Glacial Lakes Segmentation Using Multispectral Remote Sensing Data and Deep Learning Models

A dissertation submitted in
partial fulfilment for the degree of

Master of Technology

in

Computer Science

by

Debankan Das
Roll no. - **CS2336**

under the supervision of
Dr. Sarbani Palit
Computer Vision and Pattern Recognition Unit
(CVPR)

INDIAN STATISTICAL INSTITUTE, KOLKATA

June 2025

Declaration

Debankan Das, with Roll No. **CS2336**, hereby declare that the material presented in the dissertation titled **Glacial Lakes Segmentation Using Multispectral Remote Sensing Data and Deep Learning Models** represents original work carried out by me for the degree of **Master of Technology in Computer Science** at the **Indian Statistical Institute, Kolkata**.
Furthermore, I affirm that no sections of this report have been sourced or copied from external references without proper attribution. I am aware that any instances of plagiarism or the use of unacknowledged materials from third parties will be treated with the utmost seriousness and consequences.

Debankan Das
M.Tech (CS), CS2336
Indian Statistical Institute
Kolkata – 700108, India

CERTIFICATE

This is to certify that the dissertation entitled **"Glacial Lakes Segmentation Using Multispectral Remote Sensing Data and Deep Learning Models"** submitted by **Debankan Das** to the **Indian Statistical Institute, Kolkata**, in partial fulfillment of the requirements for the degree of **Master of Technology in Computer Science**, is an authentic and genuine record of the research work carried out by the candidate under my supervision and guidance. I affirm that the dissertation has met all the necessary requirements in accordance with the regulations of this institute.

Sarbani Palit 11-6-2025

Dr. Sarbani Palit
CVPR Unit
Indian Statistical Institute
Kolkata – 700108, India

Acknowledgements

I extend my sincere appreciation to Dr. Sarbani Palit, my advisor at the Computer Vision and Pattern Recognition Unit of the Indian Statistical Institute in Kolkata, for her guidance, continuous support, and inspiration. Her profound knowledge and creative suggestions have taught me a great deal in every subject and have shown me how to conduct solid research.

I am deeply grateful to all the teachers at the Indian Statistical Institute for their invaluable advice, insights, and instruction, which provided a crucial perspective to my research. Special thanks go to Subhranil Mustafi and all the other seniors at the SIPL (Signal and Image Processing Lab) for their constant mentoring.

Finally, I want to express my gratitude to my parents and extended family for their unwavering support. I also extend my sincere appreciation to all my friends for their continuous assistance and encouragement. I am thankful to everyone who has contributed to my growth and success, even if I have inadvertently missed mentioning them in the above list.

Abstract

The identification and delineation of glacial lakes through segmentation is crucial for tracking glacial changes and evaluating potential dangers from sudden flood events (GLOFs). These floods can severely impact populated areas and man-made structures downstream. Recent advances in high-quality satellite imagery have sparked increased attention toward using advanced machine learning methods, particularly deep learning, to enable precise and automated glacial lake detection.

In this study, we explore the effectiveness of deep learning-based pointwise semantic segmentation for glacial lake mapping using multisource remote sensing imagery, including both optical and synthetic aperture radar (SAR) data. We experiment with a novel stack combination that fuses optical and SAR features to evaluate its effectiveness for glacial lake segmentation. To this end, we trained and evaluated four segmentation models—U-Net, U-Net++, Attention U-Net, and DeepLabV3+—on the enriched dataset. The results demonstrate noticeable performance improvements across all models, affirming the quality and complementary nature of the fused input stack.

The study further investigates the spatiotemporal generalization capabilities of these models, examining how well they perform when transferred to unseen glacierized regions with differing topographic and climatic characteristics. A key finding is that integrating multispectral and multitemporal inputs, particularly through combined optical and SAR data, substantially enhances segmentation accuracy. Moreover, the choice of model architecture plays a critical role in achieving effective generalization across diverse terrains.

This research contributes to the development of robust and scalable tools for long-term glacial lake monitoring, providing valuable insights for climate change impact assessment and disaster risk mitigation.

Keywords: Glacial Lake Segmentation, Pixel Wise Classification, Sentinel 1 GRD Data, Landsat, NDWI, Convolutional Neural network (CNN), UNet.

Contents

Chapter 1

Introduction

Glacial lakes are water bodies formed by the accumulation of meltwater from glaciers, typically dammed by moraines or ice. Their formation and expansion are direct consequences of glacier retreat, a phenomenon accelerated by global climate change. Monitoring these lakes is crucial, as their growth can results in glacial lake outburst floods (GLOFs), posing significant risks to downstream communities.[1]

However, the remote and inaccessible nature of high-altitude glacial regions presents substantial challenges for consistent monitoring. Traditional methods, such as field surveys are labor-intensive and manual interpretation of satellite imagery is prone to errors, and often infeasible over large areas. Semi-automated approaches, while offering some improvements, still struggle with scalability and accuracy, especially in complex terrains with varying environmental conditions.

The advent of deep learning has revolutionized the field of remote sensing, offering powerful tools for feature extraction and image analysis. Convolutional Neural Networks (CNNs), in particular, have demonstrated extraordinary success in semantic segmentation tasks, enabling precise delineation of glacial lakes from satellite imagery. For instance, Jiang et al. (2024) developed a U-Net-based model utilizing multi-source satellite images to map landlocked lakes in Antarctica, achieving average F1 scores of 0.90, indicating strong spatiotemporal robustness across various study areas. [2].]

Integrating multispectral remote sensing data with deep learning models enhances the accuracy and reliability of glacial lake detection. Different studies with multi-source imagery, including PlanetScope RGB, Sentinel-2, Sentinel-1 SAR, and Landsat-8/9, to automatically map glacial lakes have shown excellent results. The models achieved high mapping accuracy demonstrating the effectiveness of combining diverse datasets. [2][3] [4]

The integration of synthetic aperture radar (SAR) data also addresses challenges posed by cloud cover and varying illumination conditions.[5] These advancements underscore the potential of deep learning models, especially when combined with multi-source remote sensing data, to overcome traditional limitations in glacial lake monitoring. By leveraging these technologies, it becomes feasible to conduct large-scale, accurate, and timely assessments of glacial lakes, which is essential for understanding glacier dynamics and mitigating associated risks.

Chapter 2

Related Work

2.1 Overview

Remote sensing data [3, 2, 4] has become a cornerstone for glacial lake segmentation due to its wide spatial coverage, frequent revisit cycles, and multispectral capabilities. The increasing accessibility of satellite imagery—such as optical data from Landsat or synthetic aperture radar (SAR) data from Sentinel-1—has enabled researchers to explore automated and accurate techniques for identifying glacial lakes. The availability of cloud-based platforms and powerful computing resources has further facilitated the training of deep learning models on large and complex remote sensing datasets.

Several studies have explored the use of various data stacks—optical, SAR, or a combination of both—for glacial lake segmentation. Optical imagery provides valuable spectral information, including indices like NDWI, which are useful for water body detection [6]. However, optical data can be hindered by cloud cover and seasonal variations. SAR data, being weather-independent and capable of penetrating cloud cover, serves as a complementary source [7]. The fusion of SAR and optical features has been shown to enhance model robustness and improve segmentation accuracy in challenging terrains [8].

With advancements in convolutional neural networks (CNNs), deep learning has emerged as a powerful approach for pixel-wise classification tasks. U-Net and its variants (such as Attention U-Net and U-Net++) have become standard architectures for medical and environmental image segmentation due to their ability to capture both local and global context through skip connections [9, 10]. Google's DeepLab family, particularly DeepLabV3 and DeepLabV3+, introduced atrous (dilated) convolutions to capture multi-scale features, pushing the state-of-the-art in semantic segmentation [11].

Recent research emphasizes model generalization across diverse geographies, temporal scales, and sensor modalities. Studies comparing different architectures and data combinations consistently highlight the importance of multi-source data integration and architecture choice in achieving reliable segmentation performance [12, 13]. This body of work lays the foundation for continued exploration into generalizable, scalable, and accurate models for glacial lake monitoring using remote sensing data.

2.2 CNN Models

In this section, we describe four distinct Convolutional Neural Network (CNN) architectures that are being used in Glacial Lake Segmentation. The models considered include U-Net [14], U-Net++ [15], Attention U-Net [16], and DeepLabV3+ [17]. A detailed description of each architecture and its relevance to our application is provided below.

2.2.1 UNet

U-Net is a widely used convolutional neural network (CNN) architecture designed for image segmentation tasks. It features a symmetrical U-shaped structure consisting of a downsampling path (encoder) and an upsampling path (decoder). The encoder extracts contextual information by reducing the spatial dimensions of the input, while the decoder reconstructs the segmentation map by upsampling and integrating features from the encoder through skip connections.

