Ecological Informatics 98 (2026) 103912

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Toward optimized deep learning-based land use/land cover classification using Landsat 8: a systematic multi-condition experiment

Juyeong Nam [a], Younju Lee [b], Changyeon Lee [b,*]

[a] Data Service Division, BigValue Co., Ltd., Seoul, Republic of Korea
[b] Department of Urban Design and Studies, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea

ARTICLE INFO

Keywords:
Land use/land cover classification
Deep learning
Landsat 8
Land cloud cover
Multi-condition experiment

ABSTRACT

High-fidelity land use/land cover (LULC) maps are fundamental for monitoring ecosystem dynamics and supporting sustainable environmental planning. While Landsat imagery supports long-term ecological analysis, its medium resolution presents significant informatics challenges, such as mixed pixels and severe class imbalances among natural habitats. This study proposes a systematic informatics framework for generating large-scale training datasets by integrating public land-cover products with Landsat 8 multispectral imagery. We evaluated 252 experimental combinations of cloud-cover thresholds, spectral bands, and loss functions across U-Net, DeepLabv3+, and Swin-UPerNet architectures. U-Net achieved the highest overall performance, particularly excelling in dominant classes like forest and farmland. Conversely, Swin-UPerNet and Focal Tversky Loss (FTL) provided a critical relative advantage in identifying "hard-to-classify" natural classes—grassland and wetland—which are inherently difficult to distinguish from dominant classes due to spectral overlap and visual ambiguity. The FTL further enhanced minority class F1-scores by up to 10%. While the RGB + NIR + SWIR1/2 combination was optimal, configurations utilizing cloud thresholds ≥20% proved that increased data volume (> 10,000 samples) effectively offsets moderate noise. Ultimately, these findings suggest that a hybrid architecture—integrating Swin-Transformer's global context modeling with U-Net's progressive upsampling—represents the most promising informatics pathway for next-generation, medium-resolution ecological monitoring.

1. Introduction

Accurate and continuous monitoring of land use/land cover (LULC) change patterns is essential for sustainable development, as it provides foundational data for assessing habitat fragmentation and ecosystem health (Dhinwa et al., 1992; Zhu et al., 2022). These observations help quantify how land use change and urbanization affect biodiversity and ecosystem services, particularly in ecologically sensitive habitats such as forests and wetlands (Redlich et al., 2022). Recently, to understand these ecological interactions and to support timely decision-making in conservation planning and prioritization, researchers have extensively used public land cover data. However, the generation of public LULC data through ground surveys is labor-intensive and costly (Chen and Yao, 2023), resulting in long-term data production intervals. This long-term data frequency does not capture intermediate changes such as urban expansion, land-cover conversion, and forest loss. This hinders the exploration of long-term LULC dynamics and their ecological consequences (Talukdar et al., 2020). Furthermore, this limitation restricts the availability of big data required by AI-based models for simulating land use change (Shao et al., 2023; Xu et al., 2024) and urban growth (Gui et al., 2025; Nam and Lee, 2025), which are essential for proactive environmental and ecological conservation management. In this context, the Landsat archive presents a valuable resource for constructing temporally continuous LULC information as it provides consistent medium-resolution observations across multiple decades.

Previously, technical limitations presented challenges in addressing these issues; however, recent studies have employed LULC classification models (LCMs) and remote sensing (RS) imagery data to enable the efficient establishment of LULC maps. In particular, deep learning-based computer vision (CV) has demonstrated superiority over statistical techniques, including maximum likelihood methods, and machine learning technologies, such as support vector machines (Boonpook et al., 2023; Zang et al., 2021). These models automatically assign an LULC class to each pixel in the RS imagery by extracting information on

* Corresponding author at: Department of Urban Design and Studies, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea.
E-mail address: idealcity@cau.ac.kr (C. Lee).

https://doi.org/10.1016/j.ecoinf.2026.103912
Received 6 May 2026; Received in revised form 29 June 2026; Accepted 1 July 2026
Available online 2 July 2026
1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

artificial ground objects, such as buildings and roads, and natural elements, including greenery and water (Zhang et al., 2025).

These models can be categorized into convolutional neural networks (CNNs) and transformer-based architectures. CNN-based models hierarchically extract image features (Guo et al., 2024a; He and Guo, 2024), and identify LULC patterns using n × n kernels that define the receptive fields (Ramos and Sappa, 2025). Previous studies employed various semantic segmentation networks, including fully convolutional networks (FCN), U-Net, and DeepLabv3+. More recently, transformer-based models—such as Vision Transformer (Dosovitskiy et al., 2020) and Swin-Transformer (Liu et al., 2021)—have been proposed. These models split input images into patches, similar to text tokens, and effectively learn long-range dependencies and global contextual information using self-attention mechanisms. These architectures demonstrated greater accuracy than that of CNNs on high-resolution benchmark datasets, such as the EuroSAT, Gaofen Image Dataset (GID), and Vaihingen dataset (Khan et al., 2024; Zhang et al., 2025). However, recent studies have increasingly focused on developing models optimized for comprehensive geometric information in high-resolution benchmark datasets (Ma et al., 2019; Zhang et al., 2025). As these datasets were collected only after the late 2010s, analyzing historical LULC prior to this period remains impossible. Additionally, whether these models exhibit superior performance when applied to medium-resolution satellite imagery, such as Landsat imagery, remains unclear.

Landsat is valuable for long-term historical analyses that support land-use planning, ecological management and policymaking because it has been continuously acquired since 1972. However, LULC classification using Landsat has four major challenges: First, although supervised deep learning requires training datasets, labeled training data are extremely limited (Vali et al., 2020). Second, researchers must determine the Land Cloud Cover (LCC) threshold during data collection for LCMs; however, the criteria for an appropriate LCC threshold and data volume remain ambiguous. Most studies have collected and used low-noise, high-quality data with an LCC < 10% (Dou et al., 2024; Goldblatt et al., 2018; Phiri et al., 2018); however, this strict filtering can substantially reduce the quantity and diversity of the training data. Specifically, the quantitative tradeoff between Landsat scene quality and sample volume has rarely been explored in deep learning models. Third, the identification of ecologically sensitive natural classes, such as grasslands and wetlands, remains a significant informatics bottleneck. Despite their high biodiversity and significant conservation value (Lamsal et al., 2015), these classes often exhibit high spectral overlap and visual ambiguity with dominant classes like forest and water, making them 'hard-to-classify' even with manual visual interpretation (Li et al., 2024; Wilkinson, 2005; Wu and Murray, 2003). Deep learning-based LCMs offer an effective method of addressing this challenge by extracting discriminative features from multispectral inputs that are inaccessible to manual visual interpretation. In this regard, investigating

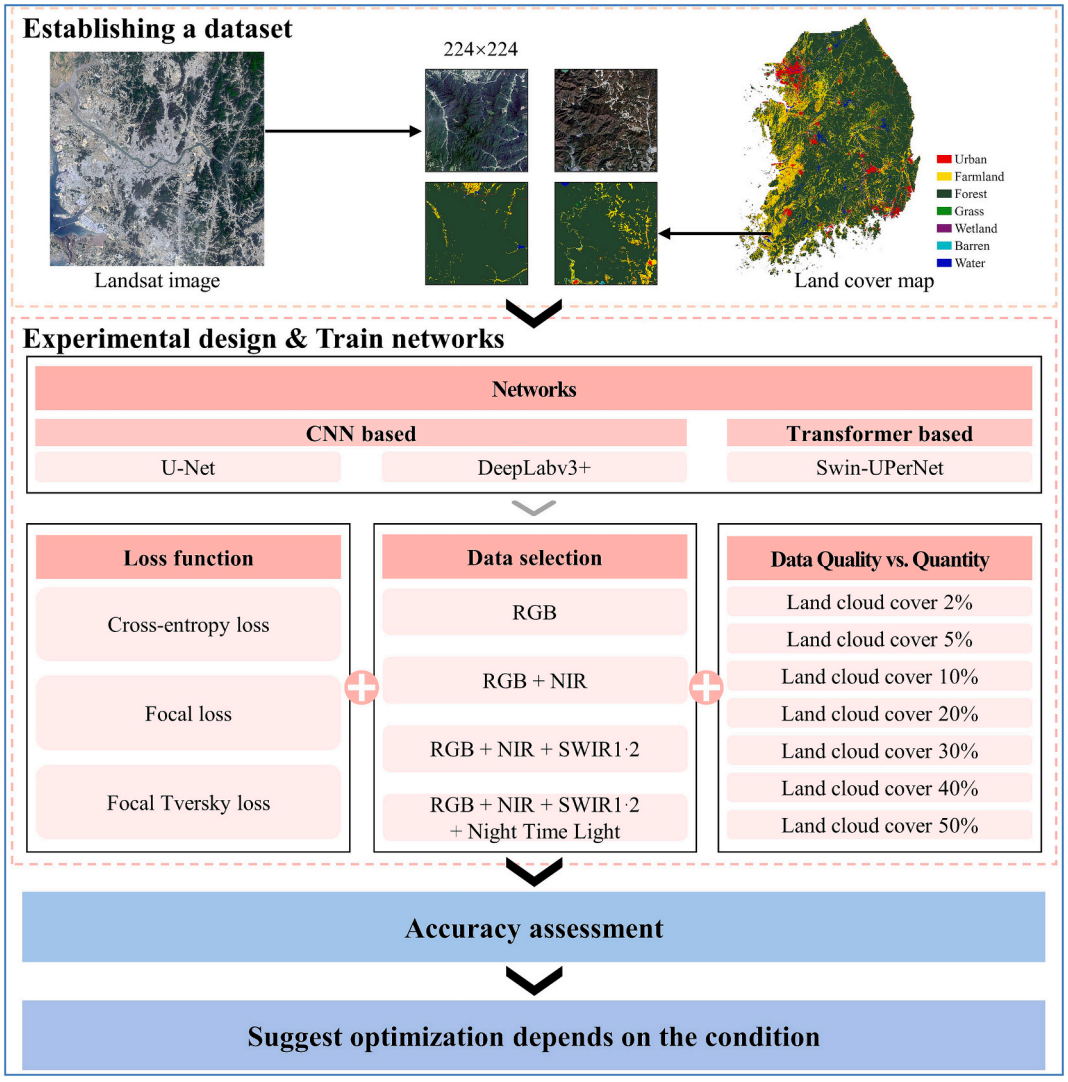


Fig. 1. Research framework.

