peer-reviewed articles was manually examined, and relevant metrics were extracted. A database with 27 fields was constructed to assist with our meta-analysis. In addition to general peer-reviewed article information such as title, authorship and journal, the database included Landsat bands used for input, spectral indexes, deep learning methods, and number of model parameters. Fields capturing accuracy metrics such as overall accuracy (OA), F1 score (F1), user accuracy (UA) and producer accuracy (PA) were also included to support quantitative analysis of deep learning accuracy performance. A single peer-reviewed article could yield multiple comparison entries in our database for each study area, which led to 352 case studies from 279 selected peer-reviewed articles. Furthermore, a single-case study could lead to multiple database entries, each representing a single direct DNN vs. traditional classifier or DNN comparison. At the end, 925 such comparison entries were extracted. Also, because not all

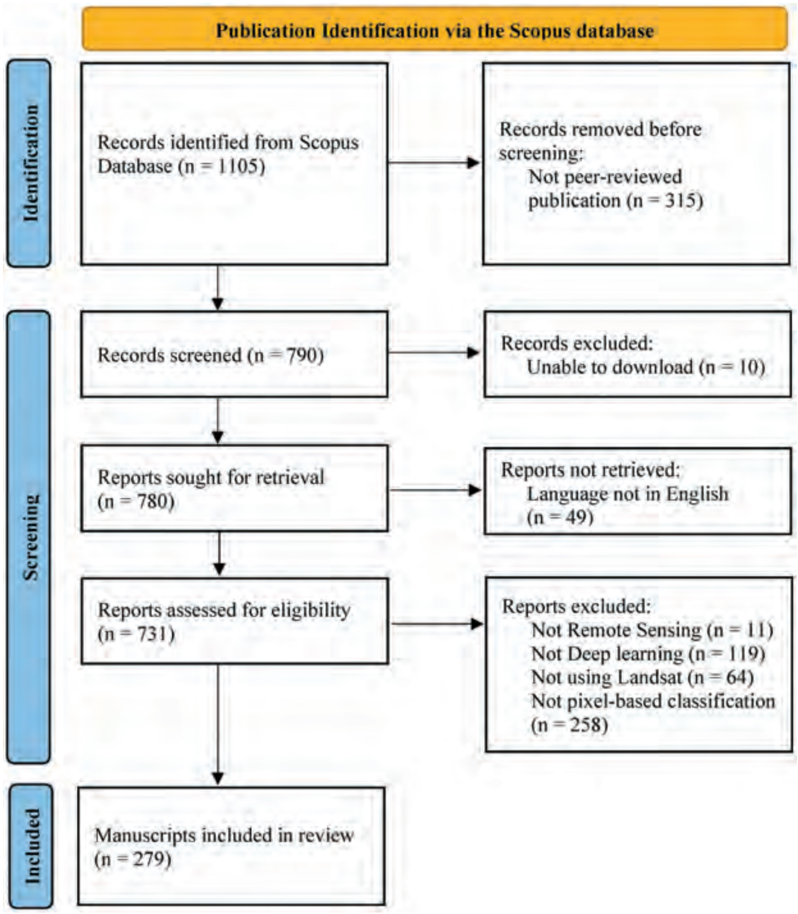

Figure 1. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram for systematic review of deep learning using Landsat in remote sensing.

papers used the same accuracy metrics, we selected metrics frequently used by most studies (OA) for subsequent quantitative analysis.

2.2. Quantitative accuracy analysis

In addition to descriptive statistics, we conducted a quantitative analysis comparing DNN with non-DNN methods. The DNN methods were grouped into CNN, UNet, CNN + RNN, and Other. The CNN category encompasses both CNN architectures featuring convolutional layers, pooling layers, and fully connected layers, and widely used pretrained models like the Residual neural network (ResNet), the Visual Geometry Group neural network (VGGNet), and the DeeplabV3+ network. UNet is distinguished from the broader CNN category due to its frequent application. CNN + RNN models combine spatial and temporal features in their architecture, with a particular focus on CNN and Long Short-Term Memory (LSTM) networks. The Other category contains the remaining methods, for example, Transformers, Deep Belief Networks (DBN) and Stacked Autoencoder (SAE)/

Autoencoder (AE). (The introductory peer-reviewed articles for the described methods are listed in Appendix C).

Here, we focus on the Single-Time Pixel-Based Classification task, meaning labelling of individual pixels at a given time, producing outputs at that time regardless of the input temporal depth. This includes land cover classifications and binary classifications such as cloud detection, water/ice/glacier detection. We could not conduct further analysis on different classifications distribution (e.g. land cover classification vs. binary classifications) because there were not enough samples based on high-quality validation datasets to support such analysis (refer to section 3.2). The input and output spatial characteristics were also included in accuracy comparison. Three distinct categories are identified: pixel input, pixel output (pixel in, pixel out); patch input patch output (patch in, patch out); and patch input, pixel output (patch in, pixel out). Here, the pixel assumes a lack of neighbouring information as input and individual pixel estimation as output. The patch refers to inclusion of neighbouring pixels as input and concurrent estimation of multiple neighbouring pixels as output. An additional dimension of comparison included the temporal dimension of inputs: multi-temporal input and mono-temporal input, depending on whether single or multiple time inputs were incorporated.

3. Results and discussion

3.1. Descriptive analysis

3.1.1. General content summarization via scientometric mapping

To organize peer-reviewed articles in thematic areas and identify possible connections, a scientometric mapping of the keywords cooccurrence network based on titles and abstracts was created using the VOSviewer software tool (Figure 2). Circle size represents frequency of occurrence, and lines thickness corresponds to connection strength. Some

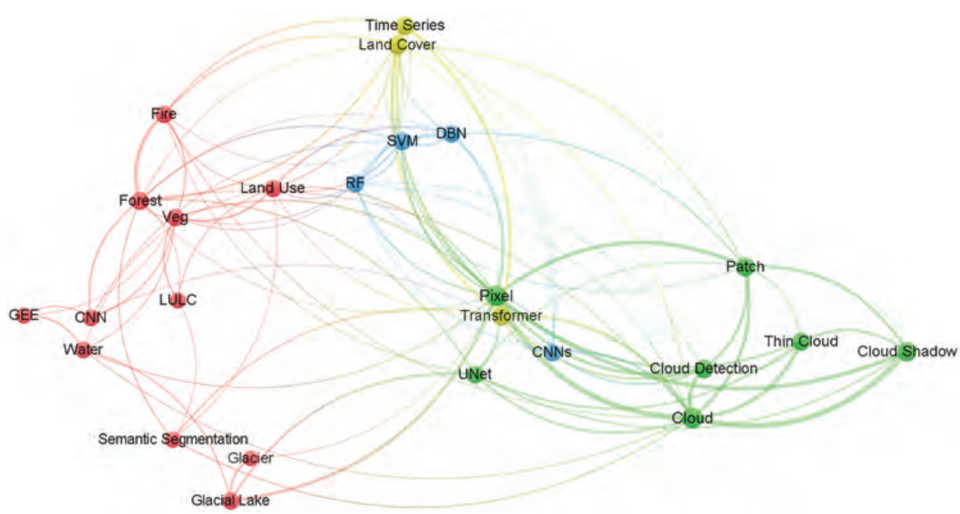

Figure 2. Scientometric mapping using titles and abstracts of papers with deep learning using Landsat data in single-time pixel-based classification. GEE refers to Google Earth Engine, and veg refers to vegetation.

terms, such as the frequent measurement metrics, classification accuracy (i.e. overall accuracy (OA), F1 score, and study area are excluded in the mapping for clarity purpose. In our analysis, semantic segmentation is seen as a special case of classification. Semantic segmentation involves dense, spatially coherent labelling of all pixels, often emphasizing object boundaries, whereas classification may assign labels per pixel without considering spatial context. Semantic segmentation proximity to Glacier Lake indicates a high interaction. Keywords of 'LULC' (land use land cover), 'Land use' and keyword of 'Land cover' indicate that there are many land use land cover classifications. Forest/Fire/Vegetation (Veg), Glacier/Glacial Lake, and Cloud/Cloud Detection/Thin Cloud/Cloud Shadow are closely grouped, as expected. These applications, in addition to water, are main applications based on classification tasks in the reviewed peer-reviewed articles. Keyword of 'Time series' being close to 'Land cover classification' is showing many applications make use of the long temporal coverage of Landsat data. Patch input is highly connected to cloud detection, while pixel input is relatively in the centre of all applications. This suggests that more applications use pixel inputs/outputs than patch inputs/outputs.

For the methods or models used in the studies for these applications, Convolutional Neural Network (CNN), UNet, and transformer are popular deep learning architectures. Random forest (RF) and support vector machine (SVM) are the main traditional machine learning methods compared with deep learning methods. The frequent occurrences of CNNs and UNets are due to their support for image classification, especially for pixel-based classification. Transformer, compared to CNNs, is a recent method achieving state-of-the-art results on various image classification tasks, by utilizing self-attention mechanisms to effectively capture long-range dependencies effectively. A full list of key citations for each identified DNN method is provided in Appendix C.