Unlike SegNet, U-Net employes skip connections that link corresponding layers of the encoder and decoder directly. These connections help preserve detailed spatial information, which is particularly important for accurately segmenting glacier ablation zones. By effectively combining low-level and high-level features, U-Net excels in tasks that demand precise boundary delineation, such as glacier segmentation. (Refer to Figure: 2.1)

2.2.2 U-Net++

UNet++ is an advanced neural network architecture designed to improve upon the traditional UNet model for image segmentation tasks, particularly in medical imaging. Unlike UNet, which uses simple skip connections between the encoder and decoder, UNet++ introduces nested skip pathways and dense convolutional blocks, creating a more refined feature aggregation mechanism. These nested connections bridge multiple encoder and decoder layers, reducing the semantic gap and enabling better gradient flow during training. Additionally, UNet++ supports deep supervision, allowing the

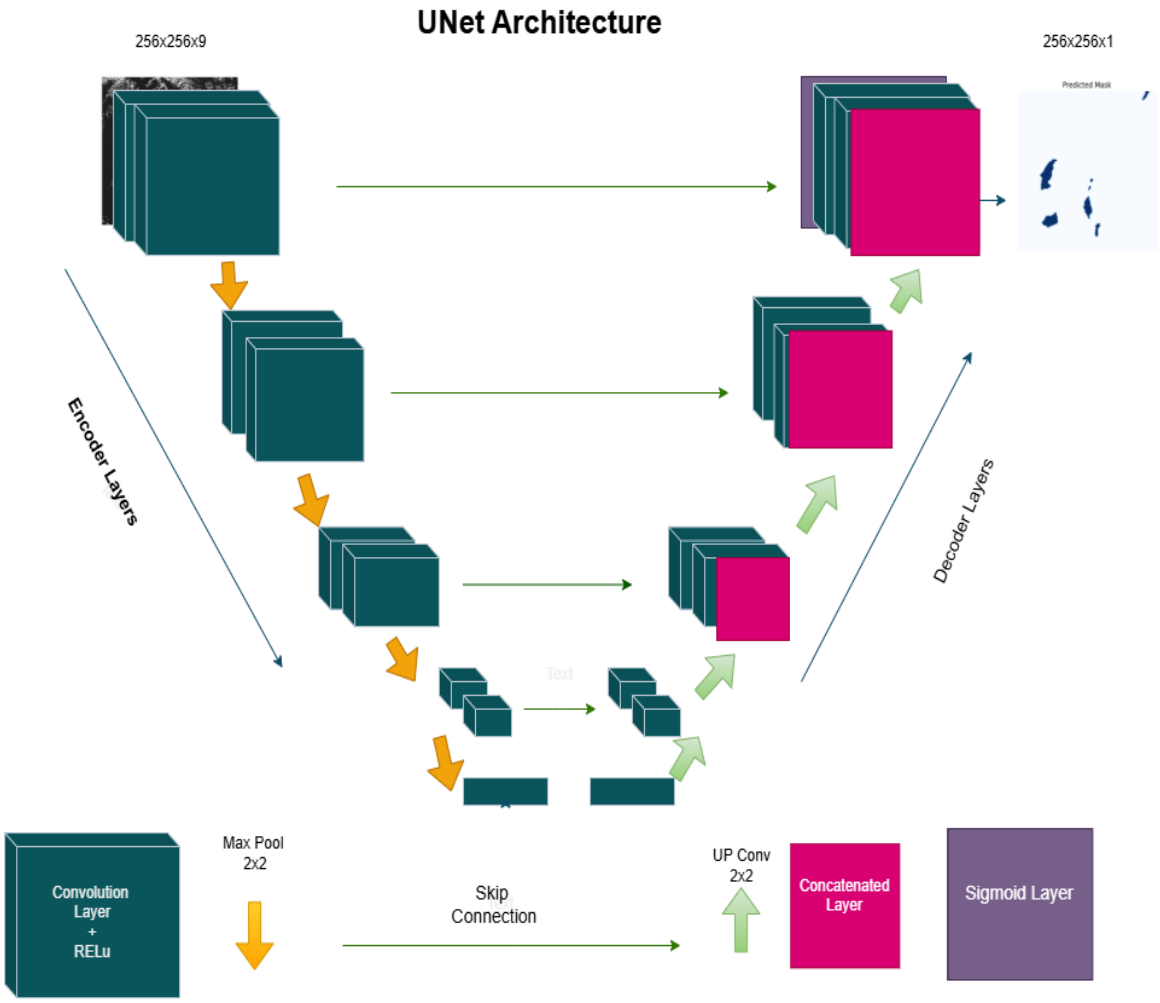


Figure 2.1: UNet Architecture

model to produce segmentation outputs at different levels, which enhances training stability and accuracy. The architecture's dense connections ensure that high-resolution features from the encoder are efficiently reused in the decoder, leading to more precise segmentation masks. Below is a conceptual visualization comparing UNet and UNet++.This design makes UNet++ particularly effective for complex segmentation tasks where fine-grained details are critical. (See figure:2.3a and 2.3b) [10]

2.2.3 Attention UNet

Attention U-Net is an modified version of the standard U-Net architecture, incorporating attention mechanisms to improve segmentation performance. These mechanisms enable the network to concentrate on the most relevant regions of the input while filtering out less important background areas. Specifically, attention gates are introduced in the skip connections between the encoder and decoder, allowing the model to selectively emphasize important features.

This selective focus results in more accurate segmentation, particularly beneficial for detecting glacier ablation zones, where the target regions may be small or highly variable. By highlighting the most critical areas of the input data, Attention U-Net achieves more refined and reliable outputs.

Through the application of four different CNN models in this study, including Attention U-Net, we aim to determine the most effective architecture for glacier ablation zone segmentation. Each model offers distinct benefits—from preserving fine spatial details in SegNet to focusing on salient features with attention mechanisms—enabling a thorough evaluation of their strengths. (Refer to 2.2)[10]

2.2.4 DeeplabV3+

DeepLabv3+ is an advanced convolutional neural network architecture designed for semantic segmentation, building upon the strengths of its predecessor, DeepLabv3. The key innovation in DeepLabv3+ is the incorporation of an encoder-decoder structure with atrous (dilated) convolutions and a modified Xception backbone, enabling precise pixel-level segmentation while maintaining computational efficiency. The encoder employs atrous spatial pyramid pooling (ASPP) to capture multi-scale contextual information by applying parallel dilated convolutions with different rates, effectively enlarging the receptive field without losing resolution. The decoder refines segmentation results by combining high-level semantic features from the encoder with low-level spatial details through skip connections, enhancing boundary accuracy. Additionally,

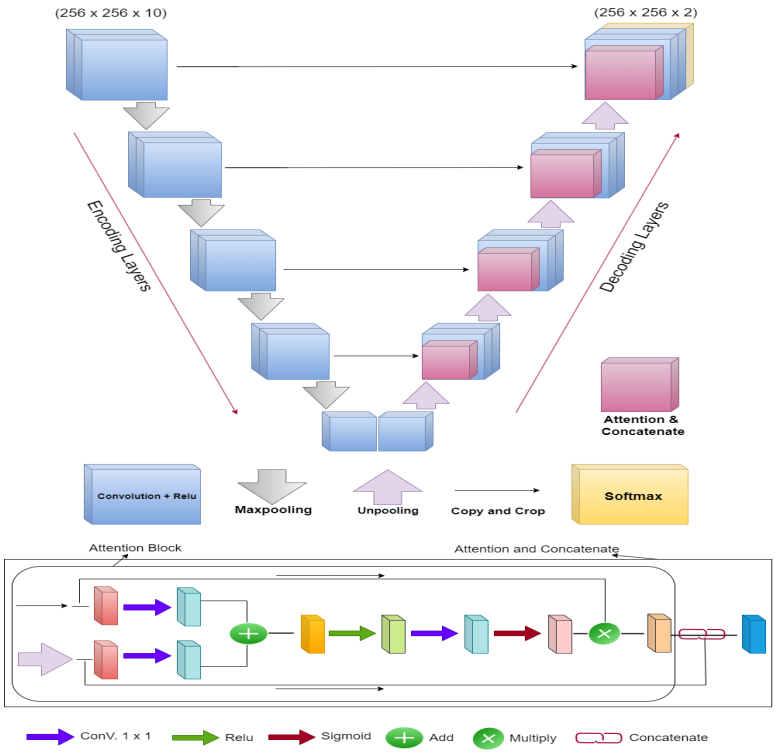


Figure 2.2: Attention U-Net Architecture

DeepLabv3+ leverages depthwise separable convolutions to reduce computational complexity while maintaining performance. This architecture excels in tasks requiring fine-grained segmentation, such as autonomous driving and medical imaging, by balancing global context understanding and local detail preservation. [17]

2.3 CNN Models Components

To gain insights into the computations of the CNN models, we outline the key components and techniques employed in our implementations:

2.3.1 Convolutional Layers (Conv2D)

Convolutional layers serve as the fundamental components of a Convolutional Neural Network (CNN). These layers perform convolution operations on the input data and forward the resulting feature maps to subsequent layers. In our implementation, 2D convolutional layers were used with 'same' padding, which maintains the original spatial dimensions of the input throughout the network by ensuring that the output size remains unchanged.

Example of a convolution operation with a stride of 1 with 'same' padding and using a

3×3 kernel:

$$\text{Input Matrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \quad \text{Kernel} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

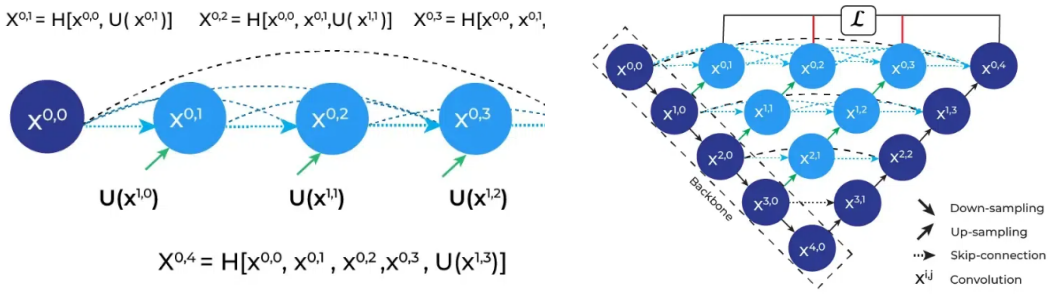
$$\text{Output Matrix} = \begin{bmatrix} 7 & 14 & 17 & 11 \\ 17 & 30 & 35 & 22 \\ 29 & 50 & 55 & 34 \\ 23 & 34 & 37 & 27 \end{bmatrix}$$

2.3.2 Nested Dense Skip Connection

UNet++ introduces **Dense nested skip pathways** [15, 10]to address the semantic gap between encoder and decoder more effectively than traditional UNet architectures. Unlike conventional skip connections that directly link corresponding encoder-decoder levels (e.g., layer 1 to layer 1), UNet++ implements a **multi-level hierarchical structure** where each decoder node (e.g., $X^{0,j}$) aggregates features from **all preceding encoder layers** ($X^{0,0}$, $X^{1,0}$, ..., $X^{i,0}$ where $i < j$) through densely connected convolutional blocks. These blocks, typically composed of Conv $\rightarrow$ BatchNorm $\rightarrow$ ReLU layers, refine features before fusion, reducing noise while preserving spatial details. The architecture employs **feature aggregation** via concatenation or summation, combining high-resolution details from shallow layers with rich semantic context from deep layers. This design offers three key advantages: (1) **reduced semantic gap** through multi-scale feature fusion, (2) **improved gradient flow** via nested connections that ease backpropagation, and (3) **enhanced boundary precision** by maintaining fine-grained details throughout the network. The result is superior segmentation accuracy, particularly in complex tasks like medical imaging where precise localization is critical.(See figure:2.3b and 2.3a) [15]