various approaches to improve performance is necessary. Finally, the contribution of Landsat's multispectral bands to classification capability remains insufficiently understanding. Yu et al. (2019) identified the optimal combination of Landsat bands using Pearson's correlation; however, it remains unclear whether this is applicable to nonlinear deep learning models. Boonpook et al. (2023) compared single-band and multispectral band usage and found that overall performance was maximized when all bands were used. However, they did not investigate various multispectral band combinations. Therefore, further studies are required to understand the improvements in the performance of the overall model and each LULC class based on these combinations.

To address this gap (Fig. 1), this study aimed to: (1) establish a training dataset by integrating Landsat 8 imagery with publicly available land-cover products without manual labeling; (2) systematically evaluate the classification performance across 252 experimental combinations of LCC thresholds, spectral band combinations, loss functions, and model architectures; and (3) develop empirically grounded guidelines for efficient model design and data utilization in medium-resolution LULC classification. The proposed framework provides foundations for efficiently generating high-frequency LULC data capable of capturing the dynamics of continuous land cover and vegetation change.

2. Materials and methods

2.1. Multispectral satellite dataset

In this study, we established training datasets for South Korea by combining land cover maps, Landsat 8 Collection 2 Level-2 (L2SP) images, and nighttime light (NTL) (Table 1). Based on Landsat 8 imagery and ground surveys, the land-cover map was a 30 m resolution government-produced product for 2018. The LULC was classified into seven major categories: urban area, farmland, forest, grassland, wetland, barren land, and water. Of these, forest is the dominant class, accounting for 68.2% of the total area, followed by farmland (19.1%) and urban area (5.7%). Grassland and wetland represent comparatively minor proportions at 2.8% and 0.7%, respectively, with all remaining classes each occupying less than 2% (Fig. 2a). South Korea is characterized by extensive forested mountain ranges extending from the eastern coast to the central interior, with farmland concentrated in the western plains and urban areas situated in flat terrain largely surrounded by farmland and forested mountains. Wetland is primarily distributed along rivers and the western and southern coastlines, and grassland is typically found adjacent to forests and farmland. Despite occupying a minor share of the landscape, natural habitats such as grassland and wetland, are essential for establishing robust informatics monitoring systems in ecologically sensitive regions. The product was utilized as ground truth (GT) label data. Landsat 8, which has been operational since 2013, is equipped with the Operational Land Imager (OLI) sensor with nine spectral bands: eight multispectral bands at 30 m spatial resolution and one panchromatic band at 15 m. As illustrated in Fig. 2b, we identified the Landsat 8 frames that cover the entirety of South Korea. From these frames, we selected six multispectral bands at 30 m resolution from 2018: visible (Blue, Green, Red; RGB), near-infrared (NIR), and shortwave infrared (SWIR1 and SWIR2), which are spectrally consistent across earlier Landsat sensors, including Landsat 7 ETM+ and Landsat 5 TM. This cross-sensor spectral compatibility ensures that the proposed framework can be extended to archival Landsat imagery to support long-term, temporally continuous LULC mapping. This spectral consistency is particularly valuable for ecological informatics applications that require long-term trend analysis of environmental change across multiple decades. According to NASA (2013), the role of each multispectral band depends on its wavelength. The RGB bands represent the visible portion of the spectrum, indicating surface color and visual appearance. The NIR band captures vegetation reflectance and is sensitive to variations in vegetation vigor. The SWIR bands are valuable for identifying impervious surfaces in urban areas because they are sensitive to soil and surface moisture. The NTL data are used to measure nighttime lighting and delineate urban areas (Bagan and Yamagata, 2015; He et al., 2025b; Keola et al., 2015). This study used NTL data acquired with an original spatial resolution of approximately 464 m (Earth Observation Group, 2024) and resampled it to 30 m using the Block Statistics tool in ArcGIS Pro (version 3.4.3) to match the Landsat grid. Unlike conventional studies that employ random pixel-wise or patch-wise sampling—which often leads to inflated performance due to spatial autocorrelation—this study adopted a strict spatial-split strategy to ensure the spatial independence of the training and test datasets. As illustrated in Fig. 2c, specific geographic grid blocks, including the major metropolitan area of Busan, were exclusively reserved for testing, while the remaining regions were used for training. This approach provides a more conservative and realistic assessment of the model's performance. Ensuring spatial independence through this strategy is a critical methodological requirement in environmental data science to mitigate the effects of spatial autocorrelation, which can otherwise lead to over-optimistic performance estimates. By evaluating the model on a spatially distinct region characterized by different urban layouts and environmental conditions, we demonstrated the model's robustness and generalizability across unseen geographic locations, rather than simple memorization of local spectral patterns. The image patch size for the training was 224 × 224 pixels. During model training, we applied data augmentation techniques, including vertical and horizontal flipping, 90° rotation, additive noise, and blurring, to improve the generalization

Table 1
Description of multispectral imagery data.

Data				Description	Resolution (m)	Source
Target (Y)	Land cover map			There are seven LULC classes (urban, farmland, forest, grass, wetland, barren, water)		The Environmental Geographic Information Service (EGIS)
Input (X)	Landsat 8 (Collection2-level 2)	RGB	Band 2 (Blue)	Used to delineate and map water bodies	30	U.S. Geological Survey
			Band 3 (Green)	Sensitive to vegetation greenness		
			Band 4 (Red)	Separates vegetation from soil and detects built-up or impervious urban structures.		
		Near Infrared (NIR)	Band 5	Exhibits strong reflectance in healthy vegetation		
		Short-Wave Infrared (SWIR)	Band 6 (SWIR1)	Sensitive to moisture content in soil and vegetation		
			Band 7 (SWIR2)	Detects moisture variations		
	Visible Infrared Imaging Radiometer Suite (VIIRS)	Nighttime light (NTL)		Measures anthropogenic nighttime illumination; serves as a proxy for urban extent	464	Earth Observation Group

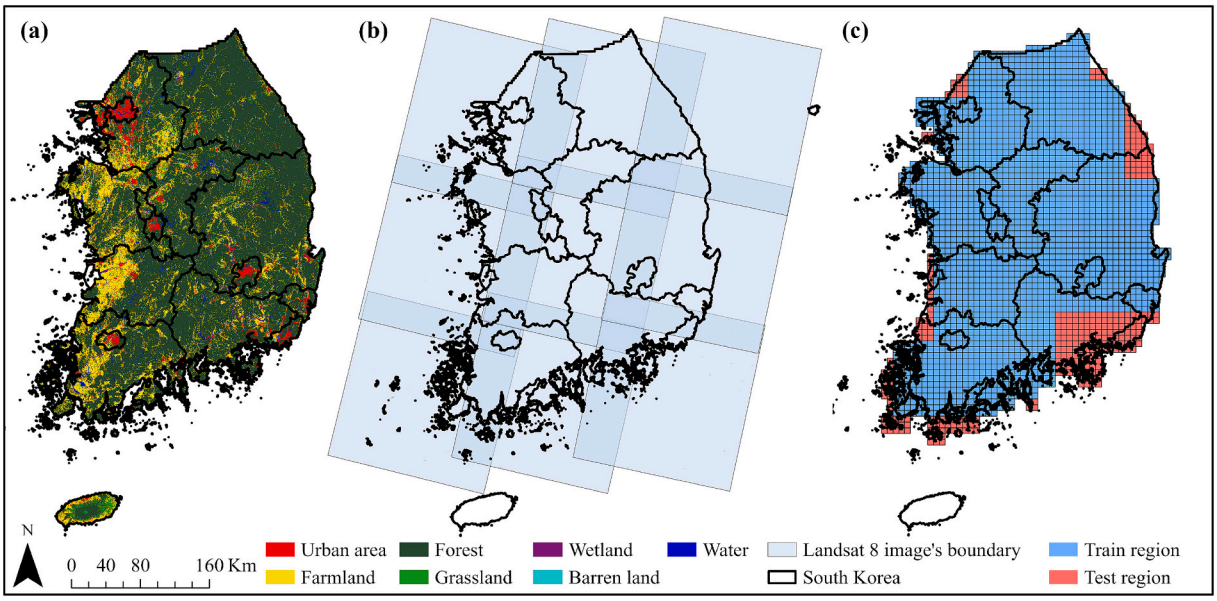


Fig. 2. Dataset in the study: (a) Land use/land cover map in the study area, (b) Boundary of the Landsat8 image, which is used in this study, and (c) training and test regions.

performance of the model.

