Enhancing Lake Identification in Alpine Periglacial Environments by Leveraging the Global Context of Transformers

Jinhao Xu[1], Min Feng[1,2], Yijie Sui[1], Yanan Su[1], Xuefei Zhang[3], Qinglin Wu[1,2], Zhimin Hu[1,4], Ruilin Wang[1,5]

[1]National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Land Satellite Remote Sensing Application Center, Ministry of Natural Resources, Beijing 100048, China

[4]College of Geography and Tourism, Chongqing Normal University, Chongqing 401331, China

[5]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China

Correspondence to: Min Feng (mfeng@itpcas.ac.cn)

Abstract. Lakes in alpine periglacial environments, as sensitive indicators of cryospheric change, are undergoing rapid expansion under global warming. Investitating their evolving distribution is essential for monitoring climate understanding impacts and assessing associated geohazards. The complex topography and heterogeneous landscapes in high-mountain regions pose significant challenges for conventional identification methods, leading to the underdetection of small lakes, elevated false positive rates, and limited ability to discriminate between lake formation types. This study introduces a Vision Transformer (ViT)-based identification framework for lakes in alpine periglacial environments, employing a two-step process of lake boundary segmentation and type classification. By leveraging ViT's global attention mechanism, the framework captures long-range spatial and spectral relationships, enhancing contextual understanding of lakes and their surroundings. Compared to CNN-based models, the ViT-based approach achieved a mean intersection over union (mIoU) of 91.01% for segmentation and an F1-score of 89.75% for classification. It significantly improved detection of small lakes (as small as 0.0001 km²), reduced artifacts from shadows, snow, ice, and river fragments, and provided a more accurate lake type classification. Applied to the Southeastern Tibetan Plateau Gorge Region, a region with high glacial lake density and outburst flood risks, the framework identified 3,266 lakes (1,708 glacial and 1,558 non-glacial), surpassing existing inventories in completeness and accuracy.

1. Introduction

The alpine periglacial environment refers to high-altitude mountain regions dominated by cold conditions, where freeze-thaw cycles, snowmelt, and low-temperature physical weathering processes prevail (Péwé, 1969; French, 2017). Lakes occurring within alpine periglacial environments serve as critical indicators of cryospheric changes (Haeberli et al., 2001; García-Rodríguez et al., 2021). Against the backdrop of global warming, the persistent net loss of ice in the cryosphere is driving the rapid expansion of lakes in alpine periglacial environments (Zhang et al., 2023; Wang et al., 2025). Following the classification framework of Yao et al., (2018), lakes in alpine periglacial environments are categorized into glacial and non-glacial lakes based on their modern glacial influence and present-day hydrological processes. Glacial lakes refer to natural water bodies that are mainly supplied by modern glacial meltwater or formed in depressions of glacier moraines, and are therefore directly coupled to ongoing glacier dynamics. These lakes are highly sensitive to climate change and commonly exhibit unstable moraine-dammed configurations, which substantially increase their susceptibility to glacial lake outburst floods (Basnett et al., 2013; Veh et al., 2022). Non-glacial lakes, in contrast, refer to all other lakes within the alpine periglacial environment that do not meet the above criteria for glacial lakes, i.e., they do not receive direct input from modern glacier meltwater and are not formed in association with active glacier moraines. Their hydrological regimes are primarily controlled by precipitation, snowmelt, or groundwater inputs under current conditions, and they generally exhibit more stable configurations and lower sensitivity to abrupt glacier-related disturbances (Luo et al., 2018; Larsen et al., 2024). Investigating the distribution, types, and evolution of lakes in alpine periglacial environments is therefore critical for understanding cryospheric responses to climate change, managing high-mountain water resources, and assessing glacial and periglacial geohazards.

Traditional field surveys, constrained by the inaccessibility of high-altitude environments and high observation costs, struggle to achieve large-scale, continuous monitoring (Nagendra and Rocchini, 2008; Avtar et al., 2020). Currently, automated identification techniques based on remote sensing data have become a fundamental method for investigating lakes in alpine periglacial environments (Liaudat et al., 2012; Romashova and Chernov, 2023). However, accurately identifying lakes in alpine periglacial environments poses three major scientific challenges: (1) Small lakes dominate in number but are difficult to detect. Globally, it is estimated that glacial lakes smaller than 0.1 km² account for over 75% of the total number (Zhang et al., 2024b). These lakes exhibit limited information in imagery and are susceptible to sub-pixel spectral mixing effects (Li and Sheng, 2012). Existing remote sensing studies typically set area thresholds at ≥ 0.001 km², a threshold primarily imposed by sensor spatial resolution and practical considerations for large-scale mapping, leaving smaller-scale lakes in alpine periglacial environments without systematic observational data (Nie et al., 2017; Chen et al., 2021); (2) Complex mountain topography and variable meteorological conditions amplify remote sensing interpretation errors. Shadows cast by steep terrain exhibit low reflectance in the visible to near-infrared bands, resembling water bodies, while seasonal snow and thin ice cover further distort the spectral characteristics of water, adding complexity to the identification of water bodies (Barbieux et al., 2018; Zhao et al., 2025); (3) Effective differentiation between glacial and non-glacial lakes remains elusive. These two lake types differ significantly in formation mechanisms and disaster susceptibility: glacial lakes depend on glacier ablation dynamics and carry high outburst risks, whereas non-glacial lakes, governed by thermodynamics or precipitation, are structurally more stable (Huggel et al., 2002; Buckel et al., 2018). Misclassification of these lake types may introduce systematic biases in analysis and assessment.

However, existing identification methods exhibit systematic limitations in addressing these challenges. Spectral thresholding, while efficient in delineating water bodies, is highly sensitive to topographic shadows, snow, and ice cover, frequently yielding false positives or missing small lakes due to complex illumination conditions in mountainous terrain (Zhao et al., 2018; Wang et al., 2020; Peppa et al., 2020). Machine learning methods enhance environmental adaptability by learning and weighting informative features during training, yet their pixel-based frameworks struggle to resolve sub-pixel spectral mixing in small lakes and lack the capacity to model spatial semantic relationships (Jain et al., 2015; Dirscherl et al., 2020; Nazakat et al., 2021). Convolutional neural networks (CNNs), currently the most widely applied and effective method for identifying lakes in alpine periglacial environments, integrate spectral and spatial features but are constrained by the strong locality assumption of convolutional kernels, limiting their ability to capture global relationships between key glacial lake indicators and topographic factors (Thati and Ari, 2022; Tang et al., 2024; Sharma and Prakash, 2025). Meanwhile, automated classification of glacial versus non-glacial lakes typically relies on proximity to glaciers (with a common threshold of 10 km), a method that overlooks lake-specific environmental traits and hydrological connectivity, resulting in substantial errors (Wang et al., 2013; Zhang et al., 2015). Shape-based methods incorporate additional morphological parameters but struggle with irregularly shaped mountain lakes, making accurate classification challenging (Feyisa et al., 2014; Jiao et al., 2012; Khandelwal et al., 2017). Spectral-based methods also face difficulties due to spectral variability caused by seasonal ice melting and the overlap of spectral signatures between ice-covered lakes and non-ice-covered lakes in specific bands, which hinders robust differentiation of their fundamental differences (Brinthan et al., 2023).

In recent years, Vision Transformer (ViT)-based methods have emerged as a promising alternative for remote sensing image analysis, dividing imagery into fixed-size patches and leveraging a Transformer architecture to capture global dependencies (Dosovitskiy et al., 2021). This architecture offers potential for integrating multi-band and multi-temporal data within a unified embedding space via self-attention mechanisms, which could enhance the discrimination of subtle spectral and spatial patterns among land features (Roy et al., 2023; Heidarianbaei et al., 2024). Recent applications in other geoscience domains underscore its adaptability and relevance for identifying lakes in alpine periglacial environments. For instance, Peng et al. (2023) applied a Transformer-based U-Net with a Local-Global Transformer encoder to glacier extraction in the Qilian Mountains, integrating Sentinel-1 SAR, Sentinel-2 multispectral data, and DEMs, achieving an overall accuracy of 0.972 by leveraging multi-source data synergy. Similarly, Zhu et al. (2023) utilized a Swin-Transformer-enhanced DeepLabv3+ for glacier and ice shelf front detection from SAR imagery, capturing dynamic calving events with a Mean Intersection over Union (mIoU) of 0.94, demonstrating ViT's strength in modeling long-range contextual dependencies. Nadachowski et al. (2024) employed a ViT architecture for glacial landform classification using DEMs across diverse terrains, attaining up to 97.5% accuracy in distinguishing subtle morphological features.

Additionally, Yan et al. (2023) developed a Transformer-based network to extract lakes from Sentinel-2 imagery in the Tibetan Plateau, reducing cloud shadow interference with an overall accuracy of 0.9954, highlighting ViT's capacity to mitigate spectral confusion. Hou et al. (2024) introduced Hydroformer, a Transformer-based temporal sequence model, for lake level reconstruction, using frequency-enhanced attention to capture temporal dependencies with an R^2 of 0.813 across varied lake sizes , evaluated on 50 lakes distributed globally. Chen et al. (2024) proposed LEFormer, a hybrid CNN-Transformer model, achieving a mIoU of 97.42% on datasets covering surface water bodies globally and lakes in the Qinghai–Tibet Plateau. These studies collectively illustrate ViT's ability to integrate multimodal data, and model complex spatial contexts, yet their application has not been extended to the detection and classification of size-heterogeneous lakes in alpine periglacial environments under strong terrain-induced spectral and spatial complexity, motivating further investigation of ViT-based methods in this domain.

This study proposes an intelligent identification framework for lakes in alpine periglacial environments based on ViT-based models and multi-dimensional remote sensing features. The framework is designed to systematically evaluate feature representation advantages of ViT models over CNNs in complex environments. The results are expected to provide methodological insights for the precise identification and classification of diverse lake types, thereby supporting more consistent mapping and characterization of periglacial landforms and related cryospheric features. The framework is applied in the Central-Eastern Himalaya (CEH, spanning central Nepal to western Bhutan) for model training and the Southeastern Tibetan Plateau Gorge region (STPG, referring to the Yarlung Zangbo River gorge and its surrounding areas in southeastern Tibet) for independent testing, to assess its effectiveness and generalization capability. In addition, the framework enables a more comprehensive survey of lakes in alpine periglacial environments within the STPG.

2. Materials and Methods

2.1 Overview

The framework proposed in this study for identifying lakes in alpine periglacial environments, as depicted in Figure 1, consists of four key steps: data preprocessing, model training, prediction and validation, and postprocessing. At its core, this framework employs a two-stage strategy—segmentation followed by classification—to detect lakes in alpine periglacial environments, diverging from traditional semantic segmentation that simultaneously conducts segmentation and classification. This shift is driven by the challenges posed by incomplete lake representations in imagery due to cropping, coupled with the high similarity among lake bodies and the often fragmented nature of environmental features. Such conditions can compromise classification accuracy in conventional workflows, potentially resulting in different regions of the same lake being assigned distinct types. To circumvent these issues, the ViT-based identification framework first segments lake outlines, then extends a defined area around these contours for secondary image cropping, before performing type classification. This ensures that the classification imagery encompasses both the complete lake body and its environmental context, thereby improving classification accuracy and consistency. Experiments were conducted using Python 3.11 and PyTorch 2.1.2 (Paszke et al., 2019) on an NVIDIA 4060TI GPU (16 GB RAM, CUDA 12.3, cuDNN 8.9.7) and an AMD Ryzen 5 7500F CPU (6 cores, 12 threads, 3.7 GHz).