2.3.3 Dilated Convolution

Dilated convolution[18], also known as atrous convolution, is a powerful extension of standard convolution used to increase the receptive field without increasing the number of parameters or losing resolution. By inserting "holes" or zeros between the kernel elements, dilated convolutions allow the network to aggregate multi-scale contextual information more effectively. This property is particularly useful in dense prediction



(a) first skip pathway of UNet++. (b) nested dense convolutional blocks

Figure 2.3: Components of UNet++.

tasks such as semantic segmentation, where capturing long-range dependencies is crucial. Unlike pooling or striding, dilation preserves spatial resolution, making it ideal for precise boundary detection. In this study, dilated convolutions are employed in the segmentation architecture to enhance the model's ability to identify glacial lake boundaries across varying scales and textures. (See Figure:2.5)

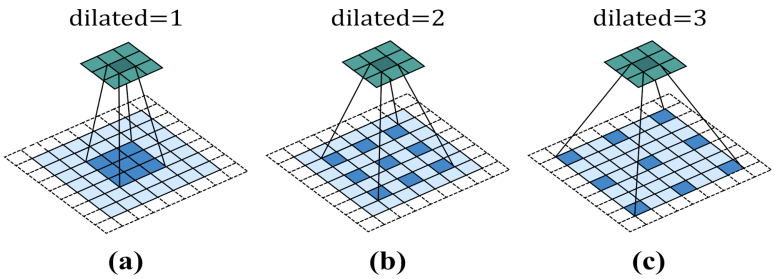


Figure 2.4: Dialated Convolution

2.3.4 Depthwise Convolution

Depthwise convolution [19] is a computationally efficient variant of standard convolution that significantly reduces the number of parameters and operations. Instead of applying convolution across all input channels simultaneously, it performs a separate spatial convolution for each channel. This operation is followed by a pointwise (1×1) convolution to combine the output channels, known collectively as depthwise separable convolution. DeepLabv3+ utilizes this approach in its Xception backbone to improve model performence without diminishing accuracy. By reducing the computational cost, depthwise convolutions make it feasible to apply DeepLabv3+ in real-time and resource-constrained environments. This design also enables deeper architectures while maintaining practical training and inference times.

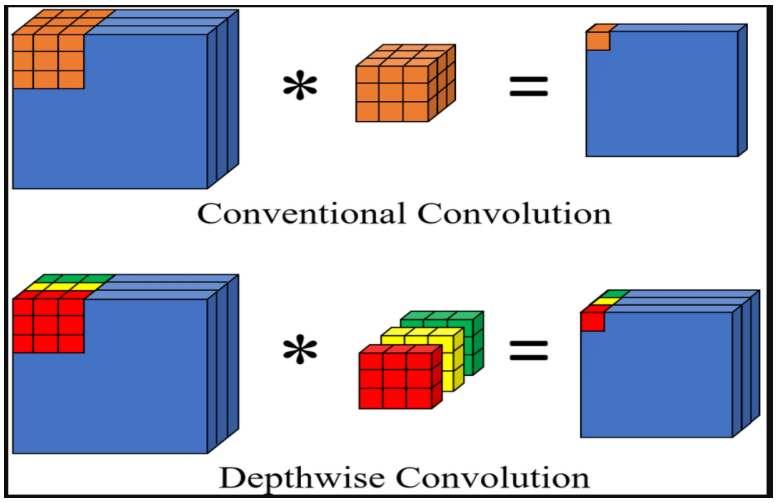


Figure 2.5: Depthwise Convolution

2.3.5 Loss Function:

The loss function measures the discrepancy between the model's predictions and the real values. To address the issue of class imbalance in our dataset, we combine Binary Cross-Entropy (BCE) Loss with Dice Loss, which focuses on maximizing the overlap (IoU) between predicted and actual masks, thereby mitigating the bias toward the more prevalent background class.

- **Binary Cross Entropy:** BCE [20] loss computes the logarithmic error between predicted probabilities ($\hat{y}$) and binary ground truth masks (y) at each pixel, driving the model to correctly classify glacial lake boundaries. Binary Cross Entropy Loss defined as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

  For each sample i, y_i represents the true binary label (either 0 or 1), while p_i denotes the model's estimated probability that sample i falls into the positive class (class 1).

- **Dice Coefficient and Dice Loss**

  - **Dice Coefficient (F1 Score):** The Dice coefficient, also known as the Sørensen–Dice index or F1 score, is a spatial overlap metric widely used in

segmentation tasks. It quantifies the similarity between two binary (or probabilistic) masks by comparing their intersection and union. The coefficient ranges from 0 (no overlap) to 1 (perfect overlap). For segmentation, it is defined as:

$$\text{Dice} = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i},$$

where:

* $p_i \in [0, 1]$ is the predicted probability for pixel i,
* $g_i \in \{0, 1\}$ is the ground truth label for pixel i,
* N is the total number of pixels.

To avoid division by zero when both masks are empty (e.g., in background regions), a small smoothing constant ϵ is often added:

$$\text{Dice}_\epsilon = \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon}.$$

- **Dice Loss:** Dice loss [21]is derived from the Dice coefficient and is designed to directly optimize region-based segmentation performance. It is particularly effective for imbalanced datasets (e.g., small glacial lakes against large background regions), as it focuses on the overlap between predictions and ground truth rather than pixel-wise errors. The loss is computed as:

$$\mathcal{L}_{\text{Dice}} = 1 - \text{Dice}_\epsilon = 1 - \frac{2\sum_{i=1}^{N} p_i g_i + \epsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon}.$$

2.3.6 Optimizer

The Adam optimizer [22] was employed in this study due to its ability to merge the strengths of both AdaGrad and RMSProp. It is known for its computational efficiency and low memory requirements, making it particularly suitable for training models on large datasets with high-dimensional parameter spaces.

2.3.7 Learning Rate

A learning rate of $\alpha = 0.001$ was applied, which determines the step size the optimizer uses when minimizing the loss function. Using a smaller value like this promotes more accurate convergence, although it typically requires a greater number of training epochs to achieve optimal results.

2.3.8 Activation Function:

The Rectified Linear Unit (ReLU) [23] activation function was employed in our model, which is mathematically expressed as:

$$\text{ReLU}(x) = \max(0, x)$$

This activation function contributes non-linear properties to the network while mitigating the vanishing gradient issue, resulting in improved training efficiency and performance.

2.3.9 MaxPooling Operation:

To downsample the input's spatial dimensions and enhance computational efficiency, MaxPooling layers were utilized. This technique also increases the model's robustness to minor positional variations by extracting the maximum activation within each local region of the feature map.

$$\text{Input Matrix} = \begin{bmatrix} 3 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 14 & 7 & 12 & 16 \end{bmatrix}$$

$$\text{MaxPooling with Stride 2} = \begin{bmatrix} \max(3,2,5,6) & \max(3,4,7,8) \\ \max(9,10,14,7) & \max(11,12,12,16) \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 14 & 16 \end{bmatrix}$$

2.3.10 Unpooling:

2.3.11 Unpooling Layers:

Unpooling layers are employed in the decoder section of networks such as SegNet to upsample feature maps. These layers restore the original spatial dimensions by repositioning the max-pooled values to their initial locations, thereby aiding in the reconstruction of high-resolution feature representations.

$$\text{Input Matrix} = \begin{bmatrix} 6 & 10 \\ 14 & 16 \end{bmatrix}$$

$$\text{Upsampling} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 10 \\ 0 & 0 & 0 & 0 \\ 14 & 0 & 0 & 16 \end{bmatrix}$$

By using these components, we optimized CNN models specifically for detecting glacier ablation zones. Every aspect—including the activation functions and advanced optimizers such as Adam—was carefully chosen to improve the model's capacity for learning intricate patterns and delivering precise predictions.

Chapter 3

Study Area and Dataset

3.1 Geographic and Environmental Summary

The glacial lake segmentation model is trained and evaluated using satellite imagery from six distinct sites spanning various regions of the Himalayan arc and its vicinity. These sites represent a diverse mix of geographic, topographic, and hydrological conditions.

- Two sites lie within the **Central and Eastern Himalaya**, including regions with significant monsoon influence.
- Two additional regions are located further east near the borders of **Arunachal Pradesh and Tibet**, areas known for high precipitation and larger, moraine-dammed glacial lakes.
- One bounding region is located westward near the **Nepal–Tibet border**, covering areas influenced by both monsoon and westerly systems.
- Another region represents the **easternmost glacial terrain**, encompassing high-elevation basins near the Indian–Tibetan plateau transition zone.

These sites have been selected to account for the variability in lake size, shape, elevation range, and sediment load, ensuring that the model can generalize well across different types of glacial lakes and environmental conditions.

Geographic Extent Summary (from KML files)

These regions collectively cover both **Western and Eastern Himalayas**, capturing diverse climatic zones ranging from *westerly-dominated cold deserts* to *monsoon-fed*

Site Label	Longitude Range	Latitude Range
Region 1	86.06 – 87.55	27.11 – 28.18
Region 2	85.37 – 85.94	28.01 – 29.00
Region 3	95.91 – 97.16	28.18 – 29.37
Region 4	94.48 – 96.13	30.13 – 30.98
Region 5	92.61 – 94.10	28.33 – 29.62
Region 6	92.79 – 93.77	28.86 – 29.66

Table 3.1: Geographic extent of glacial lake segmentation training and test sites.

glacial basins.