2.2. Deep neural networks for land use classification

In this study, we employed two CNN-based architectures, U-Net and DeepLabv3+, and a transformer-based architecture, Swin-UPerNet (Fig. 3). All the models were implemented using MMSegmentation v1.2.2 (MMSegmentation Contributors, 2024), an open-source semantic segmentation toolbox built on PyTorch. All the experiments were conducted in a unified environment using PyTorch 2.0.1, Python 3.9.21, and MMCV 2.1.0.

2.2.1. U-Net

In this study, we used the U-Net backbone implementation combined with FCNHead, which applies a final 1 × 1 convolution to the last feature map to reduce the channel dimensions to the number of LULC classes, thereby producing an LULC classification map. The U-Net, developed by Ronneberger et al. (2015), is an extensively used model for pixel-wise classification of LULC from satellite imagery. As shown in Fig. 3a, the U-Net architecture has a symmetric encoder-decoder structure for effectively segmenting complex land cover boundaries (Cao et al., 2024; Ni et al., 2023; Ramos and Sappa, 2025). The encoder employs convolutional layers to extract low-level features, including detailed boundary information, from the RS imagery. These features are progressively compressed through pooling layers to obtain high-level features representing global information. The decoder restores the image size using bilinear interpolation upsampling by a factor of two and performs pixel-wise classification. The feature maps from each encoder stage are integrated into the corresponding decoder stage using skip connections, enabling the model to learn both low-level fine-grained boundaries and high-level semantic information simultaneously. In this study, U-Net served as a robust baseline for capturing well-defined ecological boundaries, such as forest-urban transitions, which is foundational for mapping dominant land-cover types.

2.2.2. DeepLabv3+

Fig. 3b illustrates the architecture of DeepLabv3+. This model employs atrous convolutions and atrous spatial pyramid pooling (ASPP) to simultaneously capture both a large receptive field and a multiscale context (Chen et al., 2018). In contrast to standard CNNs, which have fixed receptive fields, atrous convolutions extract spatial features at various dilation rates. This capability enables the model to learn representations over a broader spatial range without increasing the number of parameters. Additionally, ASPP extracts multiscale feature maps using different dilation rates and global average pooling, and then compresses them using 1 × 1 convolutions. This module employs depthwise separable convolutions (Chollet, 2017) to reduce the computational cost.

In this study, a ResNet-50 backbone, integrated with atrous convolutions at multiple dilation rates, was employed as the encoder. In the decoder, the boundary information of the LULC classes was recovered using the low-level features extracted from the encoder. These features were concatenated with the upsampled ASPP output, and the result was subsequently upsampled four times to produce a segmentation map that matched the input resolution.

2.2.3. Swin-UPerNet

While existing CNN-based models focus on local information around pixels, transformer-based models can learn interactions in large spatial ranges through self-attention across image patches. Thus, they can capture global and contiguous land use patterns effectively. However, the Vision Transformer is computationally intensive because self-attention is applied to all patches of the input image (Dosovitskiy et al., 2020), which may be inefficient for Landsat-based studies with relatively limited data. Therefore, this study employed Swin-UPerNet, which combines a Swin-Transformer Tiny using window-based self-attention as the encoder and a UPerNet decoder to perform multiscale feature fusion. Swin-UPerNet is a model designed to simultaneously capture both local features and the global spatial context across the entire image (Liu et al., 2021).

As shown in Fig. 3c, Swin-UPerNet has a hierarchical structure in which the input image is divided into 4 × 4 patches during the patch partition stage. At each stage, the spatial resolution of feature maps was sequentially reduced to 1/4, 1/8, 1/16, and 1/32, while increasing the number of channels. Within each stage, window-based self-attention was applied within each local window, enabling efficient learning of the relationships between neighboring land-use patterns. In the feature fusion stage of the decoder, pyramid pooling was applied to integrate the global contextual information, generating a fused feature map at 1/4 the size of the original image. Finally, the segmentation output is produced

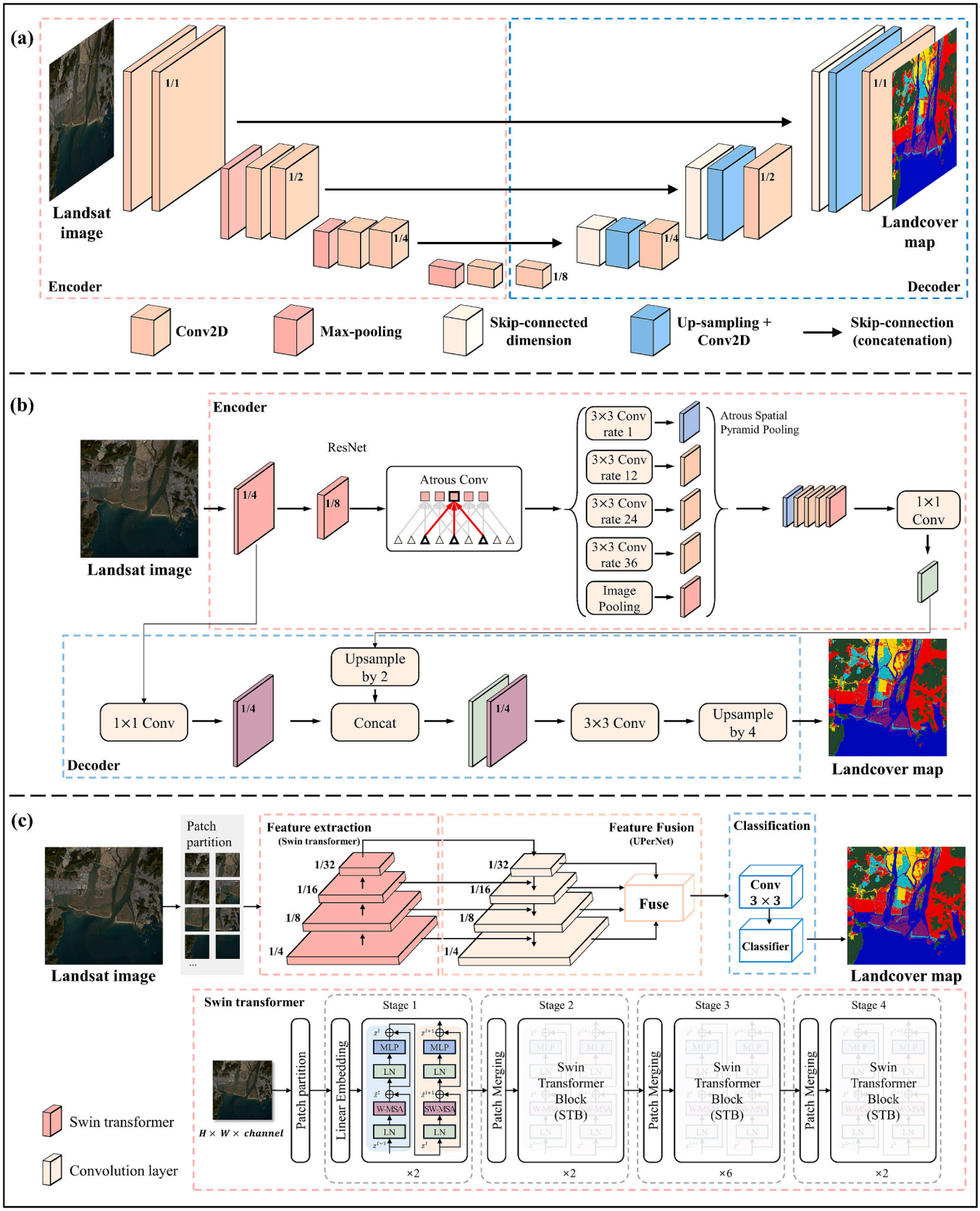


Fig. 3. Architecture of each model: (a) U-Net, (b) DeepLabv3+, and (c) Swin-UperNet.

by applying a 3 × 3 convolutional layer and a 1 × 1 classifier, followed by 4× upsampling. This global context modeling is particularly critical for distinguishing ecologically sensitive but spatially fragmented classes, such as wetlands and grasslands, which often lack distinct local spectral boundaries and are frequently confused with dominant vegetation types.

2.2.4. Parameters and hyperparameters setting

The training parameters and hyperparameter settings for U-Net, DeepLabv3+, and Swin-UPerNet used in this study were as follows: The maximum number of training epochs was set to 200, and the training was terminated if the validation loss did not improve over five epochs. These models employed the same type of fully convolutional segmentation head (FCNHead or UPerHead) for the classification stage (Table 2). The batch size for each model was determined empirically based on optimization stability and training efficiency; a batch size of 16 was used for U-Net, whereas DeepLabv3+ and Swin-UPerNet were trained with a batch size of 8. To mitigate overfitting, we applied a dropout ratio of 0.1 in the decoding heads of U-Net and DeepLabv3+, and 0.2 in the decoding and auxiliary heads of Swin-UPerNet. We used the AdamW optimizer with an initial learning rate of 1.0×10^{-4}, and adopted a warm-up cosine learning rate scheduler to ensure stable convergence of the loss (Goyal et al., 2018; He et al., 2018). Applying a consistent hyperparameter configuration across all models could ensure the reproducibility of the informatics workflow, allowing for a fair comparison of architectural performance under identical experimental conditions.