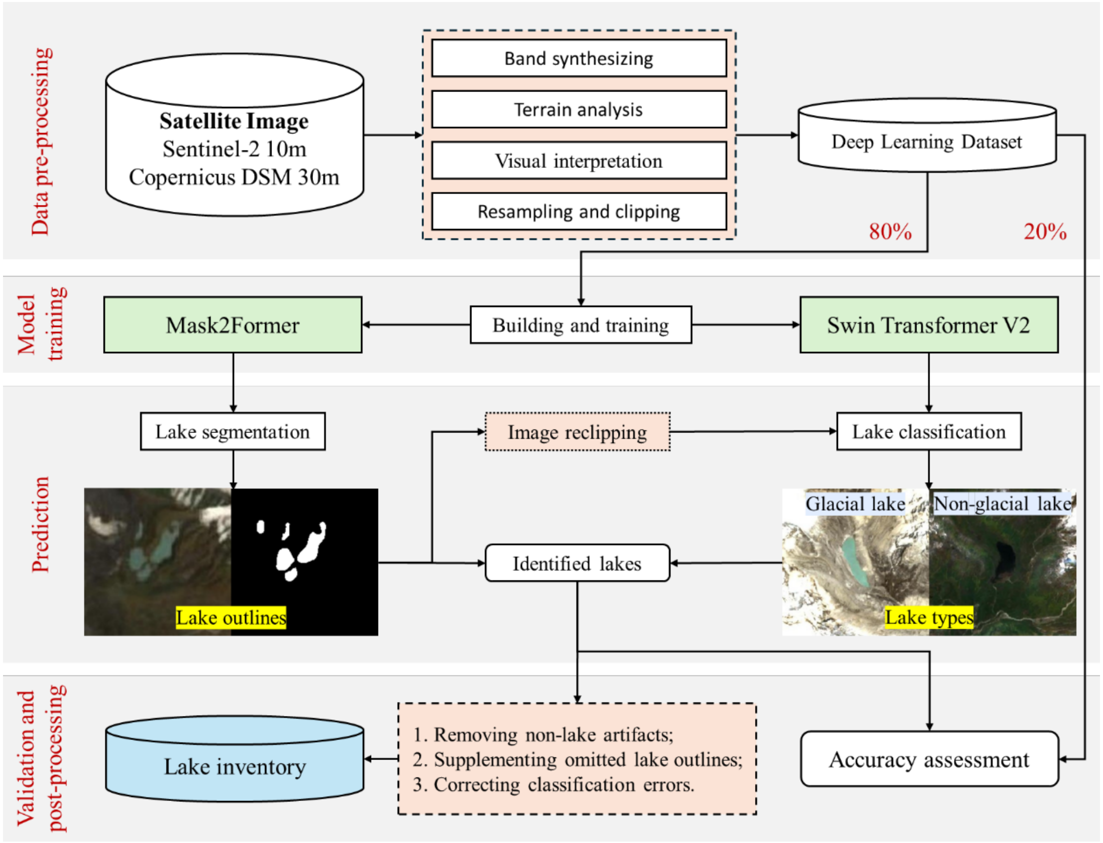


Figure 1. Flowchart of the ViT-based identification framework.

2.2 Study Sites

This study targets the CEH (Figure 2a) and the STPG (Figure 2b), both situated along the southern margin of the Tibetan Plateau. These regions host a high density of lakes in alpine periglacial environments and rank among the most dynamic zones of glacial lake evolution globally (Bajracharya et al., 2007; Ahmed et al., 2021; Furian et al., 2022). With elevations typically exceeding 4000 m asl and extensive glacier coverage, they experience concentrated summer precipitation driven by the South Asian monsoon (Wang and French, 1995; Zheng et al., 2000). Amid global warming and glacier retreat, rapid glacial lake expansion has heightened the risk of glacier lake outburst floods (GLOFs) (Bajracharya and Mool, 2009; Ahmed et al., 2021). The CEH, centered on the Himalayan main ridge, features a stepped topographic gradient, with annual precipitation of 1,500–2,500 mm (June–September) and dense populations, posing risks to downstream communities from lake outbursts (Karki et al., 2017; Xiang et al., 2024). In contrast, the STPG lies at the tectonic junction of the Himalayas, Hengduan Mountains, and Nyainqentanglha Range, characterized by intense tectonic activity, steep, fragmented terrain, and deep V-shaped valleys (Wang et al., 2014; Yu et al., 2020). It receives 2,500–4,000 mm of annual rainfall (May–October), influenced by Indian Ocean moisture and the Yarlung Zangbo vapor channel, with sparse human activity yet heightened flood potential due to extreme topographic relief (Sun and Su, 2020; Chen et al., 2024b). Deep learning samples from the CEH will be used to train models, leveraging the region’s moderate topographic variability and diverse lake characteristics to ensure comprehensive feature learning. The STPG, with its extreme terrain, higher precipitation, and complex environmental conditions, will serve as the test region to evaluate the model’s performance and generalization capability. This selection enables models to address varied topographic and climatic challenges, ensuring applicability across diverse periglacial landscapes.

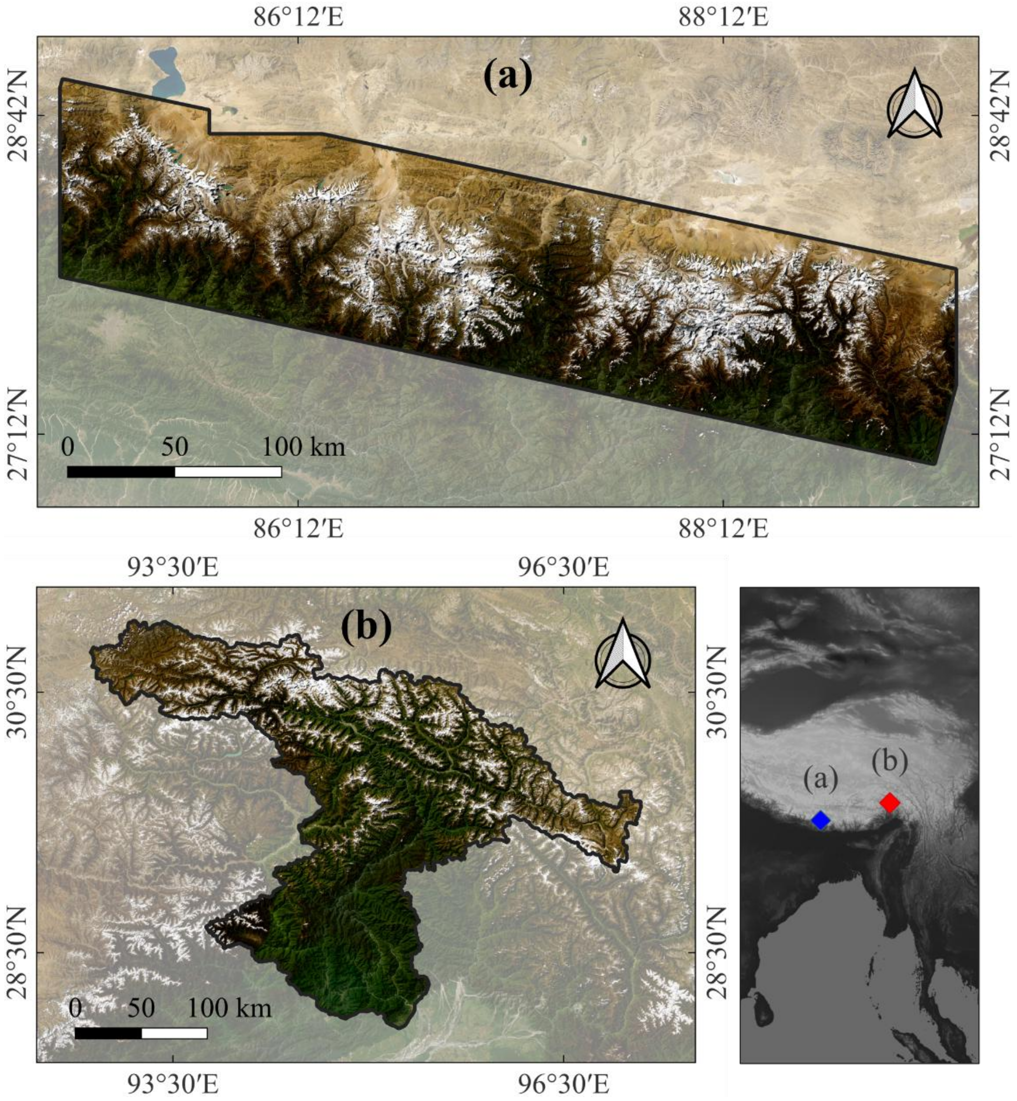


Figure 2. The location of the study sites: (a) The CEH; (b) The STPG. Base map sourced from ESRI ArcGIS World Imagery.

2.3 Data Sources and Data Pre-processing

This study utilizes 10m resolution Sentinel-2 Level-2A imagery and 30m resolution Copernicus Digital Surface Model (DSM) as data sources. The Sentinel-2 imagery was accessed via Google Earth Engine (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED, last accessed: 13 April 2025), while the Copernicus DSM was obtained from OpenTopography (https://portal.opentopography.org/raster?opentopoID=OTSDEM.032021.4326.3, last accessed: 13 April 2025).

The Sentinel-2 imagery was limited to composite observations acquired between June and October 2020, corresponding to the ablation and summer–early autumn season. During this period, lakes in alpine periglacial environments are typically ice-free, reach their maximum water extent, and exhibit the strongest spectral contrast with the surrounding terrain, thereby enabling more reliable lake identification and annotation. Seasonal composites were generated using a median compositing strategy, which suppresses transient features such as clouds and short-lived snow cover while retaining persistent surface water signals. During the ablation season, water pixels associated with lakes are consistently present across multiple acquisitions, causing the median composite to preferentially represent stable, near-maximum lake extents, although it does not strictly guarantee the absolute maximum extent for every individual lake.The utilized bands—B2 (Blue), B3 (Green), B4 (Red), B8 (Near-Infrared), and B11 (Shortwave Infrared)—generate RGB imagery, Normalized Difference Water Index (NDWI) (McFeeters, 1996), Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), and Normalized Difference Snow Index (NDSI) (Hall et al., 1995). RGB imagery provides information of lake color, shape, and location cues; NDWI enhances water body contrast for lake differentiation; NDVI reflects vegetation to prevent misidentification; and NDSI highlights snow and ice to avoid confusion while informing the glacial context. Slope and Topographic Wetness Index (TWI) (Beven and Kirkby, 1979) were derived from the DSM using the Geospatial Data Abstraction Library (GDAL, https://github.com/OSGeo/gdal, last accessed: 13 April 2025) and GRASS GIS (Version 8.4.1). Slope reflects the flatness of lake areas (near-zero for lakes), while TWI indicates potential wet areas and hydrological flow paths, aiding in assessing glacier-related lake replenishment. All data were resampled to 5-m resolution using GDAL tools and the Lanczos resample method (Lanczos, 1950), then reprojected to EPSG:3857 (WGS 84 / Pseudo-Mercator). This upsampling does not introduce new information but was applied to harmonize multi-source inputs and to facilitate smoother boundary representation for small, highly pixelated lakes during segmentation; the resulting 256 × 256 tiles still cover ~1.64 km^2, providing sufficient local spatial context.