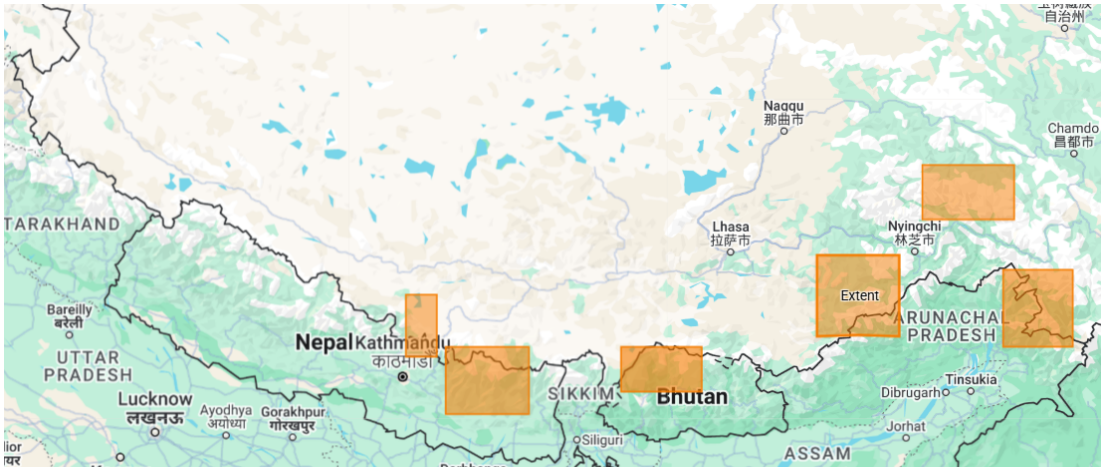


Figure 3.1: Study Area in Central and Eastern Himalaya

3.2 Dataset

Landsat 9, the newest satellite in the Landsat mission series, delivers essential Earth observation data that supports both scientific research and practical applications across various domains [24]. It is equipped with two key instruments: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS), which together acquire high-resolution multispectral imagery. These data are invaluable for tracking land use, detecting land cover changes, and assessing environmental conditions. For this study, six spectral bands from Landsat 9 were utilized:

- **Band 2**: Ideal for mapping water bodies and distinguishing between bare soil and vegetation.
- **Band 3**: Useful in evaluating the health and vigor of vegetation.

- **Band 4**: Aids in identifying different types of vegetation and defining soil boundaries.
- **Band 5**: Essential for analyzing biomass and delineating water features.
- **Band 6**: Effective in assessing soil and vegetation moisture levels.
- **Band 7**: Valuable for fire detection and examining soil and geological properties.

In Landsat imagery, a band's wavelength refers to the segment of the electromagnetic spectrum that the satellite sensor is configured to record. Each band is sensitive to a distinct spectral region—ranging from visible light to near-infrared (NIR), shortwave infrared (SWIR), and thermal infrared (TIR). The specific wavelength range of a band determines the nature of the surface features it can detect and the type of environmental information it can provide.

Spectral Band	**Range**	**Applications**
Visible Light (BGR)	0.4–0.7 µm	• True-color imagery • Vegetation analysis • Water/soil/urban mapping
Near-Infrared (NIR)	0.7–1.0 µm	• Plant health monitoring • Biomass estimation • Water detection
Shortwave Infrared (SWIR)	1.0–3.0 µm	• Moisture assessment • Geological studies • Fire detection

Besides using these landsat bands we generate a new layer using Band 3 and Band 5 which is called Normalized Difference Water Index(NDWI).This is computed from the formula:

$$\text{NDWI} = \frac{\text{Band 3} - \text{Band 5}}{\text{Band 3} + \text{Band 5}}$$

This index is instrumental in separating water bodies from other land cover types. A threshold value of 0.4 was applied to isolate water features, producing a binary mask where pixels with NDWI values above 0.4 were classified as water. This thresholding

step is vital for transforming continuous NDWI values into a clear-cut binary classification.

To overcome the challenges posed by cloud cover and to leverage the complementary information provided by SAR data, Sentinel-1 vv-polarized images were incorporated as one of the input stack layers. All Sentinel-1 data were acquired between February and May 2025.

To train and evaluate all four architectures, we employed a set of 9 spectral input layers. These include six previously described bands from Landsat-9, one processed Sentinel-1 VV polarized layer, a Landsat-derived NDWI layer, and a water mask. The Landsat-9 bands were obtained from the USGS Earth Explorer under the Landsat Collection 2 Level-2 dataset (Landsat 8-9 OLI/TIRS CS L2), while the Sentinel-1 data was acquired from the Alaska Satellite Facility.

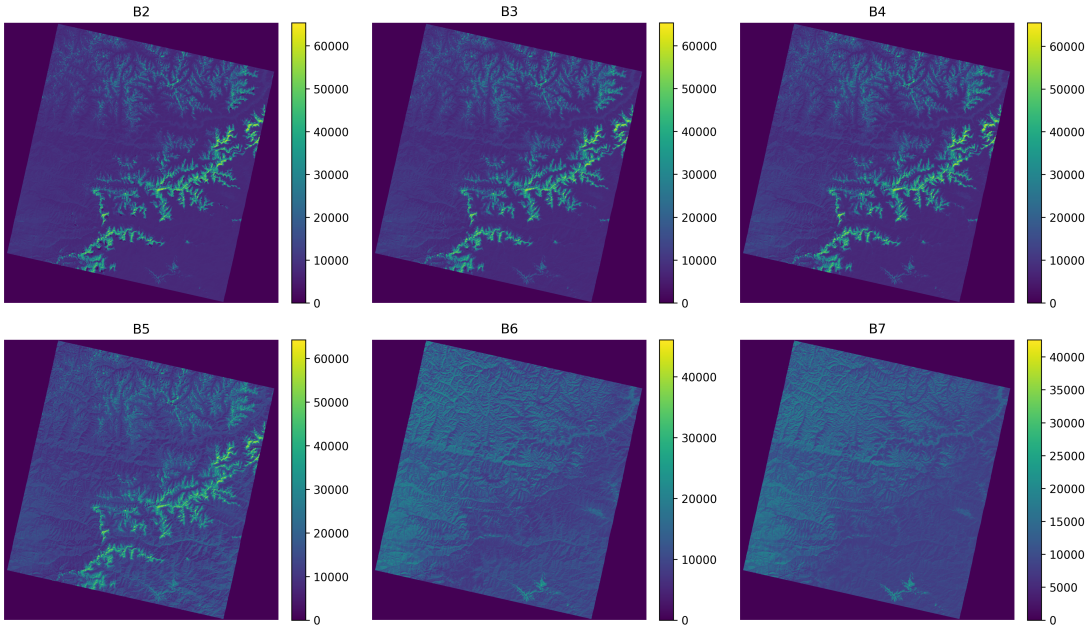
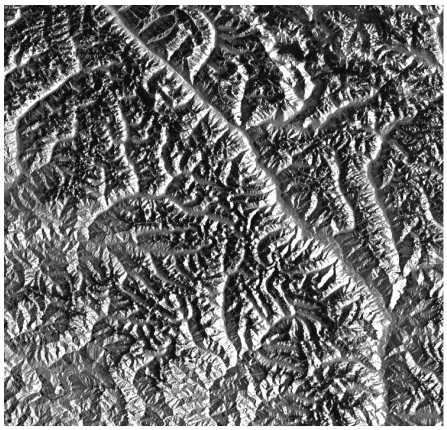
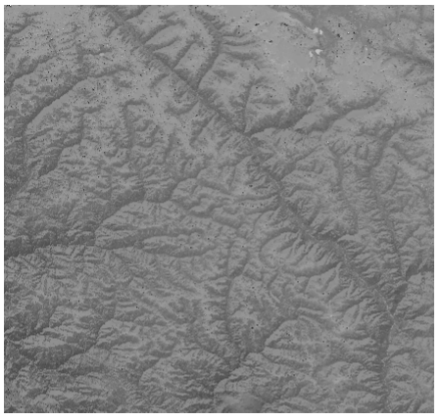


Figure 3.2: Landsat Bands (Band 2-7)

All nine layers were chosen for their distinct advantages, with each deemed essential. The Landsat-9 bands provide a broad spectrum of spectral data, crucial for analyzing different land cover and land use types. Sentinel-1's VV polarization detects surface texture and structure, which is particularly useful for identifying glacial lakes. Smooth surfaces, such as lakes, reflect low backscatter and appear darker in SAR images. This helps distinguish lakes from ice, snow, or rocky terrain, which show higher backscatter. While optical data captures spectral features, SAR provides complementary physical



(a) Sentinel 1 GRD vv polarized layer.



(b) NDWI layer.

Figure 3.3

surface information. To ensure consistency in magnitude, data values of each layer were normalized based on their lowest and highest pixel values.This normalization step ensures consistency across various data layers, which is crucial for reliable model training and evaluation.

Cloud presence, especially in high-altitude glacierized regions, poses major obstacles for analyzing optical remote sensing imagery. Clouds can hide important landscape features, resulting in gaps or inaccuracies in the data. To address this, only images with low cloud coverage were used. All Landsat imagery was acquired between May and November 2024, carefully chosen to ensure high-quality visuals. Additionally, selecting scenes from different path/row combinations enhances the model's ability to generalize across diverse spatial settings.

The Normalized Difference Water Index (NDWI) is a frequently used remote sensing index for detecting surface water bodies. It utilizes Band 3 (green) and Band 5 (near-infrared, NIR) to exploit the distinct spectral reflectance of water. Water surfaces reflect strongly in the green band but absorb heavily in the NIR band, resulting in a strong negative contrast. This contrast enables effective separation of water from surrounding land and vegetation, which typically reflect more in the NIR. The NDWI enhances water body delineation and is particularly useful in monitoring lakes, reservoirs, and seasonal flooding in remote sensing applications.

Once we have all the input band layers we need to generate ground truth mask as labels for training of deep models.For this purpose we use the Hi-MAG (High Mountain Asia Glacial-lake) database which provides detailed annual vector polygon maps of glacial lakes across High Mountain Asia from 2008 to 2017. Derived from 30-meter resolution Landsat imagery, the dataset includes lakes as small as 0.0081 km^2 and covers regions

such as the Himalayas, Karakoram, Pamirs, Tien Shan, and the Tibetan Plateau. It offers valuable information for monitoring glacial lake dynamics, including lake expansion trends and elevation distributions. The dataset is particularly useful for assessing glacial lake outburst flood (GLOF) hazards, studying glacier-lake-climate interactions, supporting hydrological and glaciological models, and observing environmental changes in high-altitude regions.