2.3. Experimental design

The experiments were conducted by training each CV model using various combinations of LCC thresholds, multispectral band data, and loss functions. A lower LCC ratio indicates a clearer surface, making land-use identification easier and reducing wavelength distortion caused by shadows, thus representing high-quality data. Conversely, a higher LCC ratio increases the probability of data errors, indicating a lower quality. In this study, the datasets were divided across ranges from high-quality data with LCC ratios of 0–2%, 0–5%, 0–10%, 0–20%, 0–30%, 0–40%, and 0–50%. This approach facilitated experimentation with appropriate levels of data quality and quantity during data acquisition for training the LCMs. Evaluating this quality–quantity trade-off is a core data informatics strategy. It identifies the optimal threshold at which increased sample volume can offset spectral noise—a critical consideration for maximizing the utility of archival satellite data.

To analyze the relationship between band usage and class-specific classification performance, we considered four band combinations: (1) RGB, (2) RGB + NIR, (3) RGB + NIR + SWIR1/2, and (4) RGB + NIR + SWIR1/2 + NTL. This design enabled us to quantify the effect of incorporating additional multispectral information on the separability of specific LULC classes.

Additionally, we used three loss functions: cross-entropy loss (CE), focal loss (FL), and focal tversky loss (FTL). The CE is the most widely used objective for training deep segmentation models. However, it can be biased toward majority land-cover classes (Lin et al., 2018). To address this issue, we employed FL, which is designed to highlight hard-to-classify pixels during training (Lin et al., 2018; Zhang et al., 2020). Nevertheless, CE and FL do not account for the uneven distribution of classes in the training data, which ultimately leads to a training bias. The Focal Tversky Loss (FTL), proposed by Abraham and Khan (2019), is an integrated loss function that combines FL and Tversky loss. The Tversky loss component is particularly beneficial for segmenting minority land cover classes that are easily overlooked during training. By jointly assigning weights to both class distribution imbalances and hard-to-classify pixels, the FTL effectively mitigates the class imbalance issue inherent in land-cover segmentation tasks. Mitigating this bias is especially vital for ecological informatics applications, where rare but ecologically sensitive habitats—such as wetlands and grasslands—are often overshadowed by dominant. In this study, the parameters of the loss function were retained at their optimized values as specified in the original study that proposed the loss function.

Table 2
Parameters and hyperparameters for each model.

Method	Backbone	Head	Batchsize	Dropout ratio	Learning rate
U-Net	U-Net	FCHead	16	0.1	0.0001
DeepLabv3+	ResNet	FCHead	8	0.1	0.0001
Swin-UPerNet	Swin-transfomer	UPerHead	8	0.2	0.0001

The overall performance of each model was assessed using Overall Accuracy (OA), Mean F1-score (MF1), and Frequency-Weighted F1-score (FWF1). Intersection-over-Union (IoU) has been commonly used in previous studies to evaluate pixel-level classification accuracy, including both boundary alignment and regional coverage. However, it is less appropriate for medium-resolution imagery, where mixed pixels are an inherent limitation. Accordingly, this study employed the F1-score as the primary performance indicator. For balanced ecological monitoring, MF1 is prioritized over OA. Unlike OA, MF1 provides an unbiased assessment of the model performance on minority natural classes. This is critical for identifying fragmented habitats that are difficult to distinguish. The OA measures the overall correctness of the classification by computing the proportion of correctly classified pixels among all pixels. The F1-score evaluates the unbiased classification performance under class imbalance conditions by calculating the harmonic mean of precision and recall. Precision refers to the proportion of pixels truly belonging to land-cover class *i* among all pixels classified as class *i*, while recall refers to the proportion of pixels correctly identified as class *i* among all pixels belonging to class *i*. MF1 represents the average of per-class F1-scores, enabling the evaluation of balanced classification performance across imbalanced classes. Additionally, FWF1 is weighted by the pixel proportion of each class, thereby reflecting the actual distribution of imbalanced land-cover classes in the assessment. These metrics are defined as follows:

$$OA(\%) = \frac{\sum_{i=1}^{n} TP_i}{Total\ number\ of\ pixels} \times 100 \tag{1}$$

$$Precision = \frac{TP_i}{TP_i + FP_i}, Recall = \frac{TP_i}{TP_i + FN_i} \tag{2}$$

$$F1_i(\%) = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \times 100 \tag{3}$$

$$MF1(\%) = \frac{1}{n} \sum_{i=1}^{n} F1 \tag{4}$$

$$FWF1(\%) = \frac{1}{Total\ number\ of\ pixels} \sum_{i=1}^{n} t_i \times F1 \tag{5}$$

where TP_i (True Positive) represents the number of pixels correctly classified as class *i*, TN_i (True Negative) represents the number of pixels that do not belong to class *i* and are correctly identified as not being class *i*, FP_i (False Positive) represents the number of pixels incorrectly identified as class *i*, and FN_i (False Negative) represents the number of pixels that truly belong to class *i* but are misclassified as other classes.

3. Experimental results

3.1. Dataset construction and class distribution

The experimental dataset in this study was constructed to quantify the quality–quantity tradeoff using scene-level Landsat LCC (Table 3).

Table 3
Training, validation and test dataset.

Land Cloud Cover (%)	Number of train data	Number of validation data	Number of total data	Land use class proportions (%) in the total train data						
				Urban	Farmland	Forest	Grass	Wetland	Barren	Water
0–2	5,325	689	6,014	5.23	18.14	65.31	2.52	1.48	1.48	5.84
0–5	6,852	1,087	7,939	5.11	18.53	64.29	2.39	1.58	1.44	6.66
0–10	7,936	1,363	9,299	5.06	18.69	64.13	2.39	1.58	1.44	6.71
0–20	10,943	1,457	12,400	5.05	19.14	64.20	2.47	1.56	1.45	6.14
0–30	11,816	1,511	13,327	5.35	19.29	63.30	2.49	1.62	1.50	6.46
0–40	12,078	1,523	13,601	5.29	19.24	63.48	2.49	1.61	1.49	6.40
0–50	14,023	1,583	15,606	5.32	19.11	63.74	2.50	1.53	1.50	6.30
Test dataset			689	6.83	13.63	55.91	2.47	1.27	1.38	18.51

Note: The proportions (%) of test dataset are based on the total data.

Under the strictest conditions (LCC 0–2%), the training and validation sets contained 5,325 and 689 samples, respectively. In contrast, an LCC of up to 50% resulted in 14,023 training and 1,583 validation samples, thereby securing approximately 2.6 times more training samples. The temporal distribution of total data samples varied substantially with the LCC threshold (Fig. S1). At LCC $\leq$ 2%, samples were concentrated in spring and late autumn (March: 1,817; November: 1,434), while June, August, and September—which fall within Korea's monsoon season—contained no observations. This seasonal concentration at low LCC thresholds introduces a potential informatics bias in ecological monitoring, as it underrepresents the peak vegetative vigor and phenological characteristics of the summer monsoon season. Conversely, relaxing the threshold to 20% recovered 1,194 samples during the June–September period. This approach facilitated an effective assessment of the combined effect of increased data volume and improved observation quality on model performance.

The class composition of the training data remained consistent across all LCC thresholds, with per-class proportions varying by no more than 1–2%. However, the degree of class imbalance was substantial. Forests (63–65%) and farmland (18–19%) constituted the two dominant classes, accounting for over 82% of the total training pixels. Urban and water each represented approximately 5–7%, while grassland (2.4–2.5%), wetland (1.4–1.6%), and barren land (1.4–1.6%) remained persistent minority classes. The forest-to-barren land and forest-to-wetland pixel ratios reached approximately 44:1, and the forest-to-grassland ratio was approximately 26:1 at the strictest LCC threshold. To ensure comparability across all experimental combinations, the test dataset was maintained constant using scenes filtered at an LCC threshold of 2% (689 samples). The proportions of farmland (13.6%) and forest (55.9%) were lower than those in the training data, while, water (18.5%) was substantially higher, reflecting the coastal geography of the test region. This configuration ensured that all performance metrics reflected the model's generalization to high-quality observations, regardless of the LCC conditions used during training.

3.2. Land use classification experimental results

We evaluated 252 experimental combinations and compared the configurations that resulted in the highest FWF1 and MF1 values (Table 4). Among the three models, U-Net achieved the highest overall performance, exceeding DeepLabv3+ and Swin-UPerNet by 0.53–0.68% in FWF1 and by 1.37–1.89% in MF1. Regarding FWF1, the optimal loss function varied by model: U-Net achieved its highest FWF1 (85.41%) with CE, whereas DeepLabv3+ and Swin-UPerNet achieved their best FWF1 values (84.88% and 85.23%, respectively) with FL. However, the highest MF1 for each model was consistently obtained with the FTL.

FTL demonstrated the most balanced class-level performance by substantially improving the minority class recall, although there were slight reductions in the dominant class F1-scores (Table S1). Across all loss functions, forest and water consistently achieved high F1-scores exceeding 91% and 95%, respectively, whereas urban and farmland F1-scores ranged between 60% and 73%. All models trained with CE and FL exhibited notably low recall for minority classes, with grass and barren below 10% and 18%, respectively, resulting in markedly reduced F1-scores for these classes. In contrast, FTL substantially improved the minority class recall, with grass increasing from $<$ 10% to 23–27% and barren increasing from $<$ 18% to 22–32% across all three models. However, this increase was accompanied by an approximately 1% decline in the F1-scores of the other classes compared with CE and FL.