3.1.2. Peer-reviewed article publication frequency over time

The distribution of peer-reviewed articles across time and deep learning methods is presented in Figure 3. The exponential increase of DNN methods using Landsat data in the last five years before 2022 is depicted. The lower number of publications in year 2023 is surprising; however, the number of 2024 publications shown in Figure 3 is only tallied until July, indicating that the total for the entire year 2024 will likely be higher than the total for 2023 and comparable to the total for 2022. Looking at DNN methods, four categories are examined. Beyond CNN, UNet and transformer, the CNN + RNN method category contains peer-reviewed articles with CNN and RNN, especially CNN and LSTM methods. The Other method category includes DNNs such as DBN and MLP. The early prevalence of CNN methods is clearly shown in Figure 3. Although CNNs have continued to be the primary deep learning approach used in the last few years, their usage, while dominant, has not seen the relative increases of UNet and Other methods. This is expected as the field evolves to expand to state-of-the-art DNN methods, partially developed in other fields (e.g. Transformers).

3.1.3. Geographic and temporal distribution of peer-reviewed article study areas

The temporal and spatial distribution of the included case studies is shown in Table 1. The table considered 11 five-year intervals in the past five decades, with values representing the number of case studies during each period. Of the 352 case studies, 245 contained both temporal and spatial information and were

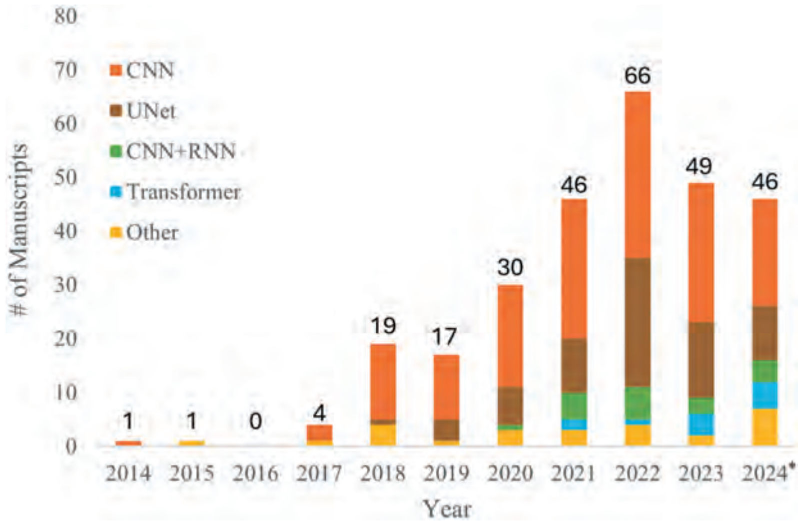

Figure 3. Landsat/deep neural networks (DNNs) peer-reviewed articles across time and methods. 2024* represents data collected before July 2024.

included. Global and continental studies are at the top rows. Most global studies were recent studies focusing on cloud detection, followed by water studies. This could be partially attributed to the availability of public global cloud reference datasets (e.g. SPARCS Cloud Dataset (Hughes and Kennedy 2019), 38-Cloud (Mohajerani, Krammer, and Saeedi 2018), 95-Cloud (Mohajerani and Saeedi 2021)) and the relative ease of water reference data creation compared to multi-class reference datasets.

Figure 4 shows a world map of case studies number in the 35 countries, among them, China, the U.S.A. and India are the most frequently studied countries. The prevalence of China and the U.S.A. can be attributed to their large size, strong scientific interest, and funding priorities (Hu 2020; X. Wang et al. 2012). Similarly, India also has large size, and diverse environmental and developmental needs supporting scientific research (Singh et al. 2024). Despite the long record of Landsat observations starting in the early 1970s, most studies concentrated on the last decade. An increase after the early 1980s could be attributed to Landsat 4 launch that with the Thematic Mapper (U.S. Geological Survey 2024) offers consistent observations with later years. Another increase in the mid-2010s could be related to the availability of Landsat 8 data (after 2013) (U.S. Geological Survey 2024) and complementary datasets from Sentinel 1 and 2 (European Space Agency, 2024a, 2024b) launched in 2014 and 2015, respectively. A peer-reviewed article reference list for each country is provided in Appendix A to support researchers interested in specific countries.

3.1.4. Additional case study characteristics

The 352 identified case studies are further summarized in Figure 4. First, the studies were organized depending on the input and output spatial types. The label patch refers to inclusion of neighbouring pixels as input and concurrent estimation of

Table 1. The distribution of identified case studies ($n = 245$) across time and country.

Years / Country	70-75	75-80	80-85	85-90	90-95	95-00	00-05	05-10	10-15	15-20	20-23
Global						2	1		7	9	2
Africa											1
Antarctica									1	1	1
Afghanistan								1	1	1	
Algeria							1	1	1	1	1
Australia									1	5	
Bangladesh									1	1	1
Brazil									4	6	1
Canada	1	1	1	1	1	1	2	3	3	3	1
China			2	3	4	6	9	16	29	41	9
Denmark				1	1	1	1	1	1	1	
Egypt								1	1	2	1
France									2	2	
Germany							1	1	1	1	1
Greece											1
Greenland	1	1	2	2	2	2	2	2	2	3	1
India				1	2	2	3	4	7	13	3
Indonesia										3	3
Iran				1	1	1	1	2	3	3	2
Iraq							1		1		1
Italy										2	
Japan										1	
Jordan									1	1	
Mexico										1	
Mongolia										2	2
Morocco										1	1
Nepal					1	1	1	1	3	2	
Netherlands										1	
Pakistan											2
Russia											1
Saudi Arabia					1	1					
South Korea									1		
Spain							1			2	
Thailand										2	1
Turkey								1	1	1	5
Ukraine					1				1	1	
USA				3	4	6	8	11	16	15	1
Vietnam									1	1	
Total	2	2	5	12	18	23	32	45	90	129	43

Africa and Antarctica are continents, not countries. Continents were used because of the relative paucity of studies in these areas.

multiple neighbouring pixels as output. The label pixel assumes a lack of neighbouring information as input and individual pixel estimation as output. More than half of the studies fall in the patch in pixel out category (59%), followed by patch input patch out (28%) and pixel in pixel out (13%). Patch in pixel out strong presence is attributed to the prevalence of CNN methods for pixel-based classification. Patch in patch out becomes popular in certain CNN studies using up sampling techniques like UNet, fully convolutional networks (FCN) (Long, Shelhamer, and Darrell 2015), DeepLabV3+. Pixel in pixel out has fewer studies due to potential loss of valuable spatial information.

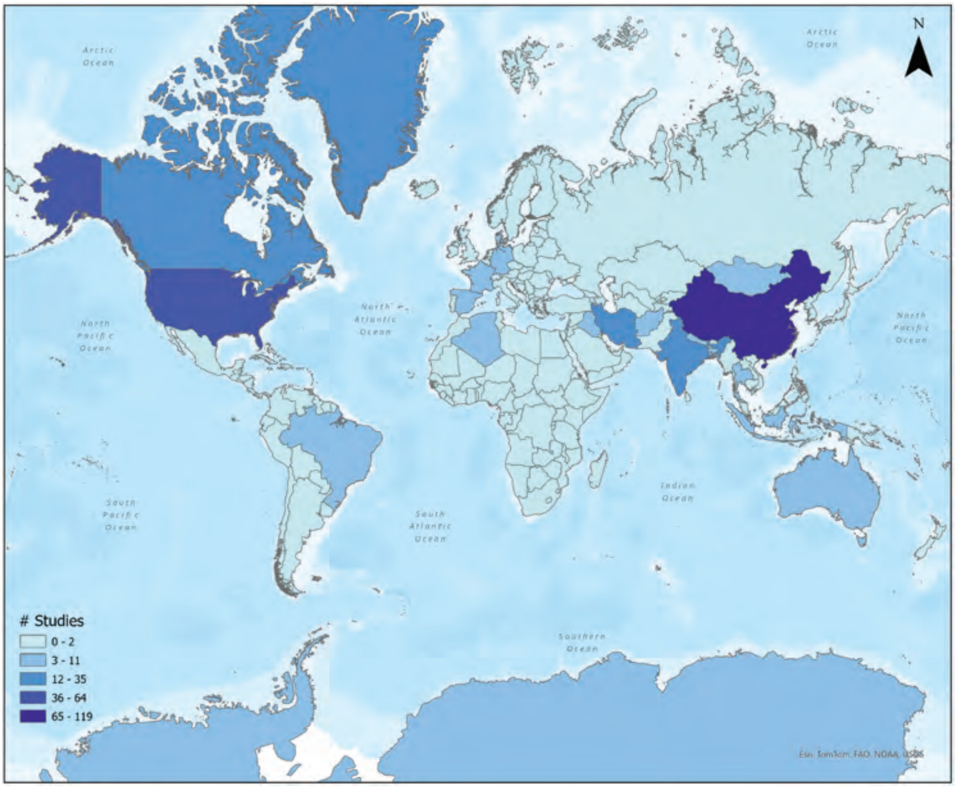

Figure 4. World map of case study counts by country. Note that map projection may distort area and distance.