Training labels were generated via visual interpretation in the CEH using RGB imagery supplemented by NDWI, adhering to the glacial lake classification system of Yao et al. (2018). Labels comprise lake outlines (0 for background, 1 for lakes) and types (0 for glacial lakes, 1 for non-glacial lakes), interpreted by two researchers experienced in glacial lake studies, yielding 5,693 labels (3,995 glacial lakes, 1,698 non-glacial lakes). Labels were interpreted by one researcher and independently checked by a second experienced researcher to ensure consistency. Potential discrepancies in lake boundaries and lake-type assignments were identified through side-by-side inspection and resolved by consensus, with reference to high-resolution imagery. Data were standardized using Z-score normalization, ensuring numerical stability and consistent feature scaling across input variables. Segmentation samples, used for lake outline detection, incorporated RGB, NDWI, NDVI, NDSI, and slope data. These were systematically cropped into 256×256 pixel tiles in a regular grid pattern to ensure comprehensive coverage, yielding 4,056 positive samples (containing lakes) and 6,045 negative samples (lacking lakes, randomly selected) to maintain data representativeness. Classification samples, designed to distinguish glacial from non-glacial lakes, included RGB, TWI, and lake outlines. For each lake outline, the boundary was extended outward by 1 km to form a region encompassing the lake and its surrounding environmental context. These regions were then cropped and resized to 256×256 pixels, balancing contextual inclusion with detail retention. The classification dataset comprised 3,995 positive samples (glacial lakes) and 1,698 negative samples (non-glacial lakes), with sample counts aligned with their respective labels.

2.4 Model Architectures and Training Parameters

Lake outline segmentation utilizes Mask2Former (Cheng et al., 2022), an advanced ViT-based model tailored for semantic segmentation, with its architecture illustrated in Figure 3a. Mask2Former enhances multi-scale feature extraction through optimized mask generation and feature interaction strategies. Its architecture consists of a backbone network for extracting multi-scale features, a Transformer decoder that refines feature maps using self-attention and cross-attention mechanisms, and a mask prediction head that reformulates segmentation as a mask classification task. By jointly modeling object-level masks and their global relationships, rather than relying solely on local pixel-wise decisions, this design reduces sensitivity to local noise and enables coherent delineation of lakes with fragmented or irregular boundaries. This design supports robust processing of high-resolution imagery and enables effective capture of detailed spatial patterns across diverse conditions.

Lake type classification employs Swin Transformer v2 (Liu et al., 2022), an advanced hierarchical ViT-based model optimized for image classification, with its architecture illustrated in Figure 3b. Swin

Transformer v2 improves feature representation with long-spaced continuous position bias and efficient computational operations. Its architecture includes a hierarchical backbone for generating multi-scale feature maps, a shifted-window self-attention mechanism for integrating local and global contextual information, and a classification head for streamlined label prediction. By aggregating information across multiple spatial scales and progressively expanding the receptive field through hierarchical attention, the model captures both lake-internal characteristics and surrounding contextual cues, which is essential for distinguishing glacial from non-glacial lakes. This structure facilitates the efficient analysis of high-resolution remote sensing imagery, with the potential to discern intricate spatial and contextual relationships.

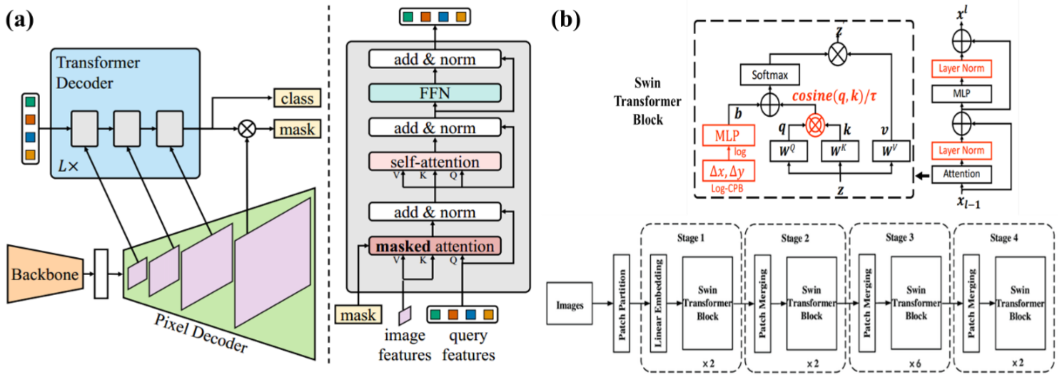


Figure 3. (a) Architecture of the Mask2Former model (Cheng et al., 2022); (b) Architecture of the Swin Transformer v2 (Liu et al., 2022).

All models were trained using a unified training strategy to ensure fair comparison. No iterative optimization or model adjustment based on test results was performed. Training was conducted with an 8:2 training-validation split, a batch size of 16, CrossEntropyLoss (Bridle, 1989), the AdamW optimizer (Loshchilov and Hutter, 2019), and a cosine schedule with warmup. Samples were shuffled before each epoch to enhance generalization. These hyperparameter settings were selected to balance training stability, convergence efficiency, and fair comparison across models, following configurations commonly used in recent remote sensing segmentation and classification studies. For comparison with CNN-based approaches, representative CNN architectures commonly used for dense prediction and image classification were selected: UNet (Ronneberger et al., 2015) and DeepLabv3+ (Chen et al., 2018) for lake outline segmentation, and ResNet (He et al., 2015) and EfficientNet (Tan and Le, 2020) for lake type classification, providing well-established and task-appropriate baselines. Details of model backbones and corresponding pre-trained weights are summarized in Table 1.

Table 1. Models evaluated in this study and their corresponding pre-trained weights.

Model	Pre-trained weights
Mask2Former (SwinTiny)	Hugging Face (facebook/mask2former-swin-tiny-ade-semantic , ADE20K)
UNet (ResNet50)	segmentation_models_pytorch (ImageNet-1K)
DeepLab v3+ (ResNet50)	segmentation_models_pytorch (ImageNet-1K)
Swin Transformer v2 (Tiny)	Hugging Face (microsoft/swinv2-tiny-patch4-window16-256, ImageNet-1K)
ResNet (50)	torchvision (ImageNet-1K)
EfficientNet (B0)	torchvision (ImageNet-1K)

2.5 Performance Assessment and Post-processing

To assess the performance of the segmentation and classification models, four metrics were employed. The mIoU was utilized to evaluate the segmentation model, while precision, accuracy, and F1-score were applied to the classification model. mIoU was computed in a pixel-based manner over the entire validation dataset, based on the total number of lake pixels, quantifying the overlap between predicted and ground-truth lake areas and explicitly assessing each model's ability to map lake areas as completely as possible. Higher mIoU values indicate more accurate and complete segmentation. Precision represents the fraction of true positive predictions among all samples classified as positive, reflecting the model's accuracy in identifying positive instances; a higher precision corresponds to fewer false positives. Recall, defined as the proportion of true positives correctly identified among all actual positive samples, measures the model's ability to detect positive instances, with higher values indicating fewer missed positives. The F1-score, computed as the harmonic mean of precision and recall, balances these two metrics and is particularly valuable when both accuracy and completeness are critical, with higher values denoting a more robust and balanced model performance. The mathematical formulations for these metrics are provided below:

$$MIoU = \frac{1}{c}\sum_{i=1}^{c}\frac{A_i \cap B_i}{A_i \cup B_i} \tag{1}$$

C represents the number of classes, A_i represents the actual segmented area for the ith class and B_i represents the predicted segmented area for the ith class.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

TP represents the number of samples correctly classified as positive, FP represents the number of samples incorrectly classified as positive , TN represents the number of samples correctly classified as negative, and FN represents the number of samples incorrectly classified as negative.

Following the performance assessment, the predicted lake outlines and types in the STPG were refined through a model-assisted post-processing workflow to generate the final lake inventory. This process integrated multi-band optical imagery, topographic information, and high-resolution Google Earth imagery as reference data. Non-lake artifacts identified by the models were removed through visual inspection, incomplete or fragmented lake outlines were manually completed, and misclassified lake types were corrected based on geomorphological context and surrounding environmental features. This post-processing step aimed to improve the consistency and completeness of the inventory while minimizing residual false positives.

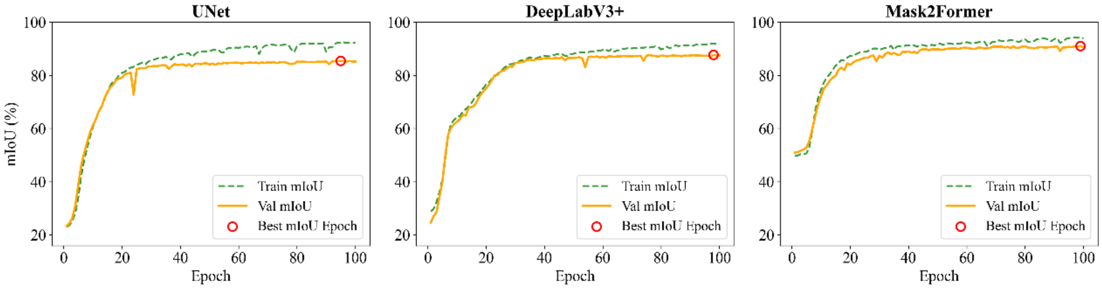
3. Results

The results presented below are based on a final lake inventory for alpine periglacial environments the STPG, which was generated by post-processing the lake outlines produced by Mask2Former and the lake-type predictions from Swin Transformer v2. Model outputs were refined through systematic correction to remove non-lake artifacts, complete fragmented lake boundaries, and adjust lake-type assignments. This finalized inventory serves as the common baseline for subsequent model performance evaluation and comparison with existing lake inventories.

Model performance is assessed from multiple perspectives, including overall segmentation and classification accuracy, as well as stratified analyses across lake size classes and elevation ranges. Lake area classes are used to evaluate model sensitivity to scale, while elevation ranges are employed as a proxy for varying environmental conditions in alpine regions. These stratified analyses provide a detailed assessment of model behavior under different geomorphological and environmental settings.

3.1 Comparative Analysis of ViT-based and CNN-based Models for Lake Segmentation

This study trained lake outline segmentation models over 100 epochs using samples from the Central–Eastern Himalaya (CEH) region. Pixel-level segmentation performance was evaluated on the CEH validation dataset. As shown in Figure 4, all three models exhibit rapid performance improvements during the early training stage and reach stable convergence after approximately 40–50 epochs, with only marginal gains thereafter. Among the tested models, Mask2Former achieved the highest validation performance and converged to a final mIoU of 91.01%. In contrast, the CNN-based models UNet and DeepLab v3+, trained with identical inputs and training settings, reached lower final mIoUs of 85.44% and 87.71%, respectively. The training and validation curves indicate stable convergence without evident overfitting for all models, while Mask2Former demonstrates a higher performance ceiling and faster convergence behavior.



Fugure 4. Training and validation performance curves of segmentation models on the CEH dataset.

Beyond pixel-level segmentation accuracy evaluated on the CEH dataset, the practical performance of the segmentation models was further examined at the lake-object level using the STPG as an application case. In this analysis, model outputs were compared against the finalized lake inventory to assess their detection behavior in a large-area mapping scenario. Under this setting, Mask2Former outperformed both CNN-based models in detection rate (Table 2) — defined as the ratio of detected lakes to the total number of lakes — achieving 93.17% compared to 87.78% for UNet and 88.89% for DeepLab v3+. It also generated substantially fewer non-lake artifacts (380, compared to 1,180 for UNet and 750 for DeepLab v3+).

Table 2. Lake segmentation performance of segmentation models.