Figure 3.4: Hi MAG vector Polygons of Glaical Lakes

The adoption of Landsat-9 over Landsat-8 for glacier ablation zone mapping presents significant benefits. The newer satellite maintains data continuity while offering superior sensor performance, particularly through its enhanced Operational Land Imager (OLI) that delivers improved spectral resolution and radiometric accuracy. These technological advancements facilitate more precise identification and analysis of glacial features, including ablation zone boundaries.

The synergistic use of Landsat-9 optical data with Sentinel-1 SAR ensures reliable, long-term observation capabilities critical for glacial lake monitoring. This multi-sensor approach provides a robust dataset for analyzing temporal patterns in cryospheric dynamics, offering essential insights into climate change impacts on glacial systems [1].

Key advantages of this combined approach include:

- Superior data quality through advanced sensor technology
- Enhanced spectral and thermal resolution for feature discrimination
- Continuous temporal coverage for change detection studies
- Complementary data fusion from optical and radar systems

These characteristics establish Landsat-9 and Sentinel-1 as optimal tools for contemporary glaciological studies and climate change research.

Chapter 4

Methodology

4.1 Data Preprocessing

After downloading the required satellite datasets, a comprehensive preprocessing pipeline was implemented to prepare the data for model training. It is important to ensure that all processing steps maintain a consistent coordinate reference system (CRS) of WGS 84 (EPSG:4326) and a spatial resolution of 30 meters. This consistency is essential for accurate spatial alignment, pixel-wise mask generation, and downstream analysis. The preprocessing workflow is described below:

1. **Sentinel-1 Preprocessing using SNAP:** The Sentinel-1 Ground Range Detected (GRD) product was processed using the Sentinel Application Platform (SNAP) [25]. The following steps were applied to the VV-polarized band, ensuring that the output adhered to the desired Coordinate Reference System (CRS) and spatial resolution:

- **Apply Precise Orbit Files:** To correct satellite positioning and improve geometric accuracy.
- **Thermal Noise Removal:** Removes additive noise that may interfere with backscatter values.
- **Radiometric Calibration:** Converts raw backscatter to calibrated sigma nought (σ°) values [26].
- **Speckle Filtering:** A Lee filter was applied to suppress speckle noise while preserving edge features [27].
- **Terrain Correction:** Geocoded the imagery using the SRTM 1-second DEM,

aligning it to **WGS 84 (EPSG:4326)** with a resampled spatial resolution of **30 meters**.

2. **Landsat and Sentinel Band Alignment**

   - **Region of Interest (ROI) Extraction:** A common spatial extent was defined and extracted from both datasets using QGIS's *Raster Extraction by Extent* tool.
   - **Cropping:** Each band was clipped to the selected ROI, ensuring identical dimensions and alignment between the datasets.
   - **NDWI Calculation and Water Mask Generation:** The Normalized Difference Water Index (NDWI) was calculated to highlight water bodies. A binary water mask was then generated by applying a threshold to the NDWI image [28].

3. **Normalization of Input Bands**

   - **Normalization Process:** All nine input layers—comprising eight Landsat-9 bands, Sentinel-1 VV polarization, NDWI, and the water mask—were normalized to a common scale.
   - **Normalization Formula:**

$$\text{Normalized Value} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

     where:

     - X = Original pixel value
     - X_{min} = Minimum pixel value in the layer
     - X_{max} = Maximum pixel value in the layer

     This scaling technique standardizes the data range, enhancing model training efficiency, promoting faster convergence, and ensuring balanced feature contributions [26].

4. **Merging Input Layers:** Once normalized, the following nine layers (Figure 4.1) were merged in the following sequence to create a three-dimensional image:

   - Landsat Band 2 (Blue)
   - Landsat Band 3 (Green)
   - Landsat Band 4 (Red)
   - Landsat Band 5 (Near Infrared)

- Landsat Band 6 (Shortwave Infrared 1)
- Landsat Band 7 (Shortwave Infrared 2)
- Sentinel-1 VV polarization processed layer
- NDWI (Normalized Difference Water Index) layer
- Water mask layer

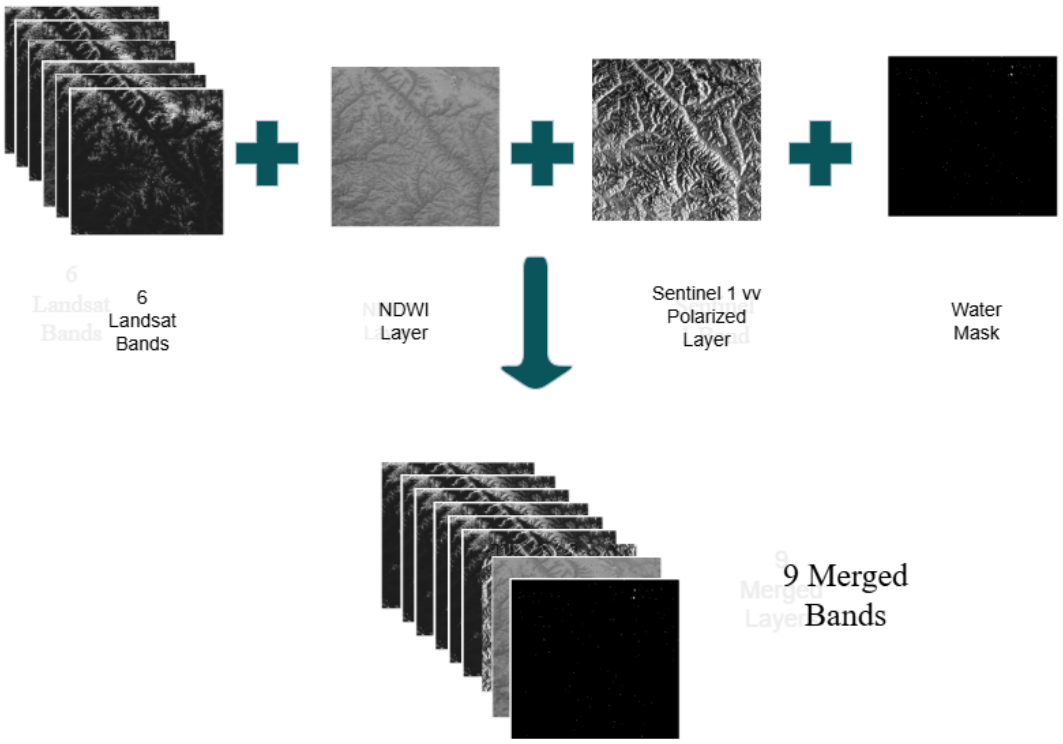


Figure 4.1: Input Structure

This stage combines spectral characteristics, topographic details, and geomorphometric attributes into a unified dataset, offering a holistic representation of the landscape. Such multi-dimensional integration enables the CNN architecture to simultaneously analyze different terrain properties and surface features, thereby enhancing its feature extraction capabilities (refer to Figure 4.1).

5. **Lake Mask Generation:**

- **Rasterization:** The lake shapefile was converted into raster format using QGIS's *Vector to Raster* processing tool. Pixels inside lake polygons were assigned a value of 1.

- **Resolution Matching:** The output raster was generated at a 30-meter resolution to match the satellite image bands.
- **CRS Alignment:** The lake mask was exported in the WGS 84 (EPSG:4326) coordinate reference system.
- **ROI Cropping:** The lake raster was clipped to the common region of interest (ROI) using the same spatial extent as the satellite bands [29].

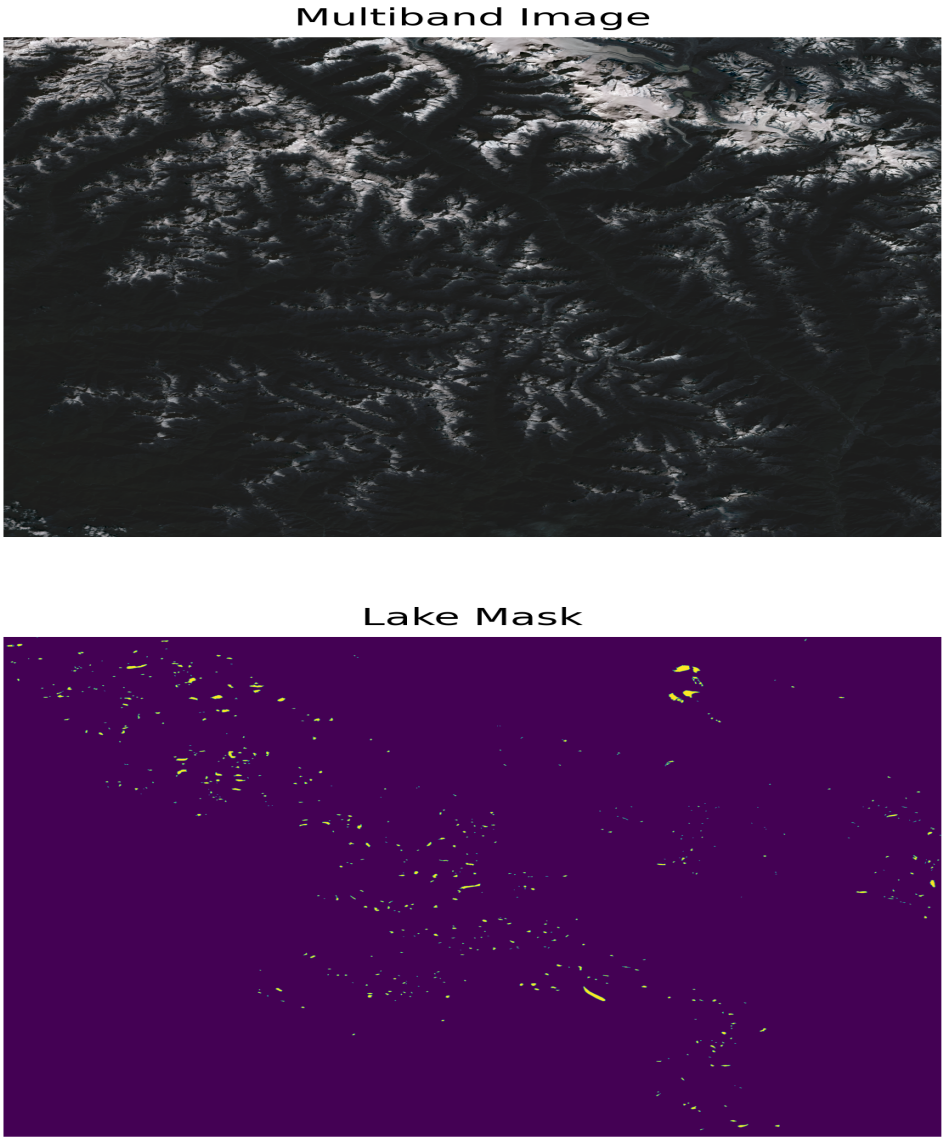


Figure 4.2: Multiband Image of Region 4 and corresponding Mask

6. **Subsampling the Image:** To train the models effectively, we constructed the

training dataset by extracting fixed-size patches of 256×256 pixels with a stride of 128 from high-resolution satellite imagery covering six distinct glaciated regions. This subsampling strategy ensured sufficient spatial coverage while introducing overlap between neighboring patches to improve model generalization. During the patch extraction process, any input tile that did not contain glacial lake pixels was discarded. This filtering step helped to maintain a meaningful representation of the target class (glacial lakes) and mitigated the impact of class imbalance in the training data.