The multi-temporal input (83%) has a considerably stronger presence compared to mono-temporal input studies (17%; Figure 5). All case studies targeted a single time classification output, but the temporal input length could vary. Most studies used multi-temporal input data because the high Landsat revisiting time allowed the studies to incorporate diverse and unique temporal features. Mono-temporal input studies are typically constrained to applications that cannot incorporate temporal information, such as cloud detection. Single country studies (94%) dominated the study area extent, followed by global (5%) and multi-country (1%).

Our meta-analysis indicated that supervised learning was the most common learning strategy (97%), which we expected based on how we filtered case studies for classification tasks. Conversely, unsupervised learning (2%) was used in a few applications like desert detection using generative adversarial networks (GAN) (Goodfellow et al. 2020). Other learning strategies labelled as hybrid (1%) of semi-supervised learning and self-supervised learning are currently used much less than supervised learning. Hybrid strategies reduced the labelling burden but often require more sophisticated models and training procedures. However, use of hybrid strategies may increase in the future.

When examining the incorporated spectral indices, the majority are designed for vegetation studies and include Normalized Difference Vegetation Index (NDVI, 43%),

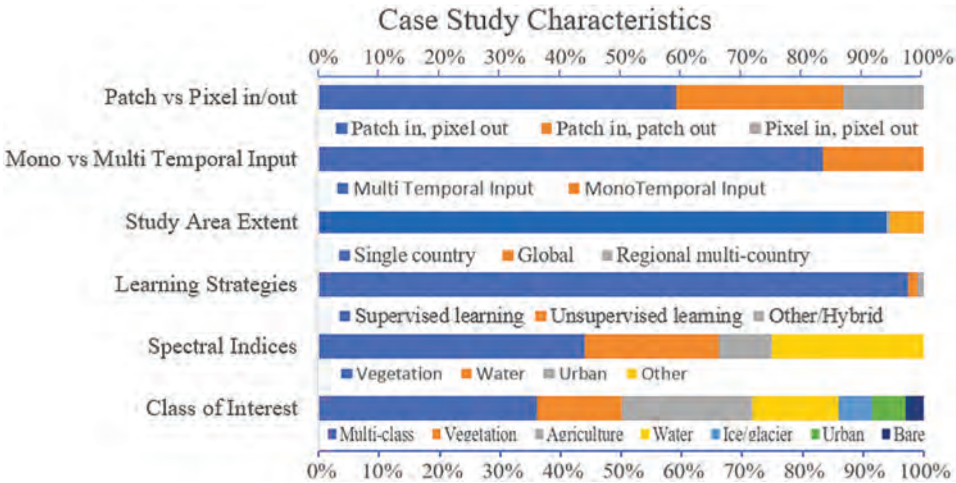

Figure 5. Additional characteristics of case studies (n = 352).

Normalized Difference Water Index (NDWI, 13%), Enhanced Vegetation Index (EVI, 9%), and Normalized Difference Snow Index (NDSI, 3%). We attribute the prevalence of NDVI to its simplicity, robustness, and diverse applications in vegetation and agriculture. Approximately three-quarters of case studies omit spectral indices because DNNs can extract features automatically from the spectral band data. However, proper indices can still help in highlighting specific known features, and reduce noise, and reducing overfitting.

The distribution of land cover/use class indicates that the multi-class has the highest proportion (37%), followed by agriculture (21%), vegetation (14%), and water (14%). Ice/glacier, urban and bare/mining classes account for smaller proportions (6%, 5% and 3%, respectively). The prominent multi-class category is mainly attributed to land cover classifications. Although multi-class products were the most widely used (e.g. Dynamic World, ESA global products), they only compose about one-third of the research conducted.

3.1.5. Landsat fusion with additional observations/data

Sensors fused with Landsat data and the corresponding tasks are shown in Figure 6. Fusion refers to the process of integrating data from multiple sources, sensors, or modalities to create a more comprehensive and enhanced representation or understanding of the observed phenomenon or area of interest. Fusion can be defined at various levels, including multiscale analysis, pixel, feature, decision, model, and hybrid. The pixel level fusion, which directly combines pixel values from multiple sensors or sources, was the most common fusion process in the reviewed peer-reviewed articles. Sentinel-2 is frequently fused with Landsat due to its data availability, spatial and temporal resolution similarity, and ease of processing. Sentinel-2 has a relatively high spatial resolution (10 m) and is suitable for various tasks because of several ongoing harmonization (Claverie et al. 2018) efforts with Landsat observations it is suitable for various tasks. The high spatial resolution optical satellites are also fused with Landsat due to their high spatial resolution (~1 m) (Li et al. 2020). Sentinel-1 is the primary radar sensor fused with Landsat

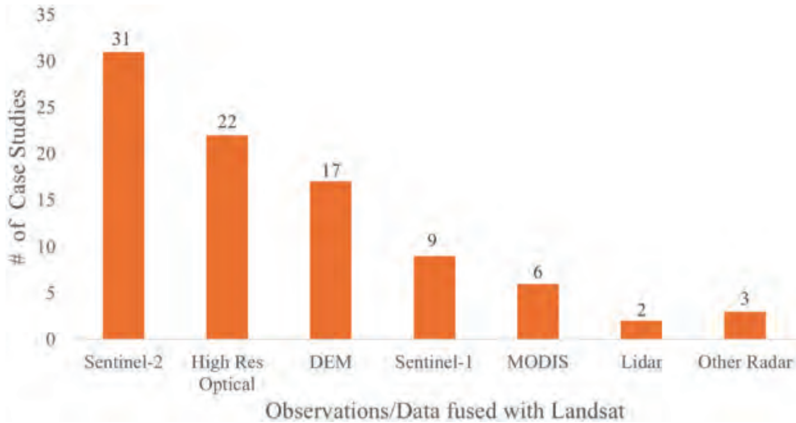

Figure 6. Number of case studies fusing Landsat with other sensors/data. High res optical refers to high resolution optical satellites (e.g. WorldView (Vidal 2008), RapidEye (Tyc et al. 2005) and Gaofen (Chen et al. 2022)). Lidar includes both spaceborne and airborne sensors (Wandinger 2005). Other radar consists of envisat (Louet and Bruzzi 1999), ALOS-1 (Iwata et al. 2007) and ALOS-2 (Kankaku, Suzuki, and Osawa 2013).

complementing optical observations with the unique radar characteristics such as all-weather imaging and penetration abilities (Carrasco et al. 2019). The high cost of Lidar and other radar, complex processing, and limited coverage restrict their fusion with Landsat (Raj et al. 2020). MODIS has a low spatial resolution (250 m-1 km), but it has a high temporal resolution and free long-term global coverage since 1999 (Justice et al. 2002). Finally, digital elevation model (DEM) data are often used together with Landsat to improve classification accuracy by providing additional topographic features such as elevation, slope, aspect and terrain roughness (Eiumnoh and Shrestha 2000).

3.1.6. Distribution of DNN methodologies

The distribution of methods for the single-time pixel-based classification is shown in Figure 7. CNN methods are widely used with various architectures such as Fully Convolutional Networks (FCN) and DeepLabV3+. FCN (Long, Shelhamer, and Darrell 2015) is commonly used in pixel-based classification due to its dense pixel predictions, efficient feature extraction and adaptability. DeepLabV3+ (Chen et al. 2018) is a state-of-the-art semantic segmentation model that extends DeepLabV3 by improving object boundary localization and detailed feature recovery. UNet is often paired with CNN due to its effectiveness in classification, particularly for small datasets, with efficient feature extraction, reduced overfitting and adaptability (Yuan, Shi, and Gu 2021). Transformers have recently demonstrated significant advantages in remote sensing tasks due to their ability to model long-range dependencies and capture global contextual information (Dosovitskiy 2020; Liu et al. 2021). Unlike CNNs, operating with fixed receptive fields, transformers leverage self-attention mechanisms to dynamically weigh the importance of all elements in the input, enabling a more flexible and comprehensive feature representation (Vaswani 2017). This is particularly beneficial for remote sensing imagery, where spatial patterns and relationships can span large areas. Also, transformers can better handle heterogeneous landscapes, complex object boundaries, and irregular shapes

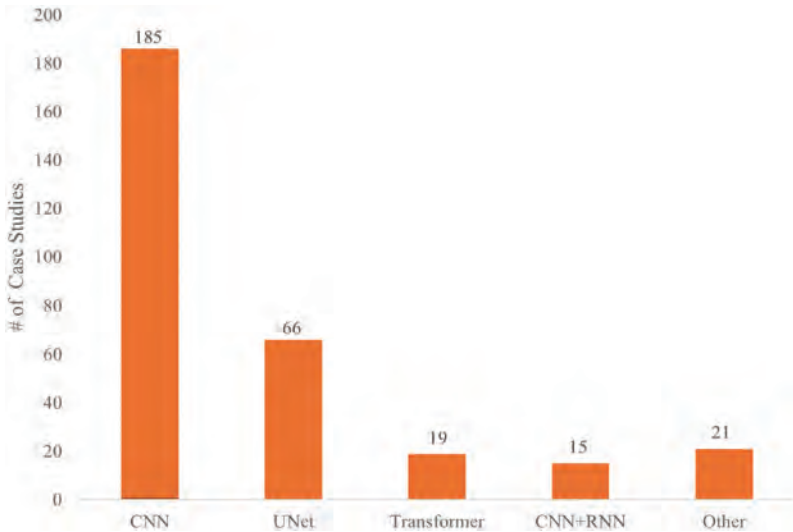

Figure 7. The distribution of methods and tasks of case studies.

common in LULC classification (Guo et al. 2022; Niu, Zhong, and Yu 2021) CNN with RNN, especially CNN and LSTM, are less widely used, possibly because of the complexity of model design and data processing for long time series.