Model	Total polygons	Detected lakes	Missed lakes	Non-lake artifacts
Mask2Former	3610	3043	223	380
UNet	4674	2867	399	1180
DeepLab V3+	4063	2903	363	750

To evaluate the segmentation performance across different lake sizes, detection rates of Mask2Former, UNet, and DeepLab v3+ were compared (Table 3). Results indicate that Mask2Former consistently outperforms the other models across all size categories, with particularly pronounced advantages for smaller lakes. For ultra-small lakes (<0.001 km²), Mask2Former achieved a detection rate of 73.42%, notably higher than UNet (61.32%) by 19.73% and DeepLab v3+ (63.42%) by 15.77%. For small lakes (0.001–0.01 km²), Mask2Former maintained a detection rate of 92.90%, surpassing UNet (86.07%) by 7.93% and DeepLab v3+ (87.52%) by 6.15%. For medium-sized lakes (0.01–0.1 km²), the performance gap narrowed, with Mask2Former achieving 98.80%, outperforming UNet (96.48%) by 2.40% and DeepLab v3+ (96.94%) by 1.92%. In large lakes (>0.1 km²), the detection rates of all three models converged, with Mask2Former and DeepLab v3+ both reaching 99.65%, while UNet lagged slightly at 99.30%, trailing Mask2Former by 0.35%.

Table 3. Number of lakes detected by segmentation models across area ranges.

Area Model	<0.001 km^2	0.001–0.01 km^2	0.01–0.1 km^2	>0.1 km^2
Final inventory	380	1522	1080	284
Mask2Former	279	1414	1067	283
UNet	233	1310	1042	282

DeepLab V3+	241	1332	1047	283

To evaluate segmentation performance across elevation gradients, detection rates of Mask2Former, UNet, and DeepLab v3+ were compared (Table 4). Results show that Mask2Former consistently outperforms the other models across all elevation ranges, with a notable advantage at lower elevations and extreme elevations. For low-elevation lakes (<4,000 m), Mask2Former achieved a detection rate of 93.65%, higher than UNet (88.38%) by 5.98% and DeepLab v3+ (90.56%) by 3.41%. For mid-elevation lakes (4,000–4,500 m), Mask2Former maintained a detection rate of 92.60%, surpassing UNet (87.92%) by 5.32% and DeepLab v3+ (88.67%) by 4.43%. For high elevation lakes (4,500–5,000 m), Mask2Former's rate reached 93.56%, outperforming UNet (89.46%) by 4.58% and DeepLab v3+ (90.15%) by 3.78%, though the gap narrowed. For extreme elevation lakes (>5,000 m), Mask2Former sustained the highest rate at 92.90%, exceeding UNet (85.70%) by 8.40% and DeepLab v3+ (86.87%) by 6.94%.

Table 4. Number of lakes detected by segmentation models across elevation ranges.

Elevation Model	<4000 m	4000–4500 m	4500–5000 m	>5000 m
Final inventory	551	662	1025	1028
Mask2Former	516	613	959	955
UNet	487	582	917	881
DeepLab V3+	499	587	924	893

3.2 Comparative Analysis of ViT-based and CNN-based Models for Lake Classification

Lake type classification performance was evaluated on the CEH dataset using precision, recall, and F1-score as metrics (Figure 5). All classification models were trained for 20 epochs, and the training curves indicate that performance improves rapidly during the initial stage and reaches stable convergence within the first 4–5 epochs, with only minor fluctuations thereafter. No pronounced overfitting is observed across the training process. Among the tested models, Swin Transformer v2 achieved the highest overall performance, with an F1-score of 89.75%, outperforming EfficientNet (82.43%) and ResNet (82.33%).

Swin Transformer v2 also exhibited the most balanced classification behavior, achieving a precision of 86.94% and a recall of 92.74%. In comparison, ResNet attained a lower precision (78.46%) despite a relatively high recall (88.62%). EfficientNet showed a different error pattern, with a moderate precision of 80.11% but a lower recall of 84.89%.

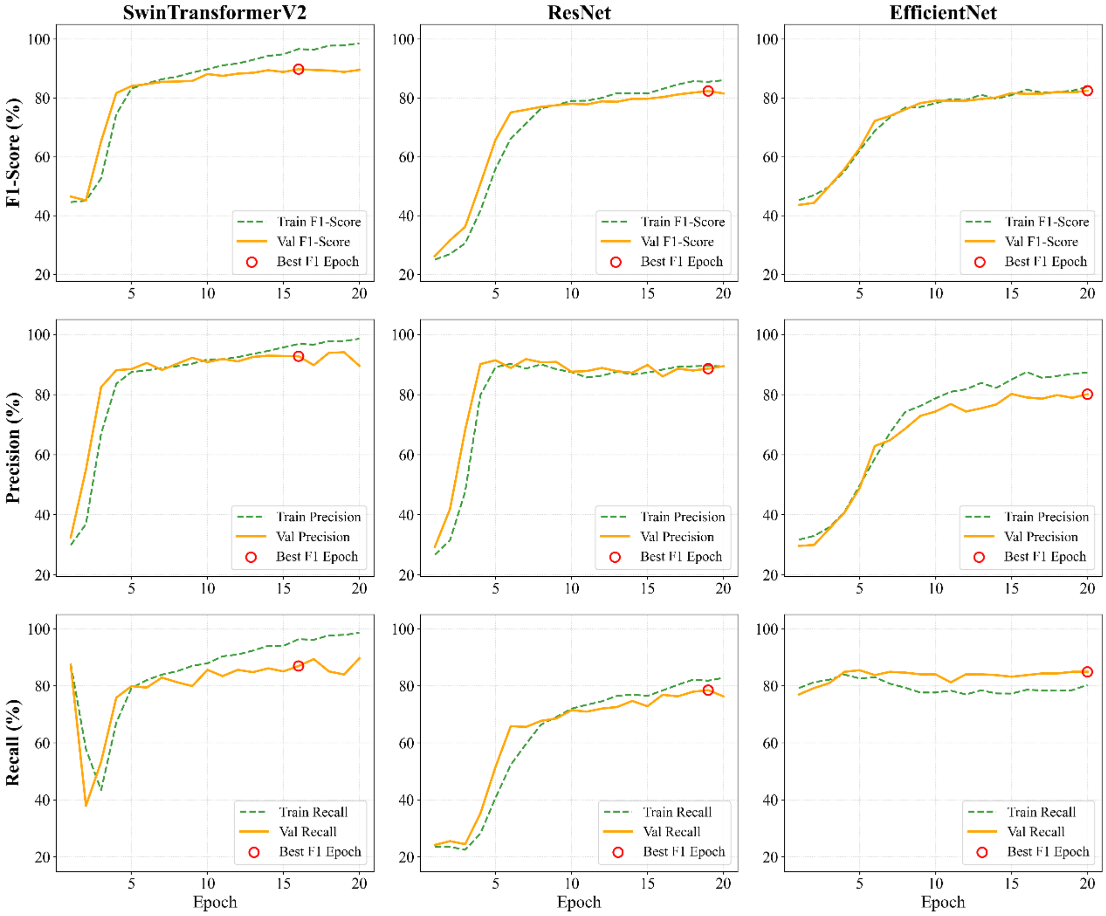


Figure 5. Training and validation performance curves of classification models on the CEH dataset.

Based on the finalized lake inventory for the STPG, the classification performance of the three models was quantified using confusion matrix statistics and F1-score (Table 5). Swin Transformer v2 achieved the highest F1-score (91.17%), indicating the best balance between omission and commission errors at the regional scale. In comparison, ResNet and EfficientNet yielded lower F1-scores of 88.39% and 88.63%, respectively. The confusion matrix reveals distinct error characteristics among the models. ResNet produced a higher number of false positives (FP = 297). EfficientNet, in contrast, shows an increased number of false negatives (FN = 149).

Table 5. Confusion matrix results for lake classification using classification models.

Model	TP	TN	FP	FN	F1-Score
Swin Transformer v2	1620	1332	226	88	91.17%
ResNet	1588	1261	297	120	88.39%
EfficientNet	1559	1307	251	149	88.63%

Classification performance was further analyzed across four lake area ranges using F1-score and confusion matrix statistics (Table 6). Swin Transformer v2 consistently achieved the highest F1-scores across all size classes, indicating robust classification performance over a wide range of lake scales. The advantage of Swin Transformer v2 was most pronounced for ultra-small (<0.001 km²) and small lakes (0.001–0.01 km²), where F1-scores exceeded those of ResNet and EfficientNet by approximately 6–8%. For medium-sized (0.01–0.1 km²) and large lakes (>0.1 km²), performance differences among models decreased, although Swin Transformer v2 still maintained the highest F1-scores.

Table 6. Confusion matrix results across area ranges for classification models.

Area	Model	TP	TN	FP	FN	F1-Sorce
<0.001 km² (n=380)	Swin Transformer v2	220	122	23	15	92.05%
	ResNet	202	95	50	33	82.96%
	EfficientNet	205	96	49	30	83.84%
0.001–0.01 km² (n=1522)	Swin Transformer v2	805	531	113	73	89.64%
	ResNet	761	437	207	117	82.45%
	EfficientNet	746	455	189	132	82.29%
0.01–0.1 km² (n=1080)	Swin Transformer v2	446	522	83	29	88.84%
	ResNet	413	481	124	62	81.62%
	EfficientNet	398	504	101	77	81.72%

>0.1 km² (n=284)	Swin Transformer v2	113	145	19	7	89.68%
	ResNet	103	139	25	17	83.06%
	EfficientNet	101	143	21	19	83.47%

Classification performance was further examined across four elevation ranges using F1-score derived from the confusion matrix (Table 7). Swin Transformer v2 consistently achieved the highest F1-scores across all elevation bands. At low elevations (<4,000 m), Swin Transformer v2 obtained an F1-score of 88.18%, substantially higher than those of ResNet and EfficientNet. Similar performance advantages were observed at mid elevations (4,000–4,500 m), where Swin Transformer v2 reached an F1-score of 89.17%. At higher elevations (4,500–5,000 m), the performance gap among models narrowed, although Swin Transformer v2 still maintained the highest F1-score (87.36%). At extreme elevations (>5,000 m), all three models achieved relatively high classification performance, with Swin Transformer v2 again yielding the best result (92.02%).

Table 7. Confusion matrix results across elevation ranges for classification models.

Elevation	Model	TP	TN	FP	FN	F1-Score
<4000 m (n=551)	Swin Transformer v2	138	376	21	16	88.18%
	ResNet	96	349	48	58	64.43%
	EfficientNet	84	362	35	70	61.54%
4000–4500 m (n=662)	Swin Transformer v2	210	401	29	22	89.17%
	ResNet	190	355	75	42	76.46%
	EfficientNet	181	365	65	51	75.73%
4500–5000 m (n=1025)	Swin Transformer v2	515	361	116	33	87.36%
	ResNet	465	323	154	83	79.69%
	EfficientNet	473	331	146	75	81.06%
>5000 m	Swin Transformer v2	721	182	72	53	92.02%

(n=1028)	ResNet	728	125	129	46	89.27%
	EfficientNet	712	140	114	62	89.00%

3.3 Lakes in Alpine Periglacial Environments in the STPG

The ViT-based identification framework identified 3,266 lakes in alpine periglacial environments in the STPG, comprising 1,708 glacial lakes and 1,558 non-glacial lakes (Figure 6). Their spatial distribution exhibits significant variability. Glacial lakes are predominantly aligned with the glacier systems of the Nyainqêntanglha Range, Himalayas, and Hengduan Mountains, extending from northwest to southeast, reflecting the primary role of glacial activity in their formation. Non-glacial lakes, conversely, are concentrated in non-glaciated regions, primarily in the northwest, central-north, and southern sectors of the study area, indicating distinct geomorphological controls on their distribution.