The highest-performing model configurations for each metric predominantly employed the RGB + NIR + SWIR1/2 band combination. A comparison of the performance improvements across various band combinations indicated that the incorporation of the NIR band into the RGB significantly enhanced MF1 for all models. Particularly notable enhancements observed for urban area, grassland, and wetland (Table S2). The RGB + NIR + SWIR1/2 combination resulted in

Table 4
Overall and land-class accuracy of the highest-performing model for each metric across all models.

Model	Loss	Band	LCC threshold	OA	MF1	FWF1	Class-wise F1							The highest metrics
							Urban	Farm land	Forest	Grass	Wet land	Barren	Water	
U-Net	CE	RGB + NIR + SWIR1/2	20	**86.55**	57.96	**85.41**	**73.03**	**69.25**	**92.98**	15.19	38.59	20.28	**96.41**	FWF1
	FTL	RGB + NIR + SWIR1/2	40	84.62	**59.93**	84.90	71.69	67.73	92.08	**22.63**	**43.30**	**25.79**	96.26	MF1
DeepLab v3+	FL	RGB + NIR + SWIR1/2	30	85.87	56.80	84.88	71.18	68.23	92.66	16.31	32.88	20.12	96.24	FWF1
	FTL	RGB + NIR + SWIR1/2 + NTL	20	83.27	58.56	83.79	66.13	60.64	91.70	20.65	38.08	21.85	95.95	MF1
Swin-UPerNet	FL	RGB + NIR + SWIR1/2	20	86.26	58.04	85.23	72.20	68.88	92.85	16.89	37.48	21.75	96.21	FWF1
	FTL	RGB + NIR + SWIR1/2	50	83.21	59.00	84.34	71.47	66.79	91.51	20.93	42.15	23.91	96.22	MF1

marginal but consistent improvements in accuracy compared to those of the RGB + NIR combination, achieving the best performance across most land-cover classes. However, the combination that included NTL demonstrated a slight decline in performance. These results experimentally demonstrated that NTL, owing to its coarser spatial resolution relative to Landsat imagery, contributed minimally to the overall model performance.

Regarding the LCC threshold, all configurations demonstrating the highest levels of performance employed an LCC threshold of ≥20% (Table 4). In terms of FWF1, the optimal thresholds were LCC 20%, 30%, and 20% for U-Net, DeepLabv3+, and Swin-UPerNet, respectively. For MF1, the optimal threshold was LCC 40%, 20%, and 50% for each model. Compared to the strictest threshold of LCC 2%, the optimal configurations improved MF1 by up to 2.10% and FWF1 by up to 0.98%, despite the progressive increase in cloud-induced spectral noise (Table S3). These findings demonstrated that the data quantity should be considered along with the data quality when constructing training datasets.

The spatial patterns of the predictions for each model are shown in Fig. 4. Fig. 4a–c present the results of the highest FWF1 models for U-Net, DeepLabv3+, and Swin-UPerNet, respectively, while Fig. 4d–f show the corresponding highest MF1 models trained with FTL.As shown in the annotated regions (1) and (2) in Fig. 4, the highest FWF1 models across all three architectures did not identify minority land cover classes such as wetland, grassland, and barren land. In contrast, the highest MF1 models trained with the FTL demonstrated a substantial improvement in the classification of these minority classes. Additionally, as shown in the annotated region (3) in Fig. 4, U-Net outperformed the other models in classifying fine-grained boundaries and spatial patterns. It captured urban pixels more precisely than the GT land cover map and distinctly represented the satellite imagery. However, as shown in the annotated region (4) in Fig. 4, arterial roads located at the urban periphery were labeled as urban areas in the GT map, yet all models consistently classified them as farmland.

The per-class recall and misclassification patterns of the highest-performing models based on MF1 are shown in Fig. 5. All three

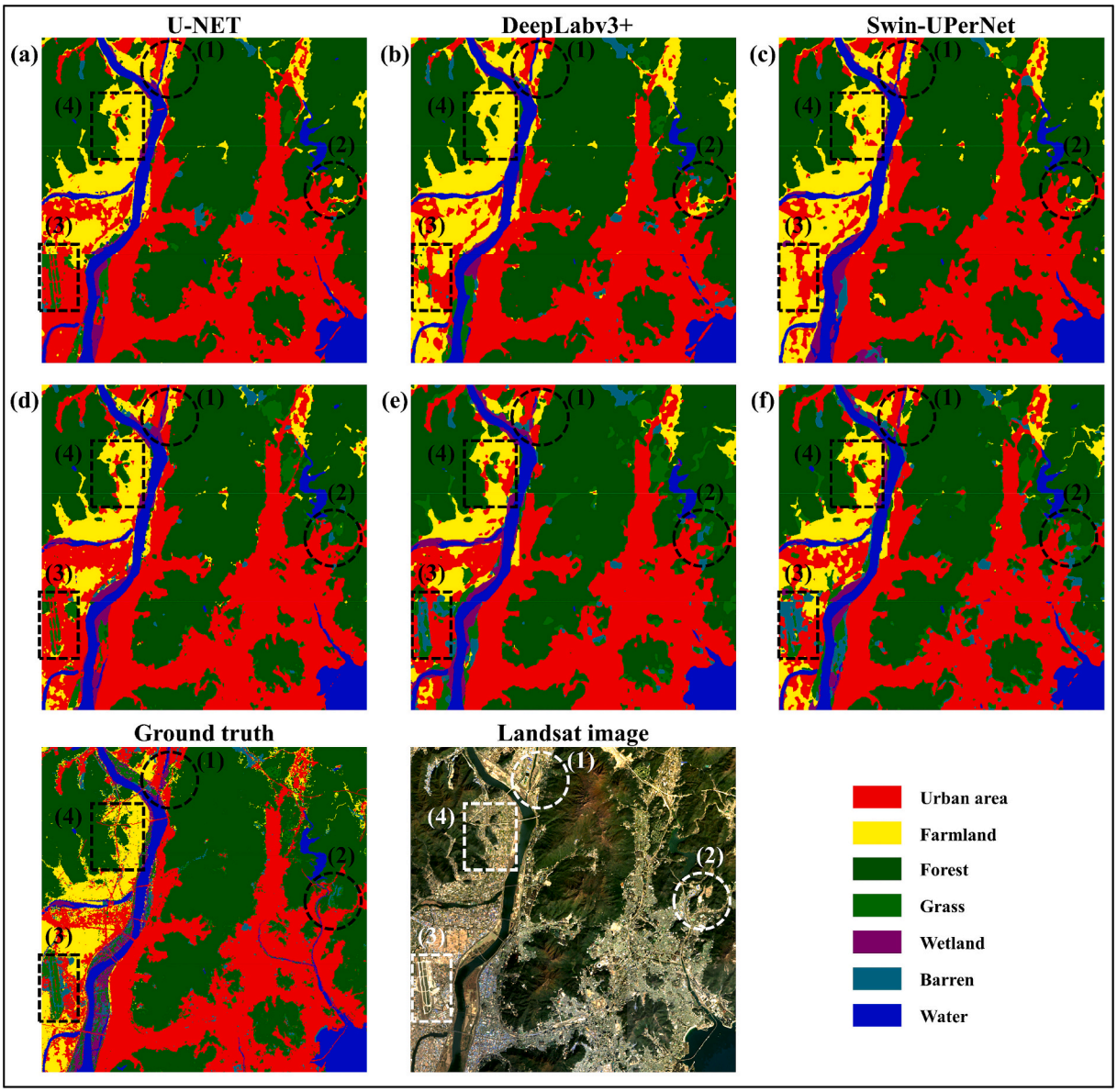


Fig. 4. Visualization of land use classification of each of the highest performing models: (a)–(c) are the results of the highest performing FWF1 models, and (d)–(f) are the results of the highest performing MF1 models: trained using (a) CE, (b) and (c) FL, and (d)–(f) FTL. Annotated regions (1)–(4) highlight specific areas discussed in the text: (1) minority class distribution, (2) minority class absence, (3) fine-grained boundary classification, and (4) arterial road misclassification.