3.2. Quantitative analysis

3.2.1. DNN accuracy improvements compared to non-DNN methods

To answer the question whether DNN methods outperform non-DNN methods, Figure 8 depicts the overall accuracy (OA) of DNNs compared to non-DNN methods across the multi-class category and three prevalent binary classifications for agriculture, vegetation, and water. The extracted accuracies refer to independent validation datasets; no training accuracies are included. Our analysis included only comparisons that were based on high-quality validation datasets that were either manually labelled by researchers involved in their study or sourced from publicly available validation datasets that have undergone rigorous verification processes. Case studies with validation datasets directly extracted from model derived LULC maps or created by automated algorithm processes were excluded. The trendlines are fitted using robust regression (Theil-Sen regression (Dang et al. 2008)) to limit the effect of outliers.

Figure 8 illustrates the accuracy improvements resulting from incorporating DNN methodologies into Landsat studies. Increases in accuracy improvements following incorporation of DNN methodologies vary, reaching up to approximately 10% for non-DNN case studies with low starting OA. Using the trendline as a guide, accuracy would improve from 80% to 85% following incorporation of DNN methodologies (more details in appendix B). Although the statistical sample per class was not sufficiently large to extract quantify conclusions, accuracies seem to increase the most for the agricultural class following incorporation of DNN implementations than other methodologies. This increased accuracy could be attributed to the stronger spatial structure of the agricultural

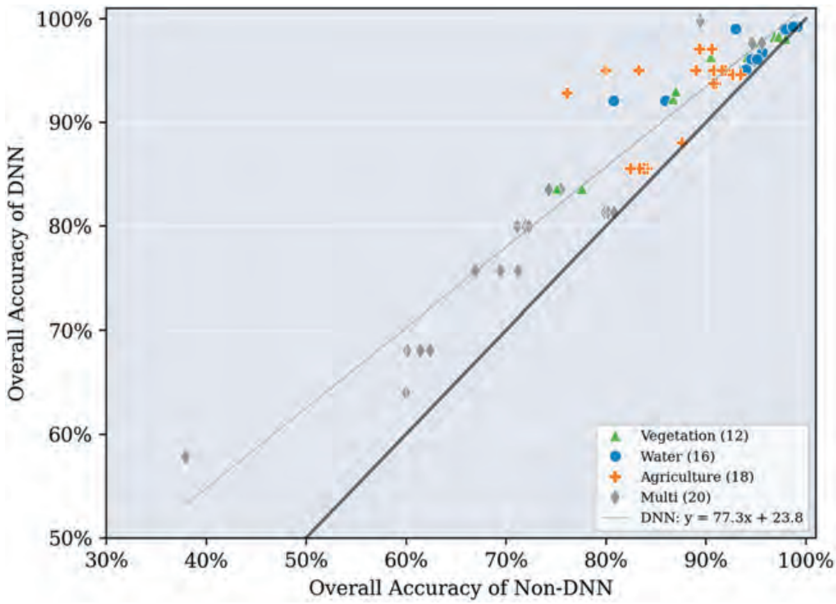

Figure 8. Comparison of DNN vs non-DNN classifications across land cover/use classes in single-time pixel-based classification. DNN includes CNN, UNet and transformer methods. Numbers in parentheses are supporting samples for each class. The black line represents the line of equality.

class, a structure that DNNs are better at capturing (Darwin et al. 2021). Alternatively, the increased accuracies for the agricultural class could also be the result from capturing distinct temporal transitions from crops that may not be as clearly present for the vegetation or other classes as all studies on the agricultural class incorporated temporal features.

3.2.2. DNN accuracy comparison between the UNet and CNN methods

The wide use of CNN and UNet methods facilitates direct comparisons of their accuracies across different classes. We concentrated on UNet methods from other semantic segmentation methods due to their popularity and their implementation in the 2023 Annual National Land Cover Database product in the United States. Figure 9 compares UNet and CNN within the multi-class category and four binary classes: water, agriculture, vegetation and urban. The CNN group includes the following methods: FCN (13), DeeplabV3+ (7), CNN (5), ResNet (1), SegNet (Badrinarayanan, Kendall, and Cipolla 2017) (3). More details are provided in appendix B. The extracted accuracies also refer to independent and high-quality validation datasets. Each sample represents an OA comparison between UNet and CNN within a study using identical input features; therefore, differences could be directly attributed to the usage of the two methods. Except for one outlier each in the multi-class and water classification task, the accuracies were consistently higher using UNet compared to CNN, as evidenced by the trendline. For example, for a CNN overall accuracy of 80%, the corresponding UNet accuracy is expected to be around 90%. Increases in accuracy incorporating UNet methodologies are more

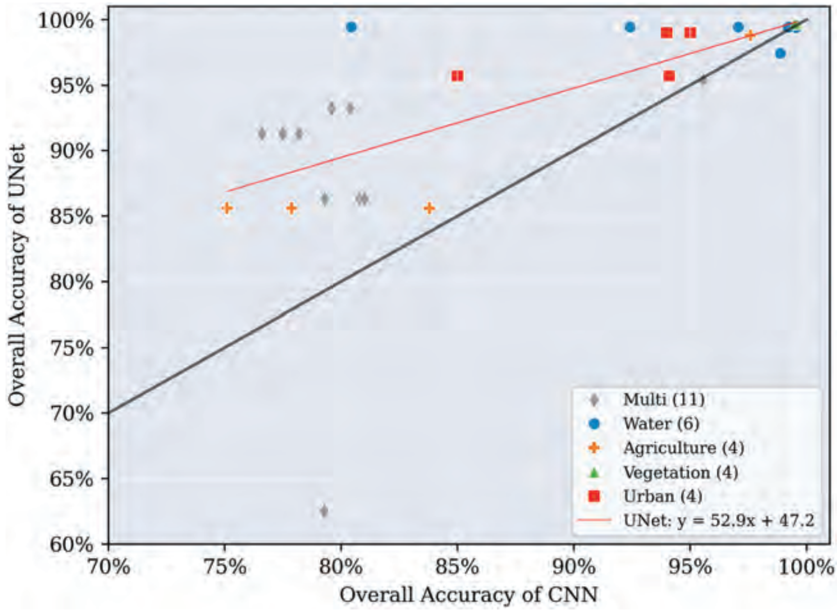

Figure 9. Overall accuracy of UNet vs CNN with class of interests and trendline. Numbers in parentheses are supporting samples for each class category. The black line represents the line of equality.

evident in underperforming CNN case studies suggesting that the UNet could offer considerable benefits when needed the most. From an architectural perspective, increased accuracies associated with UNet could be attributed to skip connections, which directly concatenate feature maps from earlier layers with those from later layers in the network to leverage both high-level information from deeper layers and low-level spatial details from earlier layers. The UNet skip layers help capture both global and local context. In contrast, CNN extracts only local patterns through their receptive fields.

Figure 10 compares accuracies of UNet and CNN methods exclusively on cloud detection. The CNN group includes FCN (8), CNN (9), ResNet (2), DeeplabV3+ (2), SegNet (1), CloudNet++ (Mohajerani and Saeedi 2021) (2). More details are available in appendix B. Using both methods, OAs were considerably higher for cloud detection than OAs for land cover/use case studies. Most OAs of UNet and CNN for cloud class exceed 90%, suggesting both models can classify clouds effectively. UNet offers a small accuracy benefit of approximately 2% for corresponding CNN OA of 90%.

Overall, Figures 9 and 10 indicate that UNet outperforms CNN in single-time pixel-based classification tasks. This superior performance can be attributed to UNet's encoder – decoder architecture with skip connections, enabling better spatial localization and preserving fine-grained details. However, UNet requires labels for large patches (e.g. 50×50 pixels), while CNNs, at least at our operating 30 m Landsat pixel, had lower spatial requirements (e.g. 7×7 pixels) and thus could support increased spatial independence of reference samples.