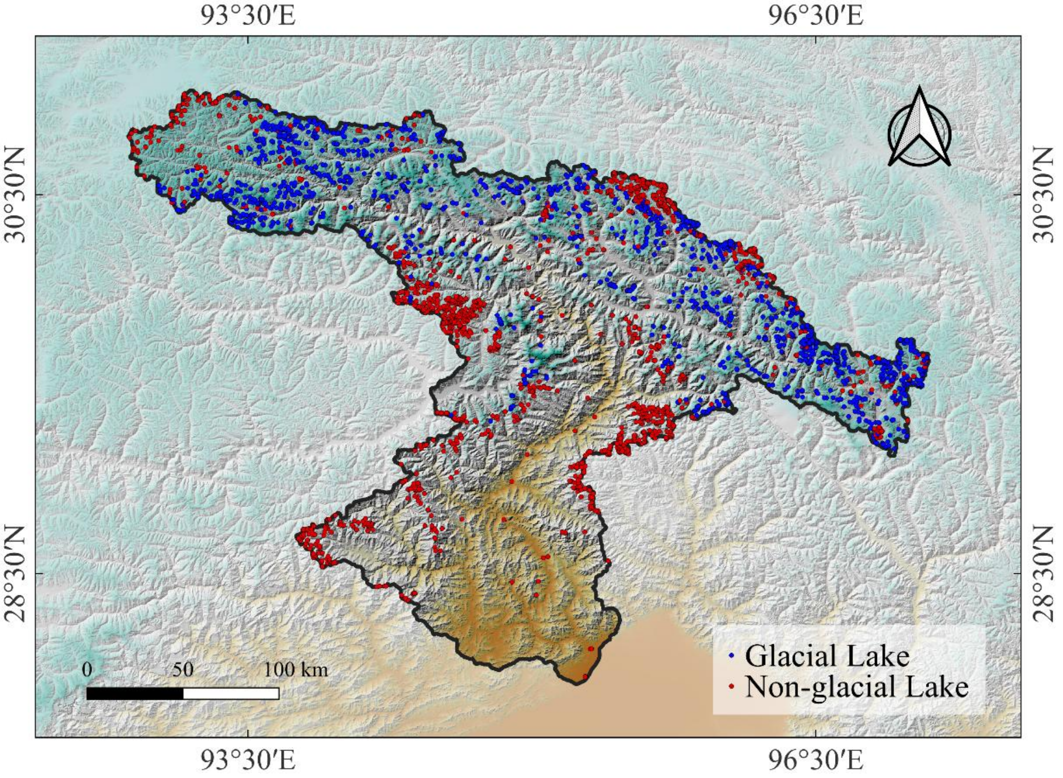
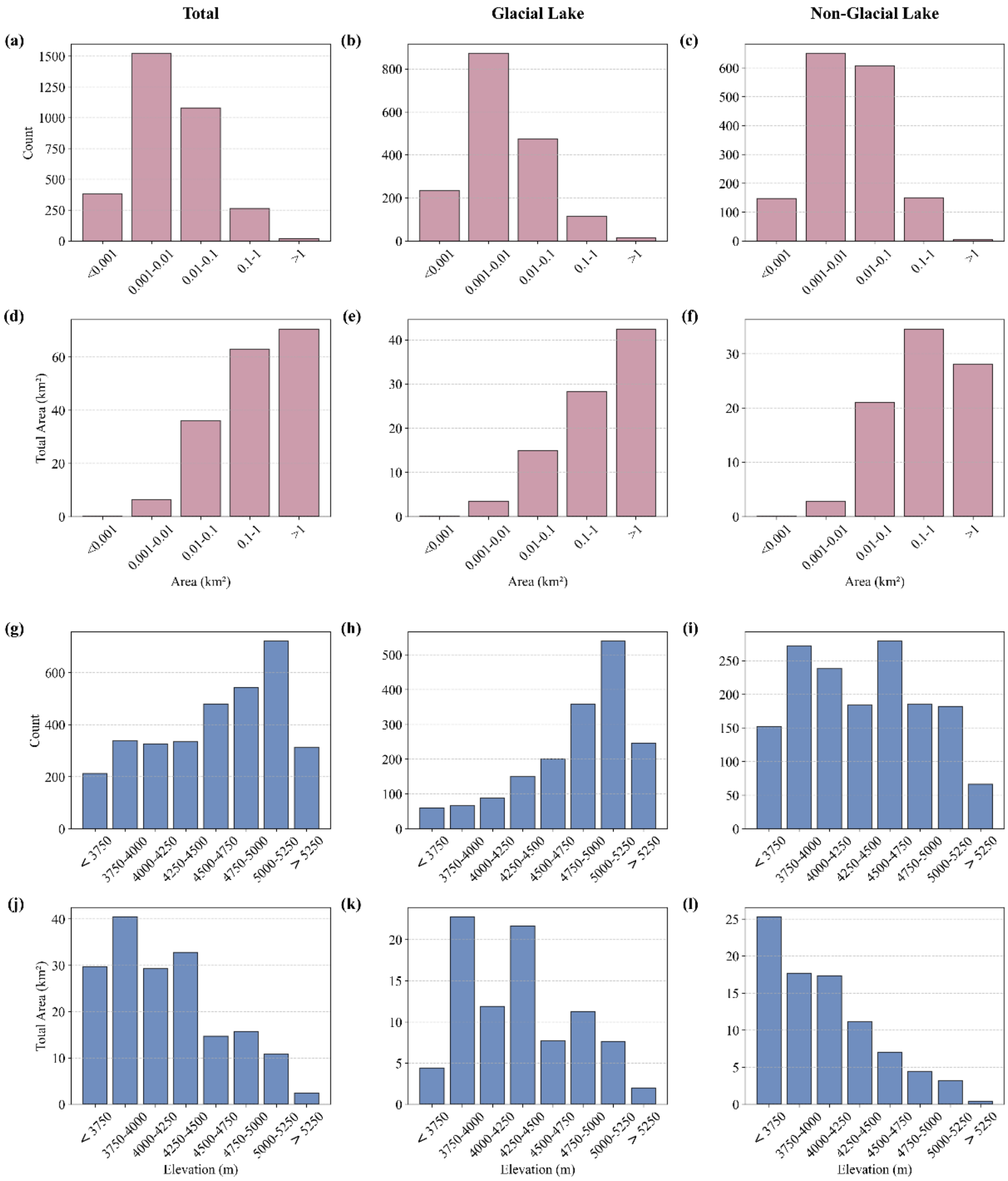


Figure 6. Lakes in Alpine Periglacial Environments in the STPG (blue dots represent glacial lakes, and red dots represent non-glacial lakes).

The identified lakes collectively occupy a total area of 175.6 km², with a mean area of 0.054 km² per lake. Glacial lakes account for 89.3 km² and non-glacial lakes for 86.3 km², with mean areas of 0.052 km² and 0.055 km² . Lake size distributions are strongly skewed toward small lakes (Figure 7a–c). In terms of lake counts, approximately 79.6 % of all lakes fall within the 0.001–0.1 km² range. Lakes smaller than 0.001 km² are relatively uncommon, whereas lakes larger than 0.1 km² constitute only a very small fraction of total lake numbers. In contrast, total lake area is dominated by a small number of larger lakes (>0.1 km²), illustrating a clear decoupling between lake abundance and area contribution. This size–frequency structure is broadly consistent for both glacial and non-glacial lakes, indicating that small lakes dominate numerically regardless of lake type, while large lakes disproportionately control total lake area (Figure 7d–f). Elevationally, lakes are distributed across a wide altitude range with a mean elevation of approximately 4,600 m (Figure 7g–i). Glacial lakes occur at systematically higher elevations, with a mean elevation of 4,822 m, whereas non-glacial lakes have a lower mean elevation of 4,356 m. Lake abundance generally increases with elevation for glacial lakes and peaks in the 5000–5250 m elevation band, whereas non-glacial lakes exhibit a comparatively more even distribution across a broad elevation range, without a pronounced peak. In terms of total lake area, lower-elevation bands contribute disproportionately to the overall area, whereas high-elevation zones, despite hosting numerous lakes, account for a comparatively smaller share of total lake area (Figure 7j–l).



420 **Figure 7.** Size- and elevation-classified distributions of lakes in alpine periglacial environments of the STPG. Panels (a–c) show lake counts by size class for all lakes, glacial lakes, and non-glacial lakes, respectively, while panels (d–f) show the corresponding total lake area. Panels (g–i) present lake counts by elevation band, and panels (j–l) show the corresponding total lake area for all lakes, glacial lakes, and non-glacial lakes.

4. Discussion

4.1 Performance Advantages of ViT-based Models over CNN-based models

Experimental results demonstrate that ViT-based models significantly outperform CNN-based models in the segmentation and classification of lakes in alpine periglacial environments. The ViT-based model excels at preserving lake boundary integrity during detection, whereas CNN-based models frequently exhibit boundary loss when processing mixed pixels, particularly in shallow zones near lake margins (as shown in Figure 8a). This disparity arises from ViT's self-attention mechanism, which constructs feature representations from a global perspective, capturing long-range pixel dependencies to delineate continuous lake boundaries accurately. In contrast, CNN-based models, constrained by localized convolutional kernels, struggle to adapt to the diverse boundary morphologies prevalent in complex terrains. To enhance information complementarity, the experiment incorporated multi-feature representations derived from Sentinel-2 imagery. For instance, when lake morphology is indistinct in RGB imagery, NDWI provides clearer boundary cues; when NDWI confounds lakes with shadowed slopes, slope data facilitates differentiation. Consequently, common non-lake artifacts caused by shadows and glacial snow are substantially suppressed in the results of both ViT-based model and CNN-based models. However, in river channel scenarios, CNN-based models still face significant challenges, often misclassifying fragmented rivers as lakes due to an over-reliance on local texture features (as shown in Figure 8b). In contrast, the ViT-based model leverages semantic scene understanding to effectively distinguish lake from non-lake features, markedly reducing artifact interference.

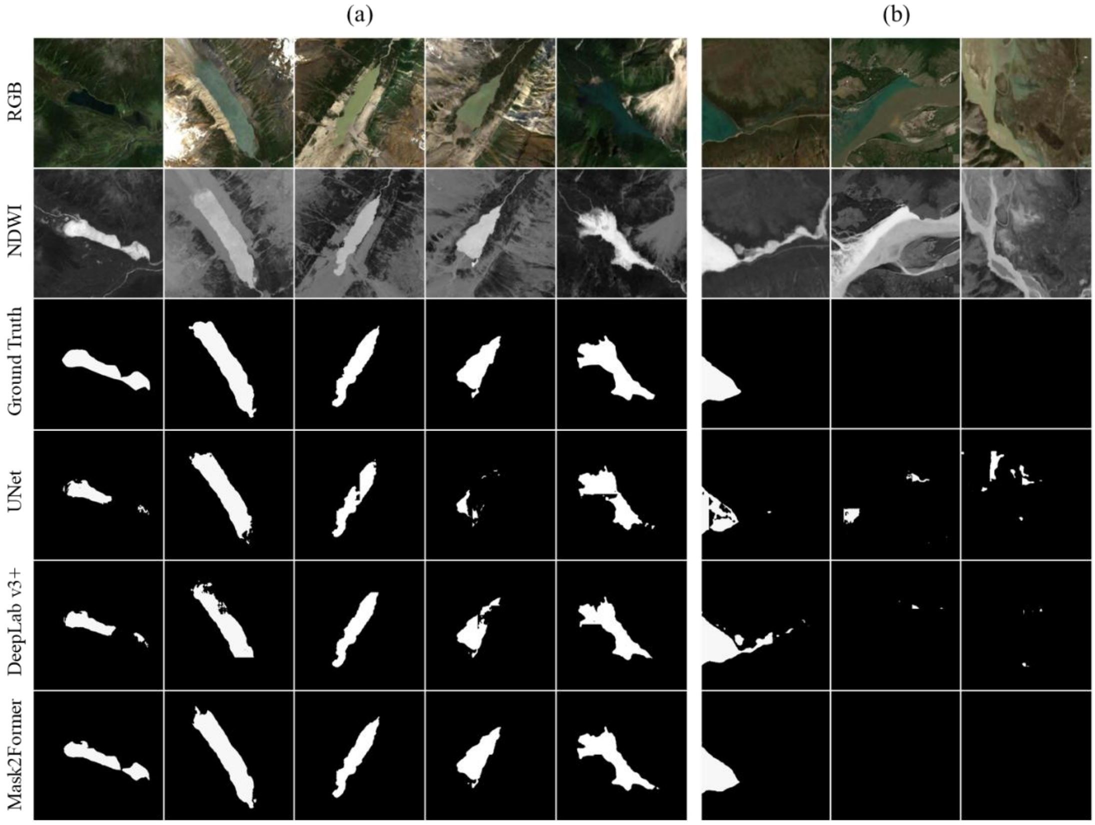


Figure 8. Comparison of the ViT-based model and CNN-based models performance in identifying lakes in alpine periglacial environments**:** (a) boundary detection integrity in mixed pixel scenarios and (b) artifact suppression in river channel scenarios.