(a)

Ground truth	Urban	Farmland	Forest	Grass	Wetland	Barren	Water
	Detected						
Urban	**75.9**	12.0	4.5	2.1	0.7	3.3	1.4
Farmland	7.9	**72.5**	12.9	3.7	1.1	1.2	0.7
Forest	1.0	5.9	**89.8**	2.2	0.2	0.3	0.5
Grass	12.0	29.8	21.6	**23.4**	7.7	2.8	2.6
Wetland	6.1	25.7	4.0	4.3	**44.7**	2.3	13.0
Barren	26.0	27.4	12.0	6.9	2.4	**22.6**	2.8
Water	0.4	0.7	0.6	0.1	1.4	0.3	**96.6**

(b)

Ground truth	Urban	Farmland	Forest	Grass	Wetland	Barren	Water
	Detected						
Urban	**68.3**	15.6	5.7	3.4	1.0	4.3	1.7
Farmland	5.1	**69.3**	16.6	5.1	1.6	1.7	0.6
Forest	0.8	5.8	**89.1**	3.0	0.2	0.5	0.5
Grass	8.0	28.7	25.1	**22.6**	8.4	4.8	2.3
Wetland	5.3	24.1	9.1	3.8	**43.5**	3.0	11.2
Barren	17.4	30.2	12.7	8.0	2.8	**26.3**	2.5
Water	0.4	0.7	0.9	0.1	1.1	0.4	**96.5**

(c)

Ground truth	Urban	Farmland	Forest	Grass	Wetland	Barren	Water
	Detected						
Urban	**71.5**	11.6	3.3	4.1	1.1	6.9	1.5
Farmland	6.2	**70.9**	10.0	7.1	1.8	3.3	0.7
Forest	0.9	6.3	**87.8**	3.3	0.2	0.9	0.5
Grass	8.7	26.3	19.4	**27.4**	8.7	7.5	2.1
Wetland	4.4	20.8	3.1	5.0	**46.3**	6.4	14.0
Barren	18.9	25.1	9.2	9.3	3.2	**32.2**	2.2
Water	0.3	0.5	0.4	0.1	1.2	1.0	**96.4**

Fig. 5. Recall and misclassification rates for the total number of classes for each model: (a), (b), and (c) U-Net, DeepLabv3+, and Swin-UPerNet, respectively.

models achieved a recall of ≥96% for water, despite its relatively low proportion in the training dataset (approximately 5%) in the overall composition. This high performance can be attributed to the spectrally distinct reflectance properties of water compared to other land cover classes, as well as its spatially clustered distribution. Forest, the most dominant land cover class, resulted in a recall of ≥87% across all three models, whereas urban and farmland areas achieved approximately 70% recall. U-Net demonstrated the highest recall for major land cover classes, including urban, farmland, forest, and water (Fig. 5a). DeepLabv3+ showed comparatively lower recall across most land cover classes among the three models (Fig. 5b), with urban areas exhibiting the most notable deficit. Furthermore, it achieved the lowest recall not only for farmland but also for grassland and wetland, suggesting a limited capacity to capture both dominant and minority classes simultaneously. Swin-UPerNet exhibited the best classification performance among the three models for minority classes such as grassland, wetland, and barren land. These results suggest that transformer-based architectures have a comparative advantage in extracting discriminative features from underrepresented land-cover classes (Fig. 5c). This comparative advantage may be attributed to the global context modeling enabled by window-based self-attention, which enables Swin-UPerNet to effectively capture the spatially dispersed and heterogeneous patterns characteristic of barren and wetland cover. Except for water, most land cover classes were misclassified as farmland. The relatively high misclassification rate of urban areas as farmland likely stems from two factors: spatial adjacency and boundary ambiguity between the two classes, and the mixed-pixel effect resulting from the labeling of peri-urban roads as urban areas in the training data. Grass was most frequently confused with farmland and forest, whereas barren land was predominantly misclassified as farmland and urban land. In addition, wetlands located along river corridors were frequently misidentified as farmlands. These misclassification patterns highlight a fundamental spectral ambiguity where grassland is frequently confused with forest and urban areas, and wetland is misidentified as water or forest. Notably, these boundaries remain challenging to distinguish even through expert visual interpretation, suggesting that the inherent limitations of medium-resolution spectral signals—rather than simple model failure—drive these informatics bottlenecks.

4. Discussion

This study systematically analyzed four factors: model architecture, loss function, spectral band combinations, and LCC thresholds across 252 experimental configurations using Landsat 8 satellite imagery. These experiments provided empirical guidance on training data composition and model design for medium-resolution LULC classification.

The performance comparison of the three models provided practical guidance for designing models suited to medium-to-low spatial resolution Landsat imagery. Despite being the most conventional segmentation architecture, U-Net achieved the best overall performance by utilizing boundary-preserving skip connections and progressive 2 times stepwise upsampling (Table 4). In contrast, architecture like DeepLabv3+ and Swin-UPerNet, which restore resolution through a single large bilinear interpolation step, inevitably sacrifice fine-grained spatial details at class boundaries (Tsiporenko et al., 2025). Unlike high-resolution imagery, medium-resolution Landsat pixels (30 m) frequently contain mixed spectral characteristics, rendering boundary information inherently ambiguous. Our findings suggest that a progressive recovery strategy effectively reduces this reconstruction uncertainty. However, Swin-UPerNet's window-based self-attention proves indispensable for identifying fragmented natural landscapes like wetlands and barren land, which lack the rigid, linear boundaries typical of urban structures. While U-Net excels at capturing well-defined edges, the Swin-Transformer's ability to learn long-range dependencies allows it to recognize spatially dispersed ecological patterns that are often lost during the rigid downsampling of standard CNNs. Therefore, these results indicate that a hybrid architecture represents the optimal pathway for medium-resolution environmental modeling. Specifically, intergrating Swin-Transformer's global context modeling with U-Net's

progressive upsampling would jointly address the dual challenges of boundary reconstruction and minority class classification.

In Landsat-based LULC classification, hybrid loss functions provide a more effective solution to class imbalances than those of single-component alternatives. Although FL assigns higher weights to hard-to-classify samples, its sample difficulty-based weighting conflates class frequency with spectral and geographic complexity. This conflation limits FL's ability to capture the distributional characteristics of extremely rare land cover classes. The FTL overcomes this limitation by jointly incorporating FL and Tversky losses to address the class imbalance. This integration improved the F1-scores of minority classes such as grassland and barren land by approximately 10% relative to those of CE and FL (Table S1). Although FTL marginally reduced the F1-scores of dominant classes, such as forest and water, this tradeoff is acceptable owing to the substantial improvements observed for minority classes. Therefore, hybrid loss functions such as FTL are particularly appropriate for LULC classification tasks. Accurate mapping of ecologically important but spatially rare land cover types—such as grass, wetland, and barren land—is essential for land use monitoring and planning.

The integration of SWIR1 and SWIR2 bands significantly enhanced the classification accuracy for urban impervious surfaces, wetlands, and barren land (Yu et al., 2019). This improvement stems from the Short-Wave Infrared (SWIR) region's sensitivity to surface moisture and material composition, providing critical discriminative power where RGB and NIR bands often overlap. For urban areas, SWIR bands effectively differentiate artificial structures from dry barren land (Wu and Murray, 2003). While both classes exhibit high reflectance in the visible and NIR spectra, SWIR bands are highly responsive to the unique spectral signatures of concrete and asphalt compared to natural soil. In wetland classification, the SWIR region serves as a ‘spectral fingerprint’ due to the strong absorption properties of water (Tian et al., 2016). The high moisture content in wetland soil and vegetation results in significantly lower SWIR reflectance than in drier forest or agricultural land, allowing the model to better separate wetlands from spectrally similar classes like grasslands. Despite these enhancements, grassland and wetland remain inherently difficult to classify due to significant spectral overlap with forest and water bodies. This spectral ambiguity is a fundamental informatics challenge; for instance, the transition from grassland to forest represents a continuum of vegetative vigor that is often indistinguishable even through expert visual interpretation. Our results suggest that future research should focus on multi-temporal informatics or hybrid loss functions to further disentangle these visually ambiguous ecological boundaries.

The results of this study empirically demonstrated the quantitative gains achievable through systematic experimentation with diverse spectral band combinations in deep learning-based models. This provides a broader understanding of multispectral data utilization in deep learning-based LCMs, complementing the single-band experiments conducted by Boonpook et al. (2023). The addition of the NIR band to the RGB improved MF1 by approximately 2–5% across all models, with particularly notable increases in the F1-scores of urban, grassland, farmland, and wetland areas (Table S2). This improvement is likely attributable to the sensitivity of the NIR band to vegetation vigor, which provides critical spectral information for discriminating urban areas, grasslands, and wetlands from farmlands. These findings are consistent with those reported by Zhang et al. (2018) and Tian et al. (2016). Additionally, as the SWIR bands contained information on the soil and impervious surface moisture content, the RGB + NIR + SWIR1/2 combination resulted in an improved classification of impervious surfaces in urban areas, farmland, and wetlands. By contrast, the marginal or negative performance changes observed with the RGB + NIR + SWIR1/2 + NTL combination underscore the importance of spatial resolution consistency among the input data. Auxiliary data such as NTL can assist in urban boundary delineation (Sharma et al., 2016; Xue et al., 2018). However, when the NTL data—originally acquired at approximately 500 m resolution—were resampled to 30 m, adjacent farmland and forest pixels were assigned the same luminance value as urban pixels. This spatial mismatch potentially hinders, rather than supports, fine-grained class discrimination. In this context, our experimental findings provide evidence for the argument by Li et al. (2024) that low-resolution auxiliary data can compromise the efficacy of classification processes in pixel-based segmentation. Furthermore, our results indicate that achieving a substantial performance enhancement requires auxiliary data with a spatial resolution equal to or exceeding that of the primary training data.