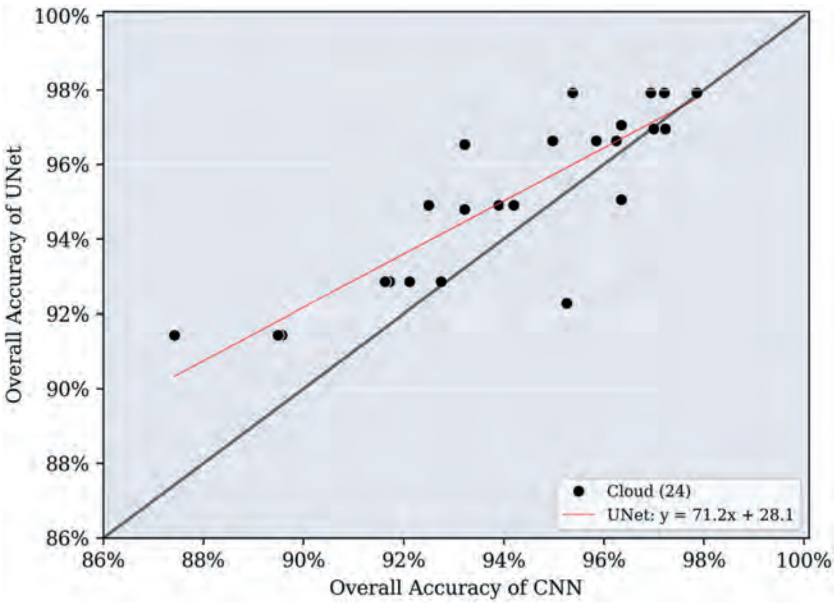

Figure 10. Overall accuracy of UNet vs CNN with cloud class and trendline. Number in parentheses is supporting samples for cloud category. The black line represents the line of equality.

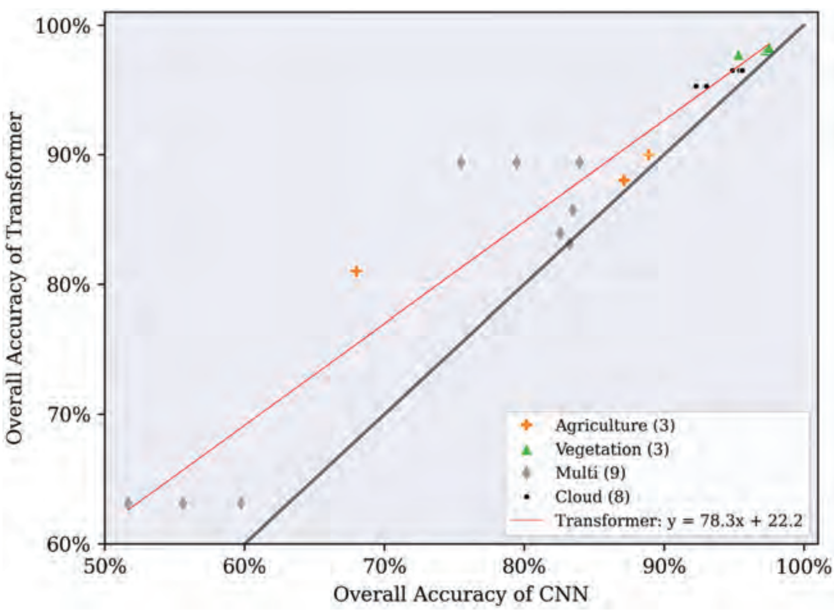

Figure 11. Overall accuracy accuracies of transformer vs CNN methods. Numbers in parentheses are supporting samples. The black line represents the line of equality.

3.2.3. DNN accuracy comparison between the transformer and CNN methods

Transformers are recent machine learning methodologies that are being increasingly used in remote sensing tasks. Here, we compare Transformers to CNN methodologies to assess if Transformers produce higher accuracies than CNNs. Due to the lower number of case studies than the previous UNet vs CNN comparison, here the land cover/use case studies were grouped with the cloud studies. The extracted accuracies again refer to independent and reliable validation datasets. Each sample represents an OA comparison with identical input features used. The architecture of Transformers is based on the self-attention mechanism, which is different from convolution kernels or filters used by CNNs; therefore, CNNs need adaptations to process certain input data. For example, when using a CNN with pixel time series input data, 1DCNN should be used to handle the temporal features instead of the typical 2DCNN used for spatial relationships (Mohajerani and Saeedi 2021; H. K. Zhang, Luo, and Li 2024). Conversely, for image or patch inputs, 2DCNN should be utilized to process the spatial features. For Transformers, temporal features of pixel time series data can be integrated through positional encoding and multi-head attention layers. For image or patch input data, the vision transformer, an adaptation of traditional transformer with patching, linear projection, and positional encoding, should be utilized to process the spatial features of input data (J. Li and Wang 2024; Z. Liu et al. 2021; Panboonyuen et al. 2021). To compare the two DNN models fairly, Figure 11 includes only comparisons where the Transformer and CNN methods were evaluated under the same feature process abilities (i.e. 1DCNN and traditional transformer for pixel time series input data, 2DCNN and vision transformer for image or patch input data).

Figure 11 shows that Transformers produced higher OAs than CNNs in the single-time pixel-based classification task (85% OA compared to 80% OA for CNN, more details in appendix B). Transformers utilize self-attention mechanisms instead of using sequence aligned convolution to allow each part of the input to attend to every other part, which enables a global understanding, that captures long-range dependencies effectively (Guo et al. 2022; Niu, Zhong, and Yu 2021; Vaswani 2017). CNNs capture local patterns effectively using receptive fields but may not capture global understanding as effectively compared to transformers in global understanding (Dosovitskiy 2020; Maurício, Domingues, and Bernardino 2023; Raghu et al. 2021). Transformers have greater flexibility compared to fixed grid-like processing of CNNs by adjusting the number or size of patch tokens. With a more standardized architecture, Transformers have a lower need for task-specific architectural engineering and can potentially simplify the design process. Therefore, Transformers have the potential to be applied more widely to remote sensing tasks. However, Transformers also face challenges as they typically require more computational resources and memory compared to CNNs, especially for large datasets, due to the quadratic complexity of the standard self-attention mechanism (Hua et al. 2022; Keles, Wijewardena, and Hegde 2023; Tang et al. 2022; Vaswani 2017).

4. Discussion

While this review provides a comprehensive overview of deep learning applications on Landsat observations, focusing on single-time pixel-based classification, several limitations should be acknowledged. The scope of quantitative analysis is limited to studies that primarily utilize CNNs, transformer-based models, and UNets, which may exclude other relevant but less widely adopted techniques. Additionally, the review emphasizes peer-reviewed articles, potentially overlooking relevant findings from grey literature or non-English publications. Finally, the fast development of deep learning methods indicates that some recent advances may not be fully represented at the time of writing. For example, recent studies have demonstrated the strong potential of vision transformers (ViTs) in remote sensing image classification, especially for hyperspectral imagery, where they often outperform traditional CNN-based models. Roy et al. (2025) introduced SimPoolFormer, a vision transformer architecture incorporating a simplified pooling mechanism, achieving competitive accuracy while reducing computational costs compared to complex ViT models. Similarly, J. Liu et al. (2023) explored an ensemble learning strategy using multiple ViT models, showing that such combinations can further enhance classification performance in hyperspectral imaging tasks. These findings underscore the growing relevance of vision transformer approaches in advancing the remote sensing field.

Another inherent limitation of this review, as with any systematic literature search/meta-analysis, is the potential presence of publication bias. Our review did not identify clear instances where non-DNN methods produced higher accuracies than DNN methods across the various tasks analysed; however, this may reflect a publication bias towards more novel or currently popular methodologies rather than a comprehensive assessment of the best tool for each task. Consequently, our review might disproportionally represent successful applications of DNNs, skewing the perceived effectiveness and applicability of the models. As such, we acknowledge that there are scenarios in which traditional ML algorithms might outperform DNNs. This is particularly true in contexts where training data are scarce for supervised learning problems and where domain-specific feature engineering plays a critical role in model performance. For example, R. Singh, Biswas, and Pal (2022) found that XGBoost models outperformed ResNet and complex CNNs for cloud masking of Sentinel 2 data when provided with hand-crafted features (89.86% F1 for XGBoost compared to 89.75% for ResNet and 88.70% for CNN). Likewise, Jeong et al. (2022) reported that an RF regressor outperformed a DNN regressor for estimating the leaf area index (LAI) of paddy rice using MODIS and climate data in two rice growing areas of South Korea (0.45 mean RMSE for RF and 0.63 for DNN over four years).

Direct comparisons between different studies are not valid, as the classification tasks in one study may be easier than in another. Therefore, we only include studies that evaluated multiple methods on the same dataset, extracting and reporting their comparisons accordingly. Although conducting separate analyses across different classification schemes is interesting, the limited number of case studies prevents us from drawing more specific conclusions beyond the classification scheme level reported by the original study authors. The limited number of case studies partially explains why we report both accuracies for DNN and non-DNN methods as a reference point, instead of simply reporting DNN improvements compared to non-DNN accuracies (the difference value).