In lake size categories, the ViT-based model exhibits pronounced superiority, particularly for ultra-small lakes (<0.001 km²), which often span just a few pixels in remote sensing imagery and are easily confused with surrounding vegetation or bare land. CNN-based models, constrained by fixed kernel sizes, struggle to resolve subpixel spectral mixing, resulting in misclassification or omission. The ViT-based model, through adaptive multiscale feature extraction, enhances detection rates and accuracy for these subtle targets. However, as lake size increases (>0.1 km²), the performance gap narrows, as larger spatial extents provide sufficient context for CNN-based models to achieve comparable segmentation and classification.

Across elevation ranges, the ViT-based model exhibits exceptional robustness, with particularly pronounced advantages over CNN-based models at low (<4,000 m) and extreme (>5,000 m) elevations. At lower elevations, dense vegetation cover often reduces lake visibility, while at extreme elevations, interference from snow and ice further obscures lake identification. CNN-based models, constrained by fixed local receptive fields and sensitivity to textural variability, struggle to effectively extract lake features in such complex environments, leading to frequent false positives. In contrast, the ViT-based model employs attention-driven feature selection to distinguish targets from backgrounds, enhancing segmentation and classification accuracy while improving generalization capacity. By modeling morphological continuity and spatial structure, the ViT-based model demonstrates robust lake identification under highly heterogeneous conditions, significantly reducing uncertainty.

Although ViT-based models generally involve higher computational complexity than CNN-based architectures, the evaluated models do not exhibit order-of-magnitude differences in computational cost, and total processing times remain within approximately 10 hours under the experimental setup. All experiments were conducted in an offline mapping framework without real-time constraints. For inventory-scale mapping tasks such as this study, model selection should prioritize segmentation accuracy and robustness over computational efficiency.

4.2 ViT-Enhanced Lake Inventory Completeness and Classification Accuracy

Compared to the previous lake inventories for the STPG from 2020 (Zhang et al., 2024a, b), this study demonstrates notable improvements in both the number of identified lakes and the overall completeness of the inventory (Figure 8). Zhang et al. (2024a), using a combination of water body indices and visual interpretation, identified 569 glacial lakes, without including non-glacial lakes. Zhang et al. (2024b), based on visual interpretation, documented 610 glacial lakes and 427 non-glacial lakes, totaling 1,037 lakes. In contrast, this study mapped 1,708 glacial lakes and 1,558 non-glacial lakes, yielding a total of 3,266 lakes, approximately three times more than that of Zhang et al. (2024b) and six times more than that of Zhang et al. (2024a). To further contextualize these improvements, Table 8 compares the proposed ViT-based classification with commonly used distance-based criteria, including the 10 km glacier proximity threshold adopted in previous inventories (Zhang et al., 2024a, b). The distance-based approach exhibits pronounced trade-offs between omission and commission errors depending on the selected threshold. While the 10 km criterion minimizes false positives, it results in a large number of missed lakes, whereas the 1 km criterion improves recall in near-glacier environments but introduces additional false positives. In contrast, the ViT-based classification achieves a more balanced performance, yielding the highest F1-score and demonstrating greater robustness across regions where glacial and non-glacial lakes spatially co-occur.

Table 8. Comparison of published lake inventories and distance-threshold-based and ViT-based classification approaches .

Method	TP	TN	FP	FN	F1-Score
Distance-based, 10 km (Zhang et al. 2024a)	428	0	141	0	85.86%
Distance-based, 10 km (Zhang et al. 2024b)	575	345	35	82	90.77%
Distance-based, 1 km (This Study)	1575	1314	133	244	89.31%
Distance-based, 10 km (This Study)	1708	654	0	904	79.09%
ViT-based (This Study)	1620	1332	226	88	91.17%

This study employs a ViT-based intelligent lake identification framework, markedly enhancing the detection of small lakes and providing more precise boundary delineation. While the published inventories captured lakes larger than 0.001 km², the ViT-based identification framework achieves an order-of-magnitude improvement, detecting lakes as small as 0.0001 km². In contrast, traditional vision interpretation are not only time-intensive and less efficient but also prone to human-induced inconsistencies, often resulting in omitted lakes or inaccurate boundaries (Blaschke, 2010; Lillesand et al., 2015). For the inventory of Zhang et al. (2024a), lakes not included in the published dataset have an average area of 0.035 km², compared to 0.129 km² for the included lakes. Collectively, these not-included lakes account for approximately 53 % of the total lake area mapped in this study. Similarly, for Zhang et al. (2024b), lakes not included in the inventory have an average area of 0.029 km², compared to 0.098 km² for the included lakes, and represent approximately 37 % of the total lake area. These differences highlight the superior capability of ViT-based models for capturing smaller lakes that were previously overlooked. These small glacial lakes, which form prolifically during glacial ablation, play a critical role in glacial lake outburst flood (GLOF) risk assessments (Yao et al., 2014; Zhang et al., 2022b). Given their abundance and widespread distribution, their potential failure poses severe threats to downstream regions, emphasizing the importance of their accurate detection.

Based on the segmentation results, Mask2Former produced a limited number of non-lake artifacts prior to post-processing across the STPG. These artifacts were explicitly reviewed and removed during the inventory refinement step. Even under a conservative assumption that some residual false positives may persist, their potential contribution is insufficient to account for the several-fold increase in lake numbers relative to previous inventories. This indicates that the observed differences are primarily driven by improved detection of small lakes rather than systematic overestimation. Residual uncertainty may persist for a small number of very small lakes affected by spectral confusion.

Lake type classification in existing glacial lake inventories has often relied on simple proximity-based criteria, assuming that lakes located within a fixed distance from glacier termini are of glacial origin (Wang et al., 2020; Zhang et al., 2022a; Ma et al., 2025). To evaluate its performance, this study applied the 10 km buffer method using glacier outlines from the Second Chinese Glacier Inventory v1.0 (Guo et al., 2015) to classify lakes in alpine periglacial environments in the STPG, and compared the results with ViT-derived classifications. To illustrate the practical implications of this approach, a representative glaciated mountain sector in the eastern STPG was selected for detailed visualization (Figure 9). The analysis shows that this distance-based method significantly overestimates both the number and spatial extent of glacial lakes. In an area approximately 30 km east of Motuo, more than 30 non-glacial lakes were misclassified as glacial simply because they fell within the 10 km buffer, despite being stable water bodies without glacial meltwater input. This discrepancy underscores the limitations of proximity-based classification. In contrast, the ViT-based model integrates spectral, morphological, and environmental features through its self-attention mechanism, enabling more accurate differentiation between glacial and non-glacial lakes—even in complex periglacial settings. This improvement reduces uncertainty in glacial lake inventories and enhances the reliability of climate risk assessments, providing a stronger basis for targeted disaster mitigation strategies.

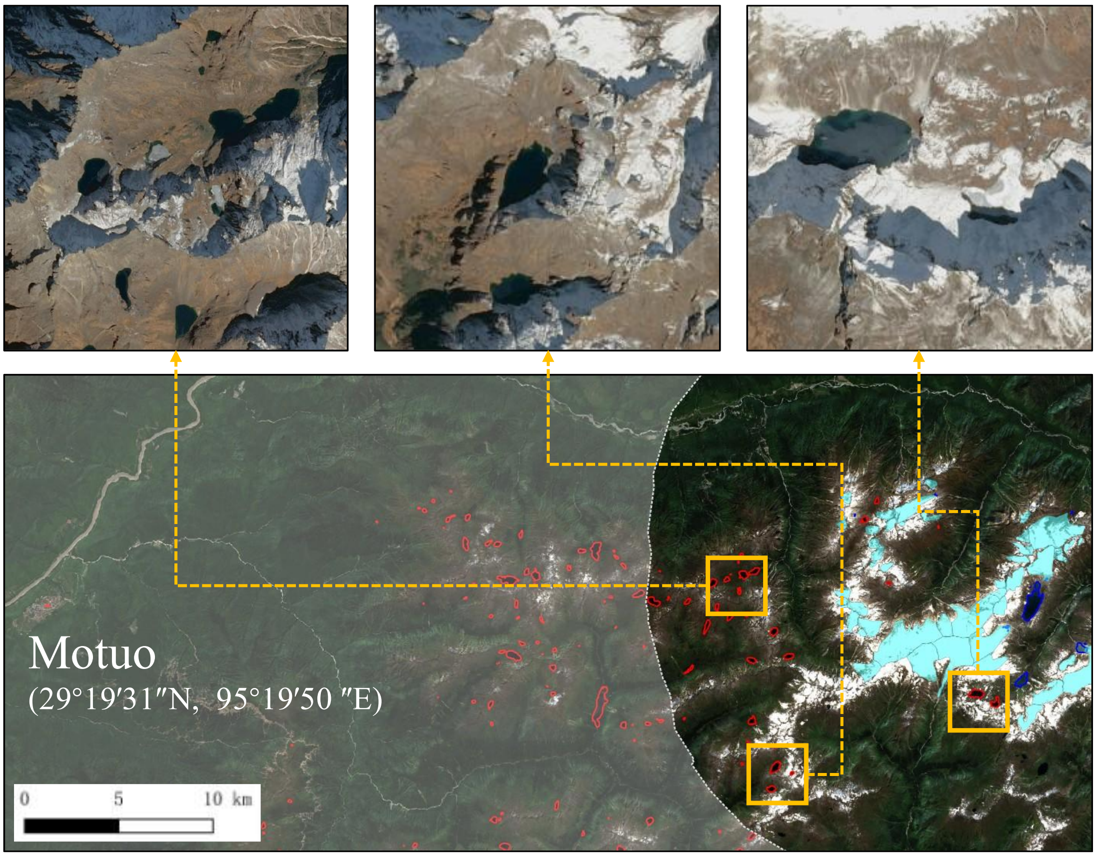


Figure 9. Distribution of lakes in alpine periglacial environments in an area located approximately 30 km east of Motuo County, Tibet, China (blue lines represent glacial lakes, red lines represent non-glacial lakes, Cyan polygons represent glacier extents, and the white translucent mask delineates areas located beyond 10 km from glaciers.). Base map sourced from ESRI ArcGIS World Imagery and Yandex Maps.

4.3 Limitations and perspectives

Despite the notable strengths of the ViT-based model in glacial lake identification, several limitations persist in this study. Glacial lakes situated far from glaciers—up to 15 km, as noted by Yao et al. (2018)—are occasionally misclassified due to the constrained spatial context resulting from cropped input images. While ViT's global attention mechanism partially mitigates this issue by capturing broader dependencies compared to conventional CNNs, the challenge of modeling long-distance spatial relationships suggests a need for multi-scale methods, such as hierarchical Transformer architectures, to enhance accuracy.