- 256×256 patch size with a stride of $128 \rightarrow 2339$ images

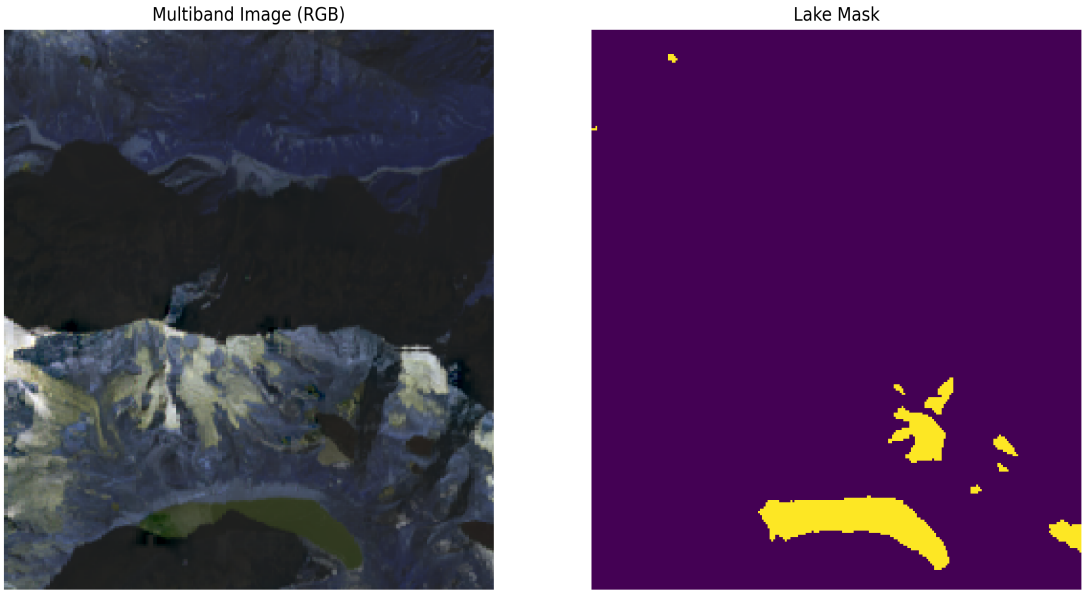


Figure 4.3: One subsampled label image of shape (256×256) and Corresponding Mask

7.Software Used for Data Preprocessing

All preprocessing steps essential for preparing the satellite imagery and glacial lake masks were conducted using a combination of open-source Geographic Information System (GIS) software and Python libraries.

- **QGIS**: Used extensively for geospatial operations such as reprojection, clipping, and visual inspection of raster and vector datasets. Its graphical interface facilitated manual verification and correction steps.
- **Python Libraries**:
  - `rasterio`: Employed for reading, writing, and manipulating georeferenced raster datasets, particularly for tasks such as resampling, masking, and band stacking.

- `numpy` and `matplotlib`: Utilized for general-purpose array manipulation, visualization, and analysis of pixel-level data.

- **Google Earth Pro**: Select images and region-of-interest (ROI) boundaries were manually extracted using Google Earth Pro for enhanced context and visualization of glacial lake environments.

4.2 Model Architecture and Experimental Setup

To perform pixel-wise segmentation of glacial lakes, four widely adopted deep learning models were trained and evaluated: **U-Net**, **U-Net++**, **Attention U-Net**, and **DeepLabV3+**. These architectures are known for their effectiveness in semantic segmentation tasks, particularly in remote sensing and medical imaging domains.

All models were implemented using the `segmentation_models.pytorch` library[30], which provides pretrained encoder backbones and configurable decoder architectures. In our implementation, a **ResNet-34** encoder with pretrained weights on ImageNet was utilized for all four models to leverage transfer learning and improve convergence speed and accuracy.

Initially, the Binary Cross-Entropy (BCE) loss function was used. However, due to the class imbalance inherent in the dataset (with glacial lakes occupying a relatively small fraction of the scene), we adopted a composite loss function combining **BCE** and **Dice Loss**, with a stability factor of 1×10^{-6} to prevent division by zero and enhance numerical stability.

Experimental Setup

- Batch Size = 8
- Learning Rate = 0.001
- Loss Function: **BCE + Dice Loss**, with stability factor 1×10^{-6}
- Train-Test Split: 80% for training and 20% for testing
- Encoder Backbone: **ResNet-34** pretrained on ImageNet
- Hardware: All models were trained using a **T4 GPU** provided by Kaggle Notebooks

Chapter 5

Experiments and Results

5.0.1 Experiments and Evaluation

We detail the experimental configuration and discuss the results of our CNN-based glacial lake segmentation study. Key metrics—including IoU, Accuracy, Precision, Recall, and F1 Score—are thoroughly examined. The model's performance is further benchmarked against existing approaches to validate its superiority.

Performance Metrics: The following metrics were used to evaluate the model's performance:

- **Intersection over Union (IoU)**: Quantifies the spatial overlap between predicted segmentation and ground truth. Computed as the ratio of intersection area to union area:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

  where P represents predicted region and G denotes ground truth.

- **Accuracy**: Measures overall prediction correctness across all classes:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

- **Precision**: Assesses the reliability of positive predictions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall (Sensitivity): Evaluates the model's ability to detect all relevant instances by

computing the proportion of true positives among all actual positive cases:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 Score: Represents the balanced harmonic mean between precision and recall, providing a single metric incorporates the information of both false positives and false negatives:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

- TP: True Positives (positives which are correctly identified)
- TN: True Negatives (negatives which are correctly identified)
- FP: False Positives (positives identified as negative)
- FN: False Negatives (negative identified as positives)

5.1 Lake Segmentation Results

To evaluate the impact of remote sensing data diversity on the robustness and generalization capability of glacial lake segmentation models, we conducted an experimental ablation study. Specifically, we designed two input configurations: one incorporating Sentinel-1 SAR (Synthetic Aperture Radar) data along with other optical bands, and another excluding the SAR layer. This comparative setup was aimed at investigating the contribution of SAR data in enhancing segmentation performance, particularly under varying surface and illumination conditions.

The results clearly indicate a noticeable performance degradation when the Sentinel-1 layer is excluded from the input stack. This highlights the complementary nature of SAR data, especially in detecting glacial features that may be obscured in optical imagery due to cloud cover, shadow, or seasonal variation.

The following tables summarize the key performance metrics—Precision, Recall, and F1 Score—for both input configurations. These findings emphasize the importance of multi-source data fusion in remote sensing-based segmentation tasks.

Table 5.1: Performance comparison with Sentinel 1 layer

Metric	UNet	UNet++	DeepLabV3+	Attention Net
Precision	0.9557	0.9647	0.9246	0.9427
Recall	0.9163	0.9329	0.9038	0.8861
F1-Score	0.9356	0.9486	0.9141	0.9136
IoU	0.721	0.786	0.7098	0.711
Epochs	140	140	160	140

Table 5.2: Performance comparison without Sentinel 1 layer

Metric	UNet	UNet++	DeepLabV3+	Attention Net
Precision	0.9432	0.9426	0.9149	0.9271
Recall	0.9077	0.9311	0.9007	0.8641
F1-Score	0.925	0.9373	0.9074	0.8944
IoU	0.708	0.752	0.6872	0.6824
Epochs	140	140	160	140

Prediction With 9 layer Input by UNet++:

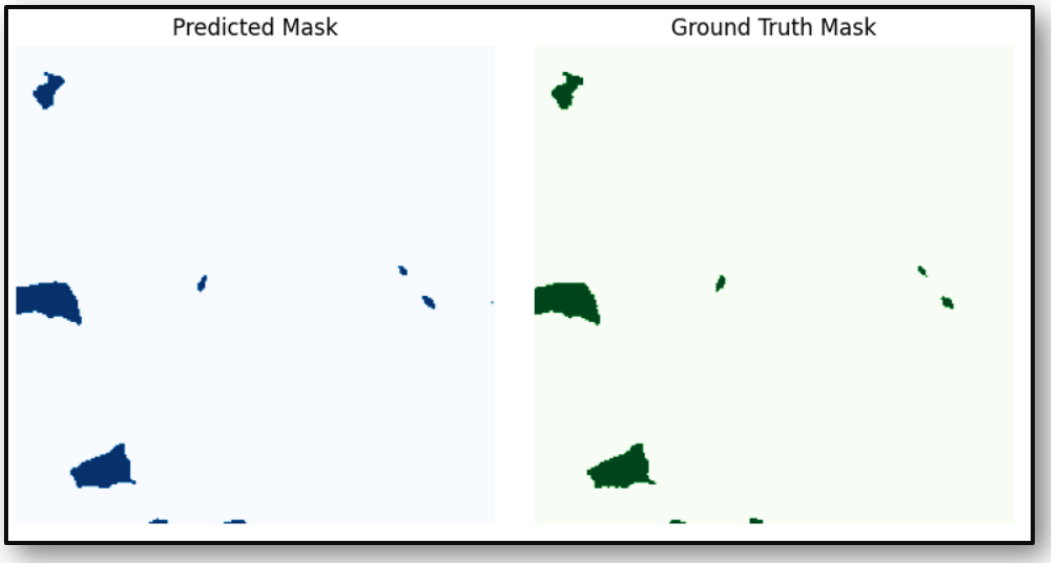


Figure 5.1: Prediction vs Ground Truth

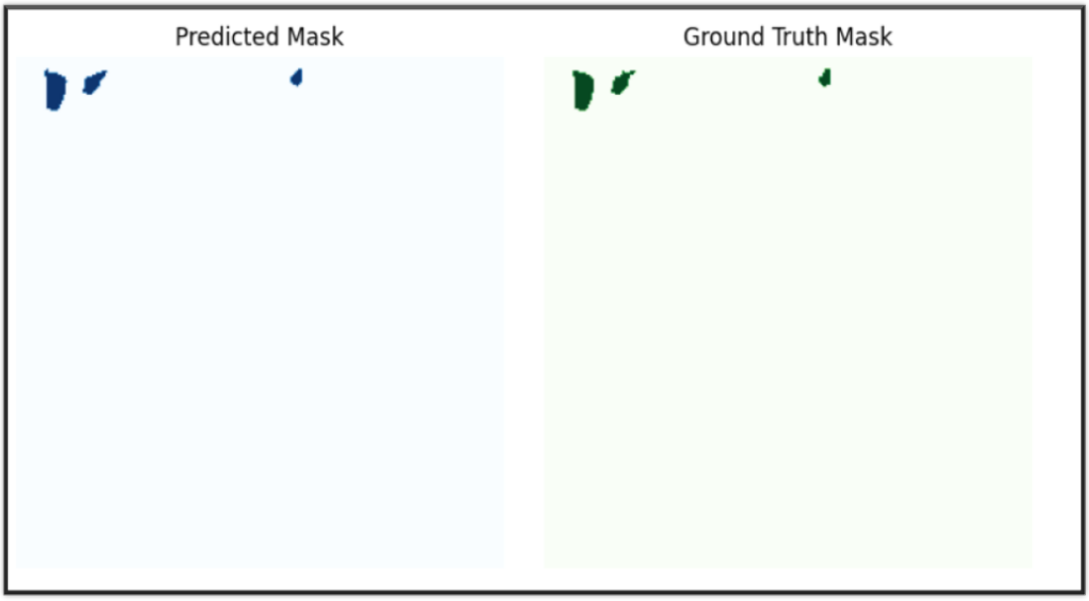


Figure 5.2: Small Lakes prediction

5.1.1 Train vs Validation Loss Plots

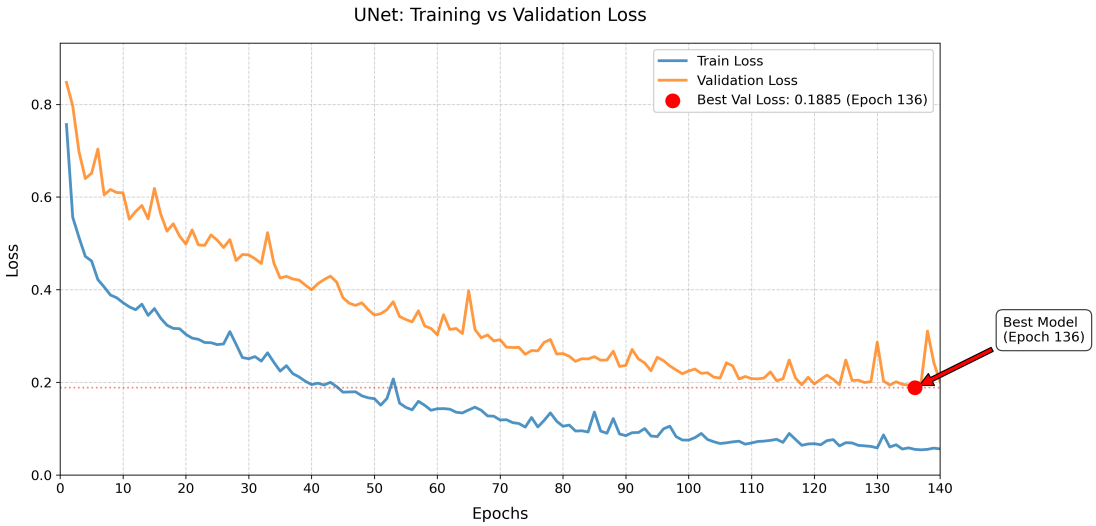


Figure 5.3: Training Vs Validation Loss of UNet

Figure 5.4: Training Vs Validation Loss of Attention Net

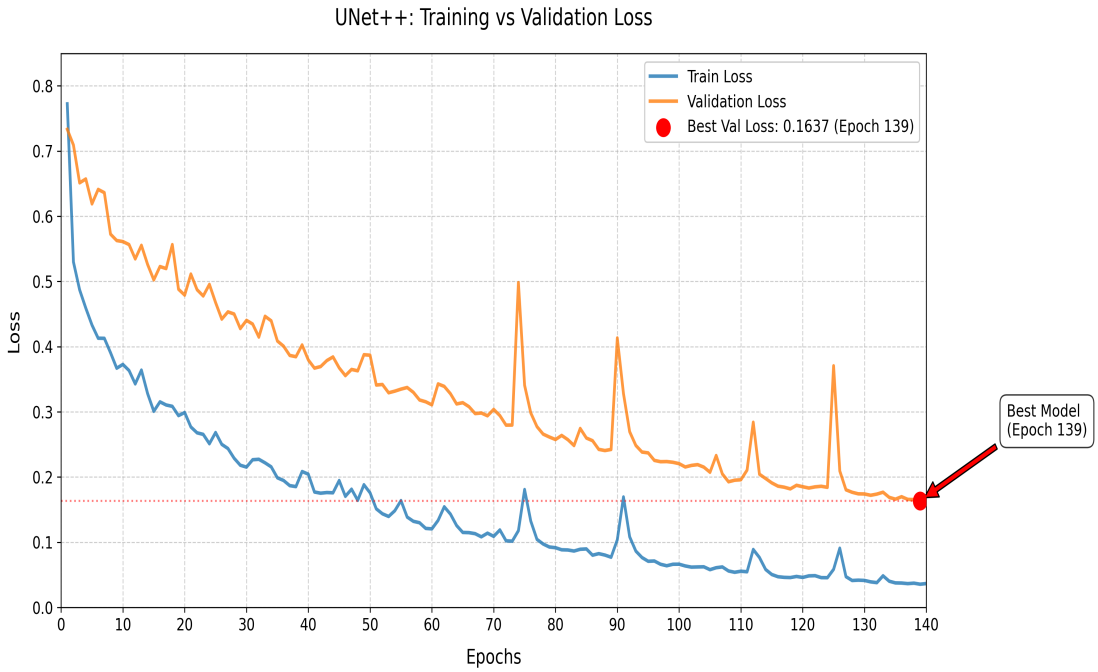


Figure 5.5: Training Vs Validation Loss of UNet++

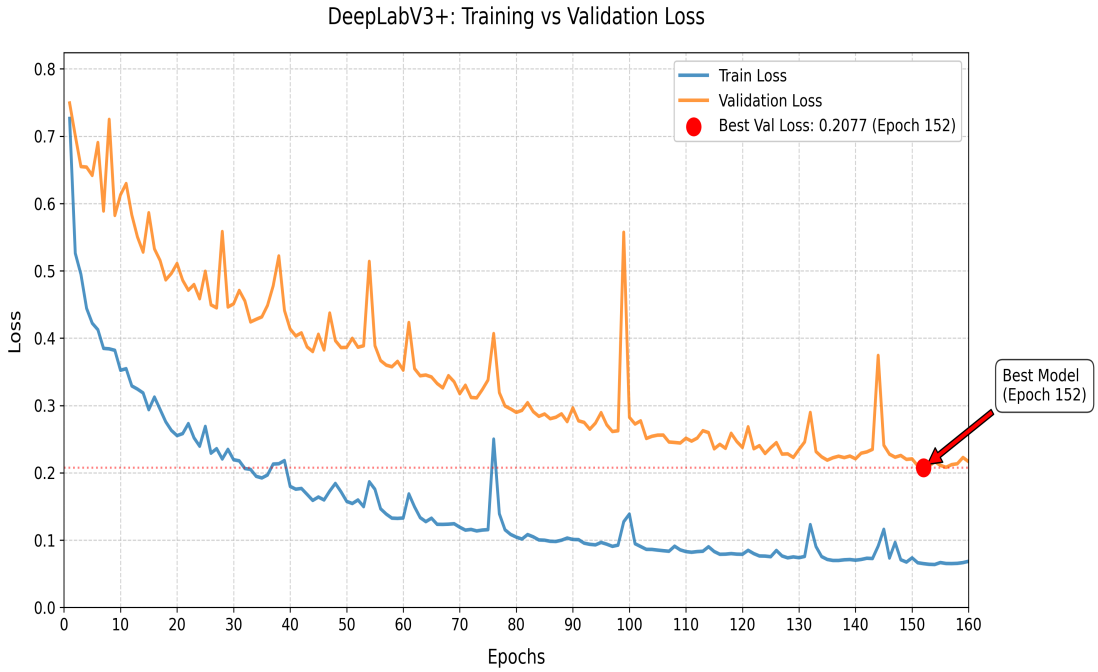


Figure 5.6: Training Vs Validation Loss of DeepLabV3+

5.2 Importance of Dice Loss in Handling Class Imbalance

Glacial lake segmentation is formulated as a binary pixel-wise classification task, where each pixel is classified as either part of a glacial lake or background. In such problems, a significant challenge arises from class imbalance, as the majority of pixels often belong to the background class, while lake pixels are sparse and unevenly distributed.

Initially, the models were trained using only Binary Cross-Entropy (BCE) loss, a standard choice for binary classification tasks. Although BCE loss enabled faster convergence—reaching a plateau after approximately 30 epochs—the segmentation performance was suboptimal. The models tended to favor the dominant background class, leading to poor delineation of lake regions. This behavior indicates that BCE loss alone struggles to guide the model effectively when the dataset is imbalanced.