In assessing the methodologies used across the reviewed studies reviewed, we noted a varied degree of reporting of optimization processes, computational considerations, preprocessing steps, and consistent evaluation criteria. Optimization processes were more varied than others, and few studies provided clear and detailed information about optimization processes. For example, Boonpook et al. (2023) used optimal hyperparameters such as learning rate and momentum for each supervised learning model when comparing different DNNs in land use and land cover classification. On the other hand, Lee et al. (2020) in land cover classification mentioned that RF and SVM were used in land cover classification without providing more information on the optimization process; for example, there is no information on the typical grid search for hyperparameter selection or testing of different learning rates for CNNs. L. Liu et al. (2022) compared RF, SVM, and DNN using predefined parameters without discussing optimization processes for pine mapping. More extensive sharing of optimization processes that match algorithmic choices could facilitate more nuanced and equitable comparisons of model performance because algorithmic choices can significantly affect the obtained results.

Relying solely on overall accuracies, our primary comparison metric, to assess classification model performance can be misleading, particularly in scenarios with imbalanced datasets or when the costs of different types of errors vary significantly. In the future, standardized reporting of a wide range of error metrics could facilitate more detailed comparisons and better decision making in the application of these technologies.

Furthermore, the field of remote sensing lacks standardized protocols in data collection, labelling and class definitions and large, consistent reference datasets. Use of reference datasets helped improved accuracy in other fields (e.g. speech and face recognition) and could facilitate direct comparisons between efforts, in place of highly segmented, incompatible, and localized comparisons. Targeted funding opportunities and coordination of international partnerships could help facilitate the development of a baseline reference dataset at Landsat spatial resolution.

Looking into the future, Analysis Ready Data Cubes (ARDCs) are playing an increasingly vital role in advancing deep learning for large-scale earth observation by providing structured, standardized, and harmonized multi-dimensional data. These data cubes, which integrate spatial, temporal, and spectral information, address the persistent challenges of data preprocessing, format inconsistencies, and varying resolutions that have traditionally hindered efficient analysis (Ferreira et al. 2020). By organizing Earth observation data into a more accessible format, ARDCs enable deep learning models to better leverage the inherent spatiotemporal relationships within the data, leading to more accurate and robust results. Recent research illustrates the effectiveness of ARDCs in advancing deep learning applications within remote sensing, such as large-scale crop mapping using diverse remote sensing sources (J. Sun et al. 2024), precise sea ice classification from a data cube using Landsat-8 imagery (Cáceres, Schwarz, and Aldenhoff 2022), and environmental monitoring for climate change mitigation (Temenos et al. 2024). These examples highlight how ARDCs are facilitating a shift towards deep learning models that can generalize across time, handle large-scale operational systems, and fully exploit the complexity of earth observation data for environmental understanding and decision-making.

5. Conclusions

This study summarized and compared deep learning methods using Landsat observations. A database was constructed for both qualitative and quantitative meta-analysis based on 352 case studies extracted from 279 peer-reviewed articles. The qualitative analysis showed that:

- CNN is the dominant DNN method; however, other DNN methods have been increasing as the field evolves and responds to the latest DNN methods from other domains (e.g. Transformers).
- Most studies were conducted in single countries, with the top 3 being China, the U.S. A. and India. Despite the long record of Landsat observations, most studies concentrated on the last decade.
- Vegetation indices, such as Normalized Difference Vegetation Index, are the most common spectral indices added as model inputs.
- Sentinel-2 was the most frequent sensor fused with Landsat.

The quantitative meta-analysis of the published works indicated that:

- An approximate 5% increase in accuracy using a DNN is expected compared to a non-DNN accuracy of 80%, with a potentially higher increase for more challenging tasks.
- UNets achieve higher OA than CNNs (approximately 10% increase in OA using UNet compared to an OA of 80% using a CNN).
- Transformers achieve a higher OA than CNNs (approximately 5% increase in OA using Transformer compared to an 80% OA using CNN), although the increases in OA are smaller compared to the increases in OA when using UNet.

There is considerable potential for moving classification to the cloud as a service via on-demand processing. This is supported by increased availability of reference data, further advancements in artificial intelligence, and cloud-based computing power becoming widely and inexpensively available. Establishing benchmark classifiers along with a framework for researchers to contribute their own reference data while harvesting the power of deep learning via transfer learning could lead to reduced use of fragmented, site-specific approaches in the field of remote sensing.

Acknowledgments

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US. Government. The authors would like to thank Pete Doucette (Director, Earth Resources Observation and Science Center (EROS), U.S. Geological Survey) for his help in reviewing the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Giorgos Mountrakis http://orcid.org/0000-0001-5958-8134

Data availability statement

Data will be made available on request.

References

Abdali, E., M. J. Valadan Zoej, A. Taheri Dehkordi, and E. Ghaderpour. 2023. “A Parallel-Cascaded Ensemble of Machine Learning Models for Crop Type Classification in Google Earth Engine Using Multi-Temporal Sentinel-1/2 and Landsat-8/9 Remote Sensing Data.” *Remote Sensing* 16 (1): 127. https://doi.org/10.3390/rs16010127.

Akter, T., M. Y. Gazi, and M. B. Mia. 2021. “Assessment of Land Cover Dynamics, Land Surface Temperature, and Heat Island Growth in Northwestern Bangladesh Using Satellite Imagery.” *Environmental Processes* 8 (2): 661–690. https://doi.org/10.1007/s40710-020-00491-y.

Alem, A., and S. Kumar. 2020. “Deep Learning Methods for Land Cover and Land Use Classification in Remote Sensing: A Review.” In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, 903–908. Noida, India: IEEE.

Badrinarayanan, V., A. Kendall, and R. Cipolla. 2017. “Segnet: A deep convolutional encoder-decoder architecture for image segmentation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615.

Ball, J. E., D. T. Anderson, and C. S. Chan. 2017. “Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools, and Challenges for the Community.” *Journal of Applied Remote Sensing* 11 (4): 1–42609. https://doi.org/10.1117/1.JRS.11.042609.

Birhanu, A., I. Masih, P. van der Zaag, J. Nyssen, and X. Cai. 2019. “Impacts of Land Use and Land Cover Changes on Hydrology of the Gumara Catchment, Ethiopia.” *Physics and Chemistry of the Earth* 112:165–174. https://doi.org/10.1016/j.pce.2019.01.006.

Boonpook, W., Y. Tan, A. Nardkulpat, K. Torsri, P. Torteeka, P. Kamsing, U. Sawangwit, J. Pena, and M. Jainaen. 2023. “Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery.” *ISPRS International Journal of geo-Information* 12 (1): 14. https://doi.org/10.3390/ijgi12010014.

Boonprong, S., C. Cao, W. Chen, X. Ni, M. Xu, and B. K. Acharya. 2018. “The Classification of Noise-Afflicted Remotely Sensed Data Using Three Machine-Learning Techniques: Effect of Different Levels and Types of Noise on Accuracy.” *ISPRS International Journal of geo-Information* 7 (7): 274. https://doi.org/10.3390/ijgi7070274.

Cáceres, A., E. Schwarz, and W. Aldenhoff. 2022. “Landsat-8 Sea Ice Classification Using Deep Neural Networks.” *Remote Sensing* 14 (9): 1975. https://doi.org/10.3390/rs14091975.

Carrasco, L., A. W. O’Neil, R. D. Morton, and C. S. Rowland. 2019. “Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine.” *Remote Sensing* 11 (3): 288. https://doi.org/10.3390/rs11030288.

Chen, L. C., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. 2017. “Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous convolution, and Fully Connected Crfs.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4): 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

Chen, L. C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. 2018. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” In *Proceedings of the European conference on computer vision (ECCV)*, edited by V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, 801–818. Munich, Germany. https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, L., H. Letu, M. Fan, H. Shang, J. Tao, L. Wu, T. Zhang, et al. 2022. “An Introduction to the Chinese High-Resolution Earth Observation System: Gaofen-1~ 7 Civilian Satellites.” *Journal of Remote Sensing* 2022. https://doi.org/10.34133/2022/9769536.

Cheng, G., J. Han, and X. Lu. 2017. “Remote Sensing Image Scene Classification: Benchmark and State of the Art.” *Proceedings of the IEEE* 105 (10): 1865–1883. https://doi.org/10.1109/JPROC.2017.2675998.

Claverie, M., J. Ju, C. Justice, J. L. Dungan, E. F. Vermote, J.-C. Roger, and S. V. Skakun. 2018. "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set." *Remote Sensing of Environment* 219:145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Collobert, R., and J. Weston. 2008. "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning." In *Proceedings of the 25th International Conference on Machine Learning*, 160–167. Helsinki, Finland. https://doi.org/10.1145/1390156.1390177.

Cui, Z., Z. Cao, J. Yang, and H. Ren. 2015. "Hierarchical Recognition System for Target Recognition from Sparse Representations." *Mathematical Problems in Engineering* 2015:1–6. https://doi.org/10.1155/2015/527095.

Dang, X., H. Peng, X. Wang, and H. Zhang. 2008. "Theil-Sen Estimators in a Multiple Linear Regression Model." *Olemiss Education*, 2.

Darwin, B., P. Dharmaraj, S. Prince, D. E. Popescu, and D. J. Hemanth. 2021. "Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review." *Agronomy* 11 (4): 646. https://doi.org/10.3390/agronomy11040646.