Regarding data quality and quantity, the LCC threshold setting necessitates careful consideration of both dimensions simultaneously rather than prioritizing quality alone. The results of this study indicated that all models achieved their best performance at an LCC threshold of $\geq$20%. This indicates that when the training dataset comprises approximately $\geq$10,000 images, the increased diversity available for model learning, combined with the generalization benefits gained from a larger data volume, can offset the quality degradation introduced by cloud-induced noise. This is consistent with the findings reported by Phinzi et al. (2023) regarding Random Forest-based LULC mapping. Additionally, the use of stringent LCC thresholds tends to concentrate training scenes in cloud-free seasons such as spring and autumn. In our dataset, the summer months (June–September) accounted for only 6% of samples at LCC $\leq$ 2%. However, at LCC 20% and LCC 50%, data samples increased to 14% and 27%, respectively. This could potentially underrepresent the spectral characteristics associated with summer crop conditions and seasonally varying vegetation phenology. This seasonal concentration may partly explain the relatively low classification accuracy for farmland and grassland, as these classes exhibit pronounced seasonal spectral variability—farmland transitions from bare soil in spring to full crop canopy in summer, while grassland spectral signatures converge with those of forest during non-growing seasons. The performance gains observed at higher LCC thresholds ($\geq$20%) may therefore reflect not only increased data volume but also improved seasonal representativeness of the training set. Strictly filtering data at LCC $<$ 10% introduces a severe seasonal informatics bias, effectively ‘blinding’ the model to the peak phenological states of the monsoon season. By relaxing the LCC threshold to 20%, we captured critical phenological signatures of summer crops and wetland vegetation, demonstrating that temporal diversity in ecological data is more vital for model robustness than the pristine quality of individual scenes. This finding has practical implications for the construction of training data in Landsat-based LULC research, which commonly employs LCC thresholds of $<$ 10%.

Although the methodology employed in this study for training dataset construction is highly practical, the quality limitations of publicly available GT data and the issue of class–label inconsistencies must be carefully considered. Specifically, as shown in the annotated region (3) in Fig. 4d, we observed cases where the model classifications represented actual land surface conditions more accurately than the public GT data. This discrepancy likely stems from temporal inconsistencies between field surveys and map production dates, which prevent public maps from adequately capturing rapid land-cover changes. Such findings suggest a paradigm shift where optimized informatics workflows serve as an automated data refinement tool; AI is no longer just a consumer of ecological data, but an active participant in validating and updating historical environmental archives. Therefore, a quality verification process for public GT data, conducted under the supervision of researchers and practitioners, is warranted in future studies. Additionally, as demonstrated in annotated region (4) (Fig. 4a-f), class misclassification can occur when linear structures like arterial roads are labeled as urban but mix with adjacent farmland in medium-resolution pixels. Incorporating a preprocessing step that leverages road network data to separate transportation infrastructure from the urban class may significantly improve classification accuracy in medium-to-low resolution environments.

5. Conclusion

This study proposed an effective informatics methodology for constructing large-scale training datasets by integrating public land-cover maps with Landsat 8 imagery. By systematically evaluating 252 experimental configurations, we established empirical guidelines for optimizing medium-resolution LULC classification, particularly for long-term ecological monitoring.

The key findings are as follows: First, U-Net achieved the best overall performance, demonstrating the efficacy of progressive upsampling and skip connections in preserving fine-grained boundaries at a 30 m resolution. Second, the Focal Tversky Loss (FTL) consistently improved F1-scores for ecologically sensitive natural classes like wetlands and grasslands by up to 10%. Third, the RGB + NIR + SWIR1/2 combination was optimal for identifying complex surface compositions, whereas auxiliary NTL data provided negligible benefits due to spatial resolution mismatch. Finally, we demonstrated that increased data volume (> 10,000 samples) at LCC thresholds ≥20% effectively offsets spectral noise and mitigates seasonal informatics bias. Overall, these findings contribute to enable high-frequency monitoring of urban and vegetation surface dynamics, thereby supporting a comprehensive understanding of the complex, nonlinear interactions between ecosystems and human activities. The resulting LULC data can be integrated into AI-based simulation models for climate change (Guo et al., 2023, 2024b) and air pollution (Guo et al., 2025; He et al., 2025a, 2026), providing an empirical foundation for evidence-based development and conservation decision-making.

Despite its practical contributions, this study is limited by its geographic focus and the persistent challenge of spectral ambiguity between natural classes. Future research should investigate hybrid architectures—combining Swin-Transformer's global context modeling with U-Net's progressive structure—and season-stratified sampling strategies to further refine automated ecological archives. These findings offer a robust informatics roadmap for reconstructing long-term environmental change patterns from archival satellite data.

CRediT authorship contribution statement

Juyeong Nam: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Younju Lee:** Visualization, Software, Investigation, Data curation, Conceptualization. **Changyeon Lee:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Changyeon Lee reports financial support was provided by National Research Foundation of Korea. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (RS-2025-16065102).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103912.

Data availability

The datasets utilized in this study were obtained from publicly available sources. Landsat 8 imagery was acquired from https://earthexplorer.usgs.gov/ (U.S. Geological Survey, 2024). The land cover map used as reference data was obtained from https://aid.mcee.go.kr/ (The Environmental Geographic Information Service (EGIS), 2024), provided by the Ministry of Climate, Energy, and Environment of the Republic of Korea. The code used in this study was developed based on the MMSegmentation framework, available at https://github.com/open-mmlab/mmsegmentation (MMSegmentation Contributors, 2024), and the modified version is publicly available at Doi: https://doi.org/10.6084/m9.figshare.32044008 (Nam et al., 2026).

References

Abraham, N., Khan, N.M., 2019. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In: IEEE 16th International Symposium On Biomedical Imaging (ISBI 2019), pp. 683–687. https://doi.org/10.48550/arXiv.1810.07842.

Bagan, H., Yamagata, Y., 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. GIsci. Remote Sens. 52, 765–780. https://doi.org/10.1080/15481603.2015.1072400.

Boonpook, W., Tan, Y., Nardkulpat, A., Torsri, K., Torteeka, P., Kamsing, P., Sawangwit, U., Pena, J., Jainaen, M., 2023. Deep learning semantic segmentation for land use and land cover types using Landsat 8 imagery. ISPRS Int. J. Geoinf. 12, 14. https://doi.org/10.3390/ijgi12010014.

Cao, Y., Jiang, N., Wang, D.-H., Wu, Y., Zhu, S., 2024. UAM-Net: an attention-based multi-level feature fusion UNet for remote sensing image segmentation. In: Liu, Q., et al. (Eds.), Pattern Recognition and Computer Vision. PRCV 2023. Lecture Notes in Computer Science, vol. 14428. Springer, Singapore, pp. 267–278. https://doi.org/10.1007/978-981-99-8462-6_22.

Chen, S., Yao, S., 2023. Identifying the drivers of land expansion and evaluating multi-scenario simulation of land use: a case study of Mashan County, China. Ecol. Inform. 77, 102201. https://doi.org/10.1016/j.ecoinf.2023.102201.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018, Lecture Notes in Computer Science, vol. 11211. Springer, Cham, pp. 833–851. https://doi.org/10.1007/978-3-030-01234-2_49.

Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258. https://doi.org/10.1109/CVPR.2017.195.

Dhinwa, P.S., Pathan, S.K., Sastry, S.V.C., Rao, M., Majumder, K.L., Chotani, M.L., Singh, J.P., Sinha, R.L.P., 1992. Land use change analysis of Bharatpur district using GIS. J. Indian Soc. Remote Sens. 20, 237–250. https://doi.org/10.1007/BF03001921.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR 2021) arXiv preprint.

Dou, P., Shen, H., Huang, C., Li, Z., Mao, Y., Li, X., 2024. Large-scale land use/land cover extraction from Landsat imagery using feature relationships matrix based deep-shallow learning. Int. J. Appl. Earth Obs. Geoinf. 129, 103866. https://doi.org/10.1016/j.jag.2024.103866.

Earth Observation Group, 2024. VIIRS Nighttime Day/Night Band Composites Version 1 (Dataset). Accessed mar 20, 2025. https://developers.google.com/earth-engine/datasets/catalog/NOAA_VIIRS_DNB_MONTHLY_V1_VCMCFG?hl=ko.

Goldblatt, R., Stuhlmacher, M.F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A.K., Cheng, W.-H., Balling, R.C., 2018. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. Remote Sens. Environ. 205, 253–275. https://doi.org/10.1016/j.rse.2017.11.026.

Goyal, P., Dollar, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K., 2018. Accurate, Large Minibatch SGD: training ImageNet in 1 Hour. arXiv preprint. https://doi.org/10.48550/arXiv.1706.02677.

Gui, B., Bhardwaj, A., Sam, L., 2025. A novel multi-scale deep learning framework for adaptive urban expansion simulation. Sustain. Cities Soc. 130, 106594. https://doi.org/10.1016/j.scs.2025.106594.

Guo, Q., He, Z., Wang, Z., 2023. Prediction of monthly average and extreme atmospheric temperatures in Zhengzhou based on artificial neural network and deep learning models. Front. For. Glob. Change 6. https://doi.org/10.3389/ffgc.2023.1249300.

Guo, Q., He, Z., Wang, Z., 2024a. Monthly climate prediction using deep convolutional neural network and long short-term memory. Sci. Rep. 14, 17748. https://doi.org/10.1038/s41598-024-68906-6.

Guo, Q., He, Z., Wang, Z., Qiao, S., Zhu, J., Chen, J., 2024b. A performance comparison study on climate prediction in Weifang City using different deep learning models. Water (Basel). 16, 2870. https://doi.org/10.3390/w16192870.

Guo, Q., He, Z., Wang, Z., 2025. Assessing the effectiveness of long short-term memory and artificial neural network in predicting daily ozone concentrations in Liaocheng City. Sci. Rep. 15, 6798. https://doi.org/10.1038/s41598-025-91329-w.

He, Z., Guo, Q., 2024. Comparative analysis of multiple deep learning models for forecasting monthly ambient PM2.5 concentrations: a case study in Dezhou City, China. Atmosphere (Basel). 15, 1432. https://doi.org/10.3390/atmos15121432.