To address this issue, Dice Loss was introduced alongside BCE loss in the training objective. Dice Loss directly optimizes the overlap between the predicted and ground truth masks, thereby emphasizing the minority class (glacial lakes) and mitigating the effects of imbalance. With the combined BCE-Dice loss, the models converged more gradually, stabilizing around 120 epochs. However, this slower convergence was accompanied by a notable improvement in segmentation accuracy, especially in correctly

identifying smaller and less prominent lake regions.

The inclusion of Dice Loss thus proved essential for enhancing model performance under class-imbalanced conditions. It complements BCE by focusing on spatial overlap, making it a robust choice for segmentation tasks involving sparse target regions such as glacial lakes.

5.3 Prediction in Different Region

Among the four models evaluated, U-Net++ consistently outperformed the others across both input configurations, demonstrating superior performance in terms of precision, recall, and F1-score. Owing to its robustness and deep feature aggregation capabilities, U-Net++ was selected as the primary architecture for further experimentation aimed at evaluating spatial generalization across diverse geomorphic settings.
To test the model's ability to generalize, we trained U-Net++ on data from the Central Himalayas (CH) and Eastern Himalayas (EH) regions, and attempted to predict glacial lakes in the Western Himalayas (WH)—a region not seen during training. However, the model's performance in WH was significantly poor, indicating that regional differences in topography, spectral characteristics, and glacial morphology hinder the model's transferability.
To further validate this observation, we retrained U-Net++ using a dataset that included samples from the WH region. In this revised setup, the model exhibited substantially improved performance, confirming that regional diversity in training data is crucial for achieving robust generalization in glacial lake segmentation tasks.

Figure 5.7: Prediction Region marked in Red Box

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this study, a 9-channel satellite image dataset was constructed by integrating optical Landsat imagery and Sentinel-1 SAR GRD data to support the task of glacial lake segmentation in the Himalayas. Four deep learning models were trained on this dataset, and experiments demonstrated that the inclusion of Sentinel-1 SAR data significantly enhanced the segmentation performance compared to models trained on Landsat-only inputs. This highlights the complementary nature of SAR data in capturing surface characteristics, especially under cloud-prone and complex terrain conditions, which are common in glacial environments.

The dataset was created by patching large satellite swaths into 256×256 image tiles with a stride of 128 pixels. This input format was effective in preserving local spatial context while maintaining computational feasibility. However, experiments with smaller patch sizes revealed that spatial context was insufficient, leading to poor segmentation performance—especially for small and irregular lake shapes. This underlines the importance of appropriate input size in pixel-wise classification tasks involving geospatial features.

When the trained models were applied to a different geological zone—the Western Himalayas—they failed to generalize effectively. This limitation was attributed to domain shift arising from differing terrain, snow cover, and lake morphology. Although incorporating a small subset of training data from the Western Himalayas slightly improved generalization, the performance remained suboptimal, indicating that further regional data diversity is necessary for robust predictions across broader geographical extents.

6.2 Future Work

To build upon the findings of this research, several future directions are proposed:

- **Incorporation of Topographic Data:**
  Integrating Digital Elevation Models (DEM) and their derivatives—such as slope, aspect, and curvature—could provide additional topographic context that helps distinguish glacial lakes from surrounding terrain features, improving segmentation accuracy.

- **Optimizing Input Patch Size and Stride:**
  Future work should explore the effect of varying patch sizes and stride values to strike a balance between spatial context and computational efficiency. Techniques like adaptive patch sampling or hierarchical input strategies may further enhance performance.

- **Domain Adaptation Across Geographical Zones:**
  To address the domain shift across Himalayan regions, techniques such as transfer learning, domain adaptation, or regional fine-tuning can be employed. Collecting a more geographically diverse dataset that includes samples from the Western Himalayas is crucial for improving generalization.

- **Multi-Temporal and Seasonal Analysis:**
  Temporal variations due to seasonal snow cover, meltwater, and vegetation changes can influence glacial lake visibility. Incorporating multi-temporal satellite data could improve robustness and enable monitoring of lake dynamics.

- **Post-Processing and Shape Refinement:**
  Incorporating morphological operations or CRF (Conditional Random Fields)-based post-processing could refine lake boundaries predicted by the segmentation models, especially in regions with fragmented or noisy predictions.

- **Integration with Hydrological and Climate Models:**
  Ultimately, glacial lake segmentation outputs can serve as inputs to broader hydrological models to assess lake expansion risks or potential GLOF (Glacial Lake Outburst Flood) hazards. Integration with climate data could support long-term monitoring frameworks.

Bibliography

[1] Bharat Raj Singh and Onkar Singh. "A Study About Realities of Climate Change: Glacier Melting and Growing Crises". In: *Climate Change*. Ed. by Bharat Raj Singh. Rijeka: IntechOpen, 2013. Chap. 2. DOI: 10.5772/54968. URL: https://doi.org/10.5772/54968.

[2] Alison L. Hardie, Umesh K. Haritashya, and Russell C. Hardie. "Deep Learning and Semantic Segmentation for the Detection and Mapping of Glacial Lakes". In: *NAECON 2024 - IEEE National Aerospace and Electronics Conference*. 2024, pp. 30–33. DOI: 10.1109/NAECON61878.2024.10670689.

[3] Saurabh Kaushik et al. "Automated mapping of glacial lakes using multi-source remote sensing data and deep convolutional neural network". In: *International Journal of Applied Earth Observation and Geoinformation* 115 (2022), p. 103085. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2022.103085. URL: https://www.sciencedirect.com/science/article/pii/S1569843222002734.

[4] Yi He et al. "An Extraction Method for Glacial Lakes Based on Landsat-8 Imagery Using an Improved U-Net Network". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 6544–6558. DOI: 10.1109/JSTARS.2021.3085397.

[5] M. Tom et al. "LAKE ICE DETECTION FROM SENTINEL-1 SAR WITH DEEP LEARNING". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-3-2020 (2020), pp. 409–416. DOI: 10.5194/isprs-annals-V-3-2020-409-2020. URL: https://isprs-annals.copernicus.org/articles/V-3-2020/409/2020/.

[6] Adnan Qayyum et al. "Monitoring supraglacial lakes using PlanetScope data and deep learning". In: *Remote Sensing* 12.13 (2020), p. 2132.

[7] Saurabh Ghosh and Aman Kaushik. "Deep learning and SAR data for water body mapping in glaciated terrain". In: *Remote Sensing Applications: Society and Environment* (2022).

[8] Andreas Kääb et al. “Satellite monitoring of glacier lake outburst floods: advances and challenges”. In: *Nature Reviews Earth & Environment* (2021).

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[10] Zongwei Zhou et al. *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*. 2018. arXiv: 1807.10165 [cs.CV]. URL: https://arxiv.org/abs/1807.10165.

[11] Liang-Chieh Chen et al. “Encoder-decoder with atrous separable convolution for semantic image segmentation”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[12] Jie Wu et al. “Deep Learning-Based Fusion of Sentinel-1 and Sentinel-2 Data for Glacier Lake Mapping”. In: *Remote Sensing* 14.2 (2022), p. 234.

[13] Jingwei He et al. “Glacial lake mapping using Landsat-8 imagery and U-Net deep learning model”. In: *Remote Sensing Letters* 12.2 (2021), pp. 137–146.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[15] Zongwei Zhou et al. *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*. 2018. arXiv: 1807.10165 [cs.CV]. URL: https://arxiv.org/abs/1807.10165.

[16] Ozan Oktay et al. *Attention U-Net: Learning Where to Look for the Pancreas*. 2018. arXiv: 1804.03999 [cs.CV].

[17] Liang-Chieh Chen et al. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. 2018. arXiv: 1802.02611 [cs.CV]. URL: https://arxiv.org/abs/1802.02611.

[18] Ismail Khalfaoui-Hassani, Thomas Pellegrini, and Timothée Masquelier. *Dilated convolution with learnable spacings*. 2023. arXiv: 2112.03740 [cs.CV]. URL: https://arxiv.org/abs/2112.03740.

[19] François Chollet. *Xception: Deep Learning with Depthwise Separable Convolutions*. 2017. arXiv: 1610.02357 [cs.CV]. URL: https://arxiv.org/abs/1610.02357.

[20] Anqi Mao, Mehryar Mohri, and Yutao Zhong. *Cross-Entropy Loss Functions: Theoretical Analysis and Applications*. 2023. arXiv: 2304.07288 [cs.LG].

[21] Carole H. Sudre et al. “Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2017, pp. 240–248. ISBN: 9783319675589. DOI: 10.1007/978-

3-319-67558-9_28. URL: http://dx.doi.org/10.1007/978-3-319-67558-9_28.

[22] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[23] Abien Fred Agarap. *Deep Learning using Rectified Linear Units (ReLU)*. 2019. arXiv: 1803.08375 [cs.NE].

[24] Jeffrey G. Masek et al. "Landsat 9: Empowering open science and applications through continuity". In: *Remote Sensing of Environment* 248 (2020), p. 111968. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2020.111968. URL: https://www.sciencedirect.com/science/article/pii/S0034425720303382.

[25] Jure Leskovec and Rok Sosič. "SNAP: A General-Purpose Network Analysis and Graph-Mining Library". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.1 (2016), p. 1.

[26] John A. Richards. *Remote Sensing Digital Image Analysis*. 5th. Springer, 2006. ISBN: 9783540251286.

[27] Jong-Sen Lee. "Digital image enhancement and noise filtering by use of local statistics". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-2.2 (1980), pp. 165–168.

[28] S.K. McFeeters. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features". In: *International Journal of Remote Sensing* 17.7 (1996), pp. 1425–1432.

[29] Rodrigo Minetto, Neucimar J. Leite, et al. "Hydrographic feature extraction from high-resolution satellite images using mathematical morphology". In: *IEEE Transactions on Geoscience and Remote Sensing* 49.12 (2011), pp. 4729–4741.

[30] Pavel Iakubovskii. *Segmentation Models Pytorch*. https://github.com/qubvel/segmentation_models.pytorch. 2019.