In addition, the analysis is based on single-season Sentinel-2 imagery acquired during the ablation and summer–early autumn period, which, although optimal for lake visibility and annotation reliability, limits the model's ability to explicitly account for seasonal variability. Summit snow cover introduces spectral confusion with glaciers in low-resolution imagery, leading to misclassifications of lakes. Incorporating time series data to account for seasonal snow variations could refine classification by providing temporal context, potentially reducing errors by distinguishing transient snow from permanent glacial features. Integrating multi-temporal imagery could address this by supplying historical context, enabling more precise identification of lakes with glacial origins.

5. Conclusion

This study proposed an intelligent framework for identifying lakes in alpine periglacial environments using ViT-based models. Compared to CNN-based models, ViT-based models demonstrated superior segmentation accuracy and classification robustness across diverse lake sizes and elevations. It effectively detected small lakes—often missed by CNN-based models—while minimizing false positives, such as mountain shadows and river fragments. The ViT-based model also distinguished glacial from non-glacial lakes with greater precision than the traditional glacier-proximity-based method, which is prone to overestimation.

When applied to the STPG, the framework produced an inventory of 3,266 lakes, comprising 1,708 glacial and 1,558 non-glacial lakes. This inventory exceeded the completeness of published datasets, highlighting the efficacy of ViT-based models in complex alpine terrains. The resulting dataset offers high-quality data to support the analysis of lake evolution and the assessment of climate-driven hydrological risks in glaciated regions.

Data availability. The training and validation datasets used in this study, including manually interpreted lake outlines and lake-type labels, are publicly available through the National Tibetan Plateau Data Center at https://doi.org/10.11888/Cryos.tpdc.303257.

Author contribution. MF and JX conceived and designed the study. JX developed the methodology, performed the analysis, and wrote the original draft. MF edited and finalized the manuscript. YJS, QW, and ZH contributed to data analysis. YNS, XZ, and RW curated the data. All authors reviewed and approved the final manuscript.

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

Financial support. This work was supported by the National Key Research and Development Program of China: [grant number 2023YFF0725005]; the TPESER Youth Innovation Key Program: [grant number TPESER-QNCX2022ZD-04]; the National Natural Science Foundation of China: [grant number 42301160]; the Science and Technology Department of Tibet Program: [grant number XZ202301ZY0035G].

References

Ahmed, R., Wani, G. F., Ahmad, S. T., Sahana, M., Singh, H., and Ahmed, P.: A review of glacial lake expansion and associated glacial lake outburst floods in the himalayan region, Earth Syst. Environ., 5, 695–708, https://doi.org/10.1007/s41748-021-00230-9, 2021.

Avtar, R., Komolafe, A. A., Kouser, A., Singh, D., Yunus, A. P., Dou, J., Kumar, P., Gupta, R. D., Johnson, B. A., and Minh, H. V. T.: Assessing sustainable development prospects through remote sensing: A review, Remote Sens. Appl. Soc. Environ., 20, 100402, https://doi.org/10.1016/j.rsase.2020.100402, 2020.

Bajracharya, S. R. and Mool, P.: Glaciers, glacial lakes and glacial lake outburst floods in the Mount Everest region, Nepal, Ann. Glaciol., 50, 81–86, https://doi.org/10.3189/172756410790595895, 2009.

Bajracharya, S. R., Mool, P. K., and Shrestha, B. R.: Impact of climate change on Himalayan glaciers and glacial lakes: case studies on GLOF and associated hazards in Nepal and Bhutan, International Centre for Integrated Mountain Development Kathmandu, 2007.

Barbieux, K., Charitsi, A., and Merminod, B.: Icy lakes extraction and water-ice classification using Landsat 8 OLI multispectral data, Int. J. Remote Sens., 39, 3646–3678, https://doi.org/10.1080/01431161.2018.1447165, 2018.

Basnett, S., Kulkarni, A. V., and Bolch, T.: The influence of debris cover and glacial lakes on the recession of glaciers in Sikkim Himalaya, India, J. Glaciol., 59, 1035–1046, https://doi.org/10.3189/2013JoG12J184, 2013.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. J., 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65, 2–16, https://doi.org/10.1016/j.isprsjprs.2009.06.004, 2010.

Bridle, J.: Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters, Adv. Neural Inf. Process. Syst., 2, https://doi.org/10.5555/2969830.2969856, 1989.

Brinthan, K., Thanujan, T., Thiruchittampalam, S., and Jayawardena, C. L.: Subclassification of Water Resources with Sentinel-2 Satellite Imagery: Spectra-Based Insight, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, 2442–2445, https://doi.org/10.1109/IGARSS52108.2023.10281682, 2023.

Buckel, J., Otto, J.-C., Prasicek, G., and Keuschnig, M.: Glacial lakes in Austria-Distribution and formation since the Little Ice Age, Glob. Planet. Change, 164, 39–51, https://doi.org/doi.org/10.1016/j.gloplacha.2018.03.003, 2018.

Chen, B., Zou, X., Zhang, Y., Li, J., Li, K., Xing, J., and Tao, P.: LEFormer: A hybrid CNN-transformer architecture for accurate lake extraction from remote sensing imagery, in: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5710–5714, https://doi.org/10.1109/ICASSP48485.2024.10446785, 2024a.

Chen, F., Zhang, M., Guo, H., Allen, S., Kargel, J. S., Haritashya, U. K., and Watson, C. S.: Annual 30 m dataset for glacial lakes in High Mountain Asia from 2008 to 2017, Earth Syst. Sci. Data, 13, 741–766, https://doi.org/10.5194/essd-13-741-2021, 2021.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation, https://doi.org/10.48550/arXiv.1802.02611, 22 August 2018.

Chen, X., Xu, X., Ma, Y., Wang, G., Chen, D., Cao, D., Xu, X., Zhang, Q., Li, L., Liu, Y., Liu, L., Li, M., Luo, S., Wang, X., and Hu, X.: Investigation of precipitation process in the water vapor channel of the Yarlung Zangbo Grand Canyon, Bull. Am. Meteorol. Soc., 105, 370–386, https://doi.org/10.1175/BAMS-D-23-0120.1, 2024b.

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R.: Masked-attention mask transformer for universal image segmentation, https://doi.org/10.48550/arXiv.2112.01527, 15 June 2022.

Dirscherl, M., Dietz, A. J., Kneisel, C., and Kuenzer, C.: Automated mapping of Antarctic supraglacial lakes using a machine learning approach, Remote Sens., 12, 1203, https://doi.org/10.3390/rs12071203, 2020.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale, https://doi.org/10.48550/arXiv.2010.11929, 3 June 2021.

Feyisa, G. L., Meilby, H., Fensholt, R., and Proud, S. R.: Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery, Remote Sens. Environ., 140, 23–35, https://doi.org/10.1016/j.rse.2013.08.029, 2014.

French, H. M.: The periglacial environment, John Wiley & Sons, 2017.

Furian, W., Maussion, F., and Schneider, C.: Projected 21st-century glacial lake evolution in High Mountain Asia, Front. Earth Sci., 10, 821798, https://doi.org/10.3389/feart.2022.821798, 2022.

García-Rodríguez, F., Piccini, C., Carrizo, D., Sánchez-García, L., Pérez, L., Crisci, C., Oaquim, A. B. J., Evangelista, H., Soutullo, A., and Azcune, G.: Centennial glacier retreat increases sedimentation and eutrophication in Subantarctic periglacial lakes: A study case of Lake Uruguay, Sci. Total Environ., 754, 142066, https://doi.org/10.1016/j.scitotenv.2020.142066, 2021.

Guo, W., Liu, S., Xu, J., Wu, L., Shangguan, D., Yao, X., Wei, J., Bao, W., Yu, P., Liu, Q., and Jiang, Z.: The second Chinese glacier inventory: data, methods and results, J. Glaciol., 61, 357–372, https://doi.org/10.3189/2015JoG14J209, 2015.

Haeberli, W., Kääb, A., Mühll, D. V., and Teysseire, P.: Prevention of outburst floods from periglacial lakes at Grubengletscher, Valais, Swiss Alps, J. Glaciol., 47, 111–122, https://doi.org/10.3189/172756501781832575, 2001.

Hall, D. K., Riggs, G. A., and Salomonson, V. V.: Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data, Remote Sens. Environ., 54, 127–140, https://doi.org/10.1016/0034-4257(95)00137-P, 1995.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, https://doi.org/10.48550/arXiv.1512.03385, 10 December 2015.

Heidarianbaei, M., Kanyamahanga, H., and Dorozynski, M.: Temporal ViT-U-Net tandem model: Enhancing multi-sensor land cover classification through transformer-based utilization of satellite image time series, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., 10, 169–177, https://doi.org/10.5194/isprs-annals-X-3-2024-169-2024, 2024.

Hou, M., Wei, J., Shi, Y., Hou, S., Zhang, W., Xu, J., Wu, Y., and Wang, H.: Hydroformer: Frequency domain enhanced multi-attention transformer for monthly lake level reconstruction with low data input requirements, Water Resour. Res., 60, e2024WR037166, https://doi.org/10.1029/2024WR037166, 2024.

Huggel, C., Kääb, A., Haeberli, W., Teysseire, P., and Paul, F.: Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps, Can. Geotech. J., 39, 316–330, https://doi.org/10.1139/t01-099, 2002.

Jain, S. K., Sinha, R. K., Chaudhary, A., and Shukla, S.: Expansion of a glacial lake, Tsho Chubda, Chamkhar Chu Basin, Hindukush Himalaya, Bhutan, Nat. Hazards, 75, 1451–1464, https://doi.org/10.1007/s11069-014-1377-z, 2015.

Jiao, L., Liu, Y., and Li, H.: Characterizing land-use classes in remote sensing imagery by shape metrics, ISPRS J. Photogramm. Remote Sens., 72, 46–55, https://doi.org/10.1016/j.isprsjprs.2012.05.012, 2012.

Karki, R., Hasson, S. ul, Schickhoff, U., Scholten, T., and Böhner, J.: Rising precipitation extremes across Nepal, Climate, 5, 4, https://doi.org/10.3390/cli5010004, 2017.

Khandelwal, A., Karpatne, A., Marlier, M. E., Kim, J., Lettenmaier, D. P., and Kumar, V.: An approach for global monitoring of surface water extent variations in reservoirs using MODIS data, Remote Sens. Environ., 202, 113–128, https://doi.org/10.1016/j.rse.2017.05.039, 2017.

Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Natl. Bur. Stand., 45, 255–282, https://doi.org/10.6028/jres.045.026, 1950.

Larsen, D. J., Blumm, A. R., Crump, S. E., Muscott, A. P., Abbott, M. B., Hangsterfer, A., and Porcelli, M.: Sedimentological characterization of earthquake-generated turbidites in fault-proximal glacial lakes: a case study from Jenny Lake, Teton range, Wyoming, Front. Earth Sci., 12, https://doi.org/10.3389/feart.2024.1391441, 2024.

Li, J. and Sheng, Y.: An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: a case study in the Himalayas, Int. J. Remote Sens., 33, 5194–5213, https://doi.org/10.1080/01431161.2012.657370, 2012.

Liaudat, D. T., Lenzano, M. G., and Castro, M.: Inventory and monitoring of rock glaciers and cryogenic processes in the Central Andes of Mendoza, Argentina: birth and extinction of a periglacial lake, in: Tenth International Conference on Permafrost–Proceedings, titleTranslation: 阿根廷门多萨中央安第斯山脉石冰川与寒冻过程编目与监测：一个冰缘湖泊的诞生与消亡, 419–425, 2012.