He, T., Zhang, Zhi, Zhang, H., Zhang, Zhongyue, Xie, J., Li, M., 2018. Bag of tricks for image classification with convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 558–567. Doi: 10.48550/arXiv.1812.01187.

He, Z., Guo, Q., Wang, Z., Li, X., 2025a. A hybrid wavelet-based deep learning model for accurate prediction of daily surface PM2.5 concentrations in Guangzhou City. Toxics 13, 254. https://doi.org/10.3390/toxics13040254.

He, Z., Guo, Q., Yin, Z., Luo, X., Sun, M., 2025b. Evolution and driving mechanisms of eco-environmental quality across different urbanization stages. Sci. Rep. 15, 22101. https://doi.org/10.1038/s41598-025-06084-9.

He, Z., Guo, Q., Zhang, Z., Feng, G., Qiao, S., Wang, Z., 2026. Forecasting daily ambient PM2.5 concentrations in Qingdao City using deep learning and hybrid interpretable models and analysis of driving factors using SHAP. Toxics 14, 44. https://doi.org/10.3390/toxics14010044.

Keola, S., Andersson, M., Hall, O., 2015. Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. World Dev. 66, 322–334. https://doi.org/10.1016/j.worlddev.2014.08.017.

Khan, M., Hanan, A., Kenzhebay, M., Gazzea, M., Arghandeh, R., 2024. Transformer-based land use and land cover classification with explainability using satellite imagery. Sci. Rep. 14, 16744. https://doi.org/10.1038/s41598-024-67186-4.

Lamsal, P., Atreya, K., Pant, K.P., Kumar, L., 2015. An analysis of willingness to pay for community-based conservation activities at the Ghodaghodi Lake Complex, Nepal. International Journal of Biodiversity Science, Ecosystem Services & Management 11, 341–348. https://doi.org/10.1080/21513732.2015.1055338.

Li, Z., Chen, B., Wu, S., Su, M., Chen, J.M., Xu, B., 2024. Deep learning for urban land use category classification: a review and experimental assessment. Remote Sens. Environ. 311, 114290. https://doi.org/10.1016/j.rse.2024.114290.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2018. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988. https://doi.org/10.48550/arXiv.1708.02002.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022. https://doi.org/10.48550/arXiv.2103.14030.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

MMSegmentation Contributors, 2024. MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark. https://github.com/open-mmlab/mmsegmentation.

Nam, J., Lee, C., 2025. Forecasting urban expansion: a dynamic urban growth model using DS-ConvLSTM to simulate multi-land regulation scenarios. Ecol. Inform. 88, 103136. https://doi.org/10.1016/j.ecoinf.2025.103136.

Nam, J., Lee, Y., Lee, C., 2026. Code for Toward Optimized Deep Learning-Based Land Use/Land Cover Classification Using Landsat 8 (Dataset). Figshare. https://doi.org/10.6084/m9.figshare.32044008.

NASA, 2013. Landsat 8 Bands. https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-bands/.

Ni, Y., Liu, J., Cui, J., Yang, Y., Wang, X., 2023. Edge guidance network for semantic segmentation of high-resolution remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 9382–9395. https://doi.org/10.1109/JSTARS.2023.3316307.

Phinzi, K., Ngetar, N.S., Pham, Q.B., Chakilu, G.G., Szabá, S., 2023. Understanding the role of training sample size in the uncertainty of high-resolution LULC mapping using random forest. Earth Sci. Inf. 16, 3667–3677. https://doi.org/10.1007/s12145-023-01117-1.

Phiri, D., Morgenroth, J., Xu, C., Hermosilla, T., 2018. Effects of pre-processing methods on Landsat OLI-8 land cover classification using OBIA and random forests classifier. Int. J. Appl. Earth Obs. Geoinf. 73, 170–178. https://doi.org/10.1016/j.jag.2018.06.014.

Ramos, L.T., Sappa, A.D., 2025. Leveraging U-Net and selective feature extraction for land cover classification using remote sensing imagery. Sci. Rep. 15, 784. https://doi.org/10.1038/s41598-024-84795-1.

Redlich, S., Zhang, J., Benjamin, C., Dhillon, M.S., Englmeier, J., Ewald, J., Fricke, U., Ganuza, C., Haensel, M., Hovestadt, T., Kollmann, J., Koellner, T., Kübert-Flock, C., Kunstmann, H., Menzel, A., Moning, C., Peters, W., Riebl, R., Rummler, T., Rojas-Botero, S., Tobisch, C., Uhler, J., Uphus, L., Müller, J., Steffan-Dewenter, I., 2022. Disentangling effects of climate and land use on biodiversity and ecosystem services—a multi-scale experimental design. Methods Ecol. Evol. 13, 514–527. https://doi.org/10.1111/2041-210X.13759.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351. Springer, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Shao, Y., Xiao, Y., Kou, X., Sang, W., 2023. Sustainable land use scenarios generated by optimizing ecosystem distribution based on temporal and spatial patterns of ecosystem services in the southern China hilly region. Ecol. Inform. 78, 102275. https://doi.org/10.1016/j.ecoinf.2023.102275.

Sharma, R.C., Tateishi, R., Hara, K., Gharechelou, S., Iizuka, K., 2016. Global mapping of urban built-up areas of year 2014 by combining MODIS multispectral data with VIIRS nighttime light data. Int. J. Digit. Earth 9, 1004–1020. https://doi.org/10.1080/17538947.2016.1168879.

Talukdar, S., Singha, P., Shahfahad, Mahato, S., Praveen, B., Rahman, A., 2020. Dynamics of ecosystem services (ESs) in response to land use land cover (LU/LC) changes in the lower Gangetic plain of India. Ecol. Indic. 112, 106121. https://doi.org/10.1016/j.ecolind.2020.106121.

The Environmental Geographic Information Service (EGIS), 2024. Land use and land cover in the late 2010s. (Dataset). Accessed Feb 29, 2024. https://aid.mcee.go.kr/.

Tian, S., Zhang, X., Tian, J., Sun, Q., 2016. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. Remote Sens. 8, 954. https://doi.org/10.3390/rs8110954.

Tsiporenko, I., Chizhov, P., Fishman, D., 2025. Going beyond U-net: assessing vision transformers for semantic segmentation in microscopy image analysis. In: Del Bue, A., Canton, C., Pont-Tuset, J., Tommasi, T. (Eds.), Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science, vol. 15638. Springer, Cham, pp. 222–238. https://doi.org/10.1007/978-3-031-91721-9_14.

U.S. Geological Survey, 2024. USGS EarthExplorer. Accessed Mar 20, 2025. https://earthexplorer.usgs.gov/.

Vali, A., Comai, S., Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: a review. Remote Sens. 12, 2495. https://doi.org/10.3390/rs12152495.

Wilkinson, G.G., 2005. Results and implications of a study of fifteen years of satellite image classification experiments. IEEE Trans. Geosci. Remote Sens. 43, 433–440. https://doi.org/10.1109/TGRS.2004.837325.

Wu, C., Murray, A.T., 2003. Estimating impervious surface distribution by spectral mixture analysis. Remote Sens. Environ. 84, 493–505. https://doi.org/10.1016/S0034-4257(02)00136-0.

Xu, X., Kong, W., Wang, L., Wang, T., Luo, P., Cui, J., 2024. A novel and dynamic land use/cover change research framework based on an improved PLUS model and a fuzzy multiobjective programming model. Ecol. Inform. 80, 102460. https://doi.org/10.1016/j.ecoinf.2024.102460.

Xue, X., Yu, Z., Zhu, S., Zheng, Q., Weston, M., Wang, K., Gan, M., Xu, H., 2018. Delineating urban boundaries using Landsat 8 multispectral data and VIIRS nighttime light data. Remote Sens. 10, 799. https://doi.org/10.3390/rs10050799.

Yu, Z., Di, L., Yang, R., Tang, J., Lin, L., Zhang, C., Rahman, Md.S., Zhao, H., Gaigalas, J., Yu, E.G., Sun, Z., 2019. Selection of Landsat 8 OLI band combinations for land use and land cover classification. In: 2019 8th International Conference on Agro-Geoinformatics. IEEE, pp. 1–5. https://doi.org/10.1109/Agro-Geoinformatics.2019.8820595.

Zang, N., Cao, Y., Wang, Y., Huang, B., Zhang, L., Mathiopoulos, P.T., 2021. Land-use mapping for high-spatial resolution remote sensing image via deep learning: a review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5372–5391. https://doi.org/10.1109/JSTARS.2021.3078631.

Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., Zhang, S., 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. Sensors 18, 3717. https://doi.org/10.3390/s18113717.

Zhang, X., Han, Liangxiu, Han, Lianghao, Zhu, L., 2020. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sens. 12, 417. https://doi.org/10.3390/rs12030417.

Zhang, Y., Huang, M., Chen, Y., Xiao, X., Li, H., 2025. Land cover classification in high-resolution remote sensing: using Swin transformer deep learning with texture features. J. Spat. Sci. 70, 205–229. https://doi.org/10.1080/14498596.2024.2386317.

Zhu, Q., Guo, X., Deng, W., Shi, S., Guan, Q., Zhong, Y., Zhang, L., Li, D., 2022. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 184, 63–78. https://doi.org/10.1016/j.isprsjprs.2021.12.005.