Deng, L., and D. Yu. 2014. "Deep Learning: Methods and Applications." *Foundations and trends® in Signal Processing* 7 (3–4): 197–387. https://doi.org/10.1561/2000000039.

Digra, M., R. Dhir, and N. Sharma. 2022. "Land Use Land Cover Classification of Remote Sensing Images Based on the Deep Learning Approaches: A Statistical Analysis and Review." *Arabian Journal of Geosciences* 15 (10): 1003. https://doi.org/10.1007/s12517-022-10246-8.

Dosovitskiy, A. 2020. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929.

Eiumnoh, A., and R. P. Shrestha. 2000. "Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand." *PE & rs- Photogrammetric Engineering & Remote Sensing* 66 (3): 297–304.

Elsevier. 2024. "Scopus Database." Accessed July 1, 2024, from https://www.scopus.com.

ESA (European Space Agency). 2024a. *Sentinel-1*. Copernicus Open Access Hub. https://scihub.copernicus.eu/dhus/1.

ESA (European Space Agency). 2024b. *Sentinel-2*. Copernicus Open Access Hub. https://scihub.copernicus.eu/dhus/2.

Ferreira, K. R., G. R. Queiroz, L. Vinhas, R. F. Marujo, R. E. Simoes, M. C. Picoli, L. M. Fonseca, et al. 2020. "Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products." *Remote Sensing* 12 (24): 4033. https://doi.org/10.3390/rs12244033.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2020. "Generative Adversarial Networks." *Communications of the ACM* 63 (11): 139–144. https://doi.org/10.1145/3422622.

Gu, Y., Y. Wang, and Y. Li. 2019. "A Survey on Deep Learning-Driven Remote Sensing Image Scene Understanding: Scene classification, Scene Retrieval and Scene-Guided Object Detection." *Applied Sciences* 9 (10): 2110. https://doi.org/10.3390/app9102110.

Guo, M. H., T. X. Xu, J. J. Liu, Z. N. Liu, P. T. Jiang, T. J. Mu, S. M. Hu, R. R. Martin, M.-M. Cheng, and S.-M. Hu. 2022. "Attention Mechanisms in Computer Vision: A Survey." *Computational Visual Media* 8 (3): 331–368. https://doi.org/10.1007/s41095-022-0271-y.

Hansen, M. C., and T. R. Loveland. 2012. "A Review of Large Area Monitoring of Land Cover Change Using Landsat Data." *Remote Sensing of Environment* 122:66–74. https://doi.org/10.1016/j.rse.2011.08.024.

Hu, A. G. 2020. "Public Funding and the Ascent of Chinese Science: Evidence from the National Natural Science Foundation of China." *Research Policy* 49 (5): 103983. https://doi.org/10.1016/j.respol.2020.103983.

Hua, W., Z. Dai, H. Liu, and Q. Le. 2022. "Transformer Quality in Linear Time." In *International Conference on Machine Learning*, edited by K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, and S. Sabato, 9099–9117. Baltimore, MD: PMLR.

Hughes, M. J., and R. Kennedy. 2019. "High-Quality Cloud Masking of Landsat 8 Imagery Using Convolutional Neural Networks." *Remote Sensing* 11 (21): 2591. https://doi.org/10.3390/rs11212591.

Iwata, T., H. Ishida, Y. Osawa, and K. Tomioka. 2007. "Advanced Land Observing Satellite (ALOS): Development and on-Orbit Status." *The Journal of Space Technology and Science* 23 (1): _1_1–_1_13.

Jeong, S., J. Ko, T. Shin, and J. M. Yeom. 2022. "Incorporation of Machine Learning and Deep Neural Network Approaches into a Remote Sensing-Integrated Crop Model for the Simulation of Rice Growth." *Scientific Reports* 12 (1): 9030. https://doi.org/10.1038/s41598-022-13232-y.

Justice, C. O., J. R. G. Townshend, E. F. Vermote, E. Masuoka, R. E. Wolfe, N. Saleous, J. T. Morisette, and J. T. Morisette. 2002. "An Overview of MODIS Land Data Processing and Product Status." *Remote Sensing of Environment* 83 (1–2): 3–15. https://doi.org/10.1016/S0034-4257(02)00084-6.

Kankaku, Y., S. Suzuki, and Y. Osawa. 2013. "Alos-2 Mission and Development Status." In *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, 2396–2399. Melbourne, Australia: IEEE.

Keles, F. D., P. M. Wijewardena, and C. Hegde. 2023. "On the Computational Complexity of Self-Attention." In *International Conference on Algorithmic Learning Theory*, 597–619. Singapore: PMLR.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems*: 25.

Kussul, N., M. Lavreniuk, S. Skakun, and A. Shelestov. 2017. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14 (5): 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Lee, J., D. Han, M. Shin, J. Im, J. Lee, and L. J. Quackenbush. 2020. "Different Spectral Domain Transformation for Land Cover Classification Using Convolutional Neural Networks with multi-Temporal Satellite Imagery." *Remote Sensing* 12 (7): 1097. https://doi.org/10.3390/rs12071097.

Li, J., and Q. Wang. 2024. "CSDFormer: A Cloud and Shadow Detection Method for Landsat Images Based on Transformer." *International Journal of Applied Earth Observation and Geoinformation* 129:103799. https://doi.org/10.1016/j.jag.2024.103799.

Li, L., L. Han, M. Ding, Z. Liu, and H. Cao. 2020. "Remote Sensing Image Registration Based on Deep Learning Regression Model." *IEEE Geoscience and Remote Sensing Letters* 19:1–5. https://doi.org/10.1109/LGRS.2020.3032439.

Li, S., W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson. 2019. "Deep Learning for Hyperspectral Image Classification: An Overview." *IEEE Transactions on Geoscience & Remote Sensing* 57 (9): 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932.

Liu, J., H. Guo, Y. He, and H. Li. 2023. "Vision transformer-Based Ensemble Learning for Hyperspectral Image Classification." *Remote Sensing* 15 (21): 5208. https://doi.org/10.3390/rs15215208.

Liu, L., Y. Guo, Y. Li, Q. Zhang, Z. Li, E. Chen, L. Yang, and X. Mu. 2022. "Comparison of Machine Learning Methods Applied on Multi-Source Medium-Resolution Satellite Images for Chinese Pine (Pinus tabulaeformis) Extraction on Google Earth Engine." *Forests* 13 (5): 677. https://doi.org/10.3390/f13050677.

Liu, X., Z. Deng, and Y. Yang. 2019. "Recent Progress in Semantic Image Segmentation." *Artificial Intelligence Review* 52 (2): 1089–1106. https://doi.org/10.1007/s10462-018-9641-3.

Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, and B. Guo. 2021. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022. Singapore.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. Boston, MA. https://doi.org/10.1109/CVPR.2015.7298965.

Louet, J., and S. Bruzzi. 1999. "ENVISAT Mission and System." *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)*, 1680–1682. Vol. 3. Hamburg, Germany: IEEE.

Makantasis, K., K. Karantzalos, A. Doulamis, and N. Doulamis. 2015. "Deep Supervised Learning for Hyperspectral Data Classification Through Convolutional Neural Networks." *2015 IEEE international geoscience and remote sensing symposium (IGARSS)*, 4959–4962. Milan, Italy: IEEE. https://doi.org/10.1109/IGARSS.2015.7326945.

Maurício, J., I. Domingues, and J. Bernardino. 2023. "Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review." *Applied Sciences* 13 (9): 5521. https://doi.org/10.3390/app13095521.

Mohajerani, S., T. A. Krammer, and P. Saeedi. 2018. "Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks." arXiv preprint arXiv:1810.05782. https://doi.org/10.48550/arXiv.1810.05782.

Mohajerani, S., and P. Saeedi. 2019. "Cloud-Net: An End-To-End Cloud Detection Algorithm for Landsat 8 Imagery." In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 1029–1032. Yokohama, Japan: IEEE. https://doi.org/10.1109/IGARSS.2019.8898776.

Mohajerani, S., and P. Saeedi. 2021. "Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery via Filtered Jaccard Loss Function and Parametric Augmentation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:4254–4266. https://doi.org/10.1109/JSTARS.2021.3070786.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and P. R. I. S. M. A. Group*. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *Annals of Internal Medicine* 151 (4): 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Mountrakis, G., and S. S. Heydari. 2023. "Harvesting the Landsat Archive for Land Cover Land Use Classification Using Deep Neural Networks: Comparison with Traditional Classifiers and Multi-Sensor Benefits." *ISPRS Journal of Photogrammetry and Remote Sensing* 200:106–119. https://doi.org/10.1016/j.isprsjprs.2023.05.005.

Mountrakis, G., and S. Shah Heydari. 2021. "Continental US Land Cover Mapping Using Deep Neural Networks, Landsat Time-Series Observations and Large Reference Datasets." *AGU Fall Meeting Abstracts* 2021:B31C–01.

NASA LP DAAC. 2024. "Moderate Resolution Imaging Spectroradiometer (MODIS)." https://modis.gsfc.nasa.gov/data/ .