Lillesand, T., Kiefer, R. W., and Chipman, J.: Remote sensing and image interpretation, 7th ed., John

Wiley & Sons, 2015.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., and Guo, B.: Swin Transformer V2: Scaling up capacity and resolution, https://doi.org/10.48550/arXiv.2111.09883, 11 April 2022.

Loshchilov, I. and Hutter, F.: Decoupled weight decay regularization, https://doi.org/10.48550/arXiv.1711.05101, 4 January 2019.

Luo, X., Kuang, X., Jiao, J. J., Liang, S., Mao, R., Zhang, X., and Li, H.: Evaluation of lacustrine groundwater discharge, hydrologic partitioning, and nutrient budgets in a proglacial lake in the Qinghai–Tibet Plateau: using ^{222}Rn and stable isotopes, Hydrol. Earth Syst. Sci., 22, 5579–5598, https://doi.org/10.5194/hess-22-5579-2018, 2018.

Ma, D., Li, J., and Jiang, L.: Efficient glacial lake mapping by leveraging deep transfer learning and a new annotated glacial lake dataset, J. Hydrol., 133072, 2025.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, Int. J. Remote Sens., 17, 1425–1432, https://doi.org/10.1080/01431169608948714, 1996.

Nadachowski, P., Łubniewski, Z., and Tęgowski, J.: Glacial landform classification with vision transformer and digital elevation model, in: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, 7254–7258, https://doi.org/10.1109/IGARSS53475.2024.10641509, 2024.

Nagendra, H. and Rocchini, D.: High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail, Biodivers. Conserv., 17, 3431–3442, https://doi.org/10.1007/s10531-008-9479-0, 2008.

Nazakat, H., ul Hassan, S. N., Khan, G., and Ali, S.: Machine learning algorithms for extraction of glacial lakes using ground range detected (GRD) data: A case study from Hunza River Basin, Pakistan, https://doi.org/10.21203/rs.3.rs-590990/v1, 2021.

Nie, Y., Sheng, Y., Liu, Q., Liu, L., Liu, S., Zhang, Y., and Song, C.: A regional-scale assessment of

Himalayan glacial lake changes using satellite observations from 1990 to 2015, Remote Sens. Environ., 189, 1–13, https://doi.org/10.1016/j.rse.2016.11.008, 2017.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An imperative style, high-performance deep learning library, https://doi.org/10.48550/arXiv.1912.01703, 3 December 2019.

Peng, Y., He, J., Yuan, Q., Wang, S., Chu, X., and Zhang, L.: Automated glacier extraction using a Transformer based deep learning approach from multi-sensor remote sensing imagery, ISPRS J. Photogramm. Remote Sens., 202, 303–313, https://doi.org/10.1016/j.isprsjprs.2023.06.015, 2023.

Peppa, M. V., Maharjan, S. B., Joshi, S. P., Xiao, W., and Mills, J. P.: Glacial Lake evolution based on remote sensing time series: A case study of Tsho Rolpa in Nepal, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., 3, 633–639, https://doi.org/10.5194/isprs-annals-V-3-2020-633-2020, 2020.

Péwé, T. L.: The periglacial environment: past and present, McGill-Queen’s Press, 1969.

Romashova, K. V. and Chernov, R. A.: Inventory methodology of periglacial lakes in Spitzbergen (Svalbard), Probl. Arct. Antarct., 69, 158, https://doi.org/10.30758/0555-2648-2023-69-2-157-170, 2023.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: convolutional networks for biomedical image segmentation, https://doi.org/10.48550/arXiv.1505.04597, 18 May 2015.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W.: Monitoring vegetation systems in the Great Plains with ERTS, in: NASA. Goddard Space Flight Center 3rd ERTS-1 Symp, NTRS Author Affiliations: Texas A&M Univ.NTRS Report/Patent Number: PAPER-A20NTRS Document ID: 19740022614NTRS Research Center: Legacy CDMS (CDMS), 1974.

Roy, S. K., Deria, A., Hong, D., Rasti, B., Plaza, A., and Chanussot, J.: Multimodal Fusion Transformer for Remote Sensing Image Classification, IEEE Trans. Geosci. Remote Sens., 61, 1–20, https://doi.org/10.1109/TGRS.2023.3286826, 2023.

Sharma, A. and Prakash, C.: Quantifying glacial lake changes using deep learning models in the northwestern Himalayan region from 1992 to 2020, Int. J. Environ. Res., 19, 95, https://doi.org/10.1007/s41742-025-00769-9, 2025.

Sun, H. and Su, F.: Precipitation correction and reconstruction for streamflow simulation based on 262 rain gauges in the upper Brahmaputra of southern Tibetan Plateau, J. Hydrol., 590, 125484, https://doi.org/10.1016/j.jhydrol.2020.125484, 2020.

Tan, M. and Le, Q. V.: EfficientNet: Rethinking model scaling for convolutional neural networks, https://doi.org/10.48550/arXiv.1905.11946, 11 September 2020.

Tang, Q., Zhang, G., Yao, T., Wieland, M., Liu, L., and Kaushik, S.: Automatic extraction of glacial lakes from Landsat imagery using deep learning across the Third Pole region, Remote Sens. Environ., 315, 114413, https://doi.org/10.1016/j.rse.2024.114413, 2024.

Thati, J. and Ari, S.: A systematic extraction of glacial lakes for satellite imagery using deep learning based technique, Measurement, 192, 110858, https://doi.org/10.1016/j.measurement.2022.110858, 2022.

Veh, G., Lützow, N., Kharlamova, V., Petrakov, D., Hugonnet, R., and Korup, O.: Trends, breaks, and biases in the frequency of reported glacier lake outburst floods, Earths Future, 10, https://doi.org/10.1029/2021EF002426, 2022.

Wang, B. and French, H. M.: Permafrost on the Tibet plateau, China, Quat. Sci. Rev., 14, 255–274, https://doi.org/10.1016/0277-3791(95)00006-B, 1995.

Wang, P., Scherler, D., Liu-Zeng, J., Mey, J., Avouac, J.-P., Zhang, Y., and Shi, D.: Tectonic control of Yarlung Tsangpo Gorge revealed by a buried canyon in Southern Tibet, Science, 346, 978–981, https://doi.org/10.1126/science.1259041, 2014.

Wang, X., Ding, Y., Liu, S., Jiang, L., Wu, K., Jiang, Z., and Guo, W.: Changes of glacial lakes and implications in Tian Shan, central Asia, based on remote sensing data from 1990 to 2010, Environ. Res. Lett., 8, 044052, https://doi.org/10.1088/1748-9326/8/4/044052, 2013.

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z., and Tang, Z.: Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images, Earth Syst. Sci. Data, 12, 2169–2182, https://doi.org/10.5194/essd-12-2169-2020, 2020.

Wang, Y., Zheng, D., Zhang, G., Carrivick, J. L., Bolch, T., Ren, W., Guo, L., Su, J., Yuan, S., and Li, X.: Patterns and change rates of glacial lake water levels across High Mountain Asia, Natl. Sci. Rev., 12, nwaf041, https://doi.org/10.1093/nsr/nwaf041, 2025.

Xiang, Y., Zeng, C., Zhang, F., and Wang, L.: Effects of climate change on runoff in a representative Himalayan basin assessed through optimal integration of multi-source precipitation data, J. Hydrol. Reg. Stud., 53, 101828, https://doi.org/10.1016/j.ejrh.2024.101828, 2024.

Yan, X., Song, J., Liu, Y., Lu, S., Xu, Y., Ma, C., and Zhu, Y.: A Transformer-based method to reduce cloud shadow interference in automatic lake water surface extraction from Sentinel-2 imagery, J. Hydrol., 620, 129561, https://doi.org/10.1016/j.jhydrol.2023.129561, 2023.

Yao, X., Liu, S., Sun, M., and Zhang, X.: Study on the glacial lake outburst flood events in Tibet since the 20th century, J. Nat. Resour., 29, 1377–1390, https://doi.org/10.11849/zrzyxb.2014.08.010, 2014.

Yao, X., Liu, S., Han, L., Sun, M., and Zhao, L.: Definition and classification system of glacial lake for inventory and hazards study, J. Geogr. Sci., 28, 193–205, https://doi.org/10.1007/s11442-018-1467-z, 2018.

Yu, G.-A., Lu, J., Lyu, L., Han, L., and Wang, Z.: Mass flows and river response in rapid uplifting regions–A case of lower Yarlung Tsangpo basin, southeast Tibet, China, Int. J. Sediment Res., 35, 609–620, https://doi.org/10.1016/j.ijsrc.2020.05.006, 2020.

Zhang, G., Yao, T., Xie, H., Wang, W., and Yang, W.: An inventory of glacial lakes in the Third Pole region and their changes in response to global warming, Glob. Planet. Change, 131, 148–157, https://doi.org/10.1016/j.gloplacha.2015.05.013, 2015.

Zhang, G., Bolch, T., Yao, T., Rounce, D. R., Chen, W., Veh, G., King, O., Allen, S. K., Wang, M., and

Wang, W.: Underestimated mass loss from lake-terminating glaciers in the greater Himalaya, Nat. Geosci., 16, 333–338, https://doi.org/10.1038/s41561-023-01150-1, 2023.

Zhang, G., Carrivick, J. L., Emmer, A., Shugar, D. H., Veh, G., Wang, X., Labedz, C., Mergili, M., Mölg, N., and Huss, M.: Characteristics and changes of glacial lakes and outburst floods, Nat. Rev. Earth Environ., 5, 447–462, https://doi.org/10.1038/s43017-024-00554-w, 2024a.

Zhang, M., Chen, F., Guo, H., Yi, L., Zeng, J., and Li, B.: Glacial lake area changes in High Mountain Asia during 1990–2020 using satellite remote sensing, Research, 2022, 2022/9821275, https://doi.org/10.34133/2022/9821275, 2022a.

Zhang, T., Wang, W., Gao, T., An, B., and Yao, T.: An integrative method for identifying potentially dangerous glacial lakes in the Himalayas, Sci. Total Environ., 806, 150442, https://doi.org/10.1016/j.scitotenv.2021.150442, 2022b.

Zhang, T., Wang, W., and An, B.: Heterogeneous changes in global glacial lakes under coupled climate warming and glacier thinning, Commun. Earth Environ., 5, 374, https://doi.org/10.1038/s43247-024-01544-y, 2024b.

Zhao, C., Wei, H., Feyisa, G. L., de Castro Tayer, T., Ma, G., Wu, H., and Pan, Y.: Evaluating spectral indices for water extraction: Limitations and contextual usage recommendations, Int. J. Appl. Earth Obs. Geoinformation, 139, 104510, https://doi.org/10.1016/j.jag.2025.104510, 2025.

Zhao, H., Chen, F., and Zhang, M.: A systematic extraction approach for mapping glacial lakes in high mountain regions of Asia, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 11, 2788–2799, https://doi.org/10.1109/JSTARS.2018.2846551, 2018.

Zheng, D., Zhang, Q., and Wu, S.: Mountain geoecology and sustainable development of the Tibetan Plateau, Springer Science & Business Media, 2000.

Zhu, Q., Guo, H., Zhang, L., Liang, D., Wu, Z., Liu, Y., and Lv, Z.: GLA-STDeepLab: SAR enhancing glacier and ice shelf front detection using swin-TransDeepLab with global–local attention, IEEE Trans.

Geosci. Remote Sens., 61, 1–13, https://doi.org/10.1109/TGRS.2023.3324404, 2023.