Neupane, B., T. Horanont, and J. Aryal. 2021. "Deep learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and meta-Analysis." *Remote Sensing* 13 (4): 808. https://doi.org/10.3390/rs13040808.

Niu, Z., G. Zhong, and H. Yu. 2021. "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing* 452:48–62. https://doi.org/10.1016/j.neucom.2021.03.091.

Ouhami, M., A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals. 2021. "Computer vision, iot and Data Fusion for Crop Disease Detection Using Machine Learning: A Survey and Ongoing Research." *Remote Sensing* 13 (13): 2486. https://doi.org/10.3390/rs13132486.

Panboonyuen, T., K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul. 2021. "transformer-Based Decoder Designs for Semantic Segmentation on Remotely Sensed Images." *Remote Sensing* 13 (24): 5100. https://doi.org/10.3390/rs13245100.

Raghu, M., T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy. 2021. "Do Vision Transformers Seem Like Convolutional Neural Networks?" *Advances in Neural Information Processing Systems* 34:12116–12128.

Raj, T., F. Hanim Hashim, A. Baseri Huddin, M. F. Ibrahim, and A. Hussain. 2020. "A Survey on lidar Scanning Mechanisms." *Electronics* 9 (5): 741. https://doi.org/10.3390/electronics9050741.

Roy, S. K., A. Jamali, J. Chanussot, P. Ghamisi, E. Ghaderpour, and H. Shahabi. 2025. "Simpoolformer: A Two-Stream Vision Transformer for Hyperspectral Image Classification." *Remote Sensing Applications: Society & Environment* 37:101478. https://doi.org/10.1016/j.rsase.2025.101478.

Saha, J., S. S. Ria, J. Sultana, U. A. Shima, M. M. H. Seyam, and M. M. Rahman. 2024. "Assessing Seasonal Dynamics of Land Surface Temperature (LST) and Land Use Land Cover (LULC) in Bhairab, Kishoreganj, Bangladesh: A Geospatial Analysis from 2008 to 2023." *Case Studies in Chemical and Environmental Engineering* 9:100560. https://doi.org/10.1016/j.cscee.2023.100560.

Shafique, A., G. Cao, Z. Khan, M. Asad, and M. Aslam. 2022. "Deep Learning-Based Change Detection in Remote Sensing Images: A Review." *Remote Sensing* 14 (4): 871. https://doi.org/10.3390/rs14040871.

Shi, W., M. Zhang, R. Zhang, S. Chen, and Z. Zhan. 2020. "Change Detection Based on Artificial Intelligence: State-Of-The-Art and Challenges." *Remote Sensing* 12 (10): 1688. https://doi.org/10.3390/rs12101688.

Shirmard, H., E. Farahbakhsh, R. D. Müller, and R. Chandra. 2022. “A Review of Machine Learning in Processing Remote Sensing Data for Mineral Exploration.” *Remote Sensing of Environment* 268:112750. https://doi.org/10.1016/j.rse.2021.112750.

Singh, R., M. Biswas, and M. Pal. 2022. “Cloud Detection Using Sentinel 2 Imageries: A Comparison of XGBoost, RF, SVM, and CNN Algorithms.” *Geocarto International* 38 (1): 1–32. https://doi.org/10.1080/10106049.2022.2146211.

Singh, V. K., P. Singh, A. Kanaujia, and A. Nandy. 2024. “Who is Funding Indian Research? A Look at Major Funding Sources Acknowledged in Indian Research Papers.” arXiv preprint arXiv:2411.15045.

Sun, A. Y., and B. R. Scanlon. 2019. “How Can Big Data and Machine Learning Benefit Environment and Water Management: A Survey of Methods, Applications, and Future Directions.” *Environmental Research Letters* 14 (7): 073001. https://doi.org/10.1088/1748-9326/ab1b7d.

Sun, J., X. Yao, S. Yan, Q. Xiong, G. Li, and J. Huang. 2024. “Large-Scale Crop Mapping Based on Multi-Source Remote Sensing Intelligent Interpretation: A Spatiotemporal Data Cubes Approach.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, New Jersey, USA.

Sutskever, I., O. Vinyals, and Q. V. Le. 2014. “Sequence to Sequence Learning with Neural Networks.” *Advances in Neural Information Processing Systems* 27.

Tang, S., J. Zhang, S. Zhu, and P. Tan. 2022. *Quadtree Attention for Vision Transformers*. arXiv preprint arXiv:2201.02767. https://doi.org/10.48550/arXiv.2201.02767.

Temenos, A., N. Temenos, I. N. Tzortzis, I. Rallis, A. Doulamis, and N. Doulamis. 2024. “C2a-Dc: A Context-Aware Adaptive Data Cube Framework for Environmental Monitoring and Climate Change Crisis Management.” *Remote Sensing Applications: Society & Environment* 34:101171. https://doi.org/10.1016/j.rsase.2024.101171.

Tesfaw, B. A., B. Dzwairo, and D. Sahlu. 2023. “Assessments of the Impacts of Land use/Land Cover Change on Water Resources: Tana sub-basin, Ethiopia.” *Journal of Water and Climate Change* 14 (2): 421–441. https://doi.org/10.2166/wcc.2023.303.

Tsagkatakis, G., A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides. 2019. “Survey of deep-Learning Approaches for Remote Sensing Observation Enhancement.” *Sensors (Switzerland)* 19 (18): 3929. https://doi.org/10.3390/s19183929.

Tyc, G., J. Tulip, D. Schulten, M. Krischke, and M. Oxfort. 2005. “The rapideye Mission Design.” *Acta Astronautica* 56 (1–2): 213–219. https://doi.org/10.1016/j.actaastro.2004.09.029.

U.S. Geological Survey. 2024. “Landsat-7 Image.” https://earthexplorer.usgs.gov/3.

U.S. Geological Survey. 2024. “The Landsat Archive.” https://earthexplorer.usgs.gov/4.

Vali, A., S. Comai, and M. Matteucci. 2020. “Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review.” *Remote Sensing* 12 (15): 2495. https://doi.org/10.3390/rs12152495.

Vaswani, A. 2017. “Attention is All You Need.” *Advances in Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.1706.03762.

Vidal, C. 2008. “What is a worldview?” In *Nieuwheid denken: De wetenschappen en het creatieve aspect van de werkelijkheid*, edited by H. Van Belle and J. Van der Veken, 71–86. Leuven, Belgium: Acco.

Wandinger, U. 2005. “Introduction to Lidar.” In *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*, edited by C. Weitkamp, 1–18. New York, NY: Springer New York.

Wang, P., B. Bayram, and E. Sertel. 2022. “A Comprehensive Review on Deep Learning Based Remote Sensing Image Super-Resolution Methods.” *Earth-Science Reviews* 232:104110. https://doi.org/10.1016/j.earscirev.2022.104110.

Wang, X., D. Liu, K. Ding, and X. Wang. 2012. “Science Funding and Research Output: A Study on 10 Countries.” *Scientometrics* 91 (2): 591–599. https://doi.org/10.1007/s11192-011-0576-6.

Woodcock, C. E., S. A. Macomber, M. Pax-Lenney, and W. B. Cohen. 2001. “Monitoring Large Areas for Forest Change Using Landsat: Generalization Across Space, Time and Landsat Sensors.” *Remote Sensing of Environment* 78 (1–2): 194–203. https://doi.org/10.1016/S0034-4257(01)00259-0.

Yuan, X., J. Shi, and L. Gu. 2021. “A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery.” *Expert Systems with Applications* 169:114417. https://doi.org/10.1016/j.eswa.2020.114417.

Zhang, H. K., D. Luo, and Z. Li. 2024. “Classifying Raw Irregular Time Series (CRIT) for Large Area Land Cover Mapping by Adapting Transformer Model.” *Science of Remote Sensing* 9:100123. https://doi.org/10.1016/j.srs.2024.100123.

Zhang, X., T. Zhang, G. Wang, P. Zhu, X. Tang, X. Jia, and L. Jiao. 2023. “Remote Sensing Object Detection Meets Deep Learning: A Metareview of Challenges and Advances.” *IEEE Geoscience and Remote Sensing Magazine* 11 (4): 8–44. https://doi.org/10.1109/MGRS.2023.3312347.

Zhang, Z., J. Geiger, J. Pohjalainen, A. E. D. Mousa, W. Jin, and B. Schuller. 2018. “Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments.” *ACM Transactions on Intelligent Systems and Technology (TIST)* 9 (5): 1–28. https://doi.org/10.1145/3178115.

Zhu, X. X., D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer. 2017. “Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources.” *IEEE Geoscience and Remote Sensing Magazine* 5 (4): 8–36. https://doi.org/10.1109/MGRS.2017.2762307.

Zubkova, M., M. Lötter, F. Bronkhorst, and L. Giglio. 2024. “Assessment of the Effectiveness of Coarse Resolution Fire Products in Monitoring Long-Term Changes in Fire Regime within Protected Areas in South Africa.” *International Journal of Applied Earth Observation and Geoinformation* 132:104064. https://doi.org/10.1016/j.jag.2024.104064.