Deep Learning–Based Glacier Semantic Segmentation Using Multispectral Satellite Imagery: Binary and Multi-Class Modeling

Project submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology in Computer Science and Engineering

by

Chinmaya Kumar Behera

Regd.No.: 2202081078

under the Supervision of

Dr. Bighnaraj Naik

Department of Computer Science and Engineering

Veer Surendra Sai University of Technology

Burla, Odisha–768018

[2025–26]

Department of Computer Science and Engineering

Veer Surendra Sai University of Technology, Burla, Odisha

Certificate of Examination

Registration Number: 220208108
Name: Chinmaya kumar behera
Title of the Project: Deep Learning–Based Glacier Semantic Segmentation Using Multispectral Satellite Imagery: Binary and Multi-Class Modeling

This is to certify that the project report entitled **"Deep Learning–Based Glacier Semantic Segmentation Using Multispectral Satellite Imagery: Binary and Multi-Class Modeling"** submitted by **Chinmaya kumar Behera** Regd.no. **2202081078** of **Computer Science and Engineering** under the Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Odisha has been examined by us. We are satisfied with the volume, quality and correctness of the work.

Internal Examiners **External Examiners**

HOD, Dept. of CSE
VSSUT, BURLA

HOP, UG
Dept. of CSE

Department of Computer Science and Engineering

Veer Surendra Sai University of Technology, Burla, Odisha

Supervisor's Certificate

This is to certify that the Project work entitled **"Deep Learning–Based Glacier Semantic Segmentation Using Multispectral Satellite Imagery: Binary and Multi-Class Modeling"** is being submitted by **Chinmaya kumar Behera**(Regd. No.: **2202081078**), to the Department of Computer Science and Engineering, Veer Surendra Sai University of Technology, Burla, in partial fulfillment of the requirement for the degree of Bachelor of Technology in Computer Science and Engineering during the academic year 2025–2026. It is an original report carried out by her under my supervision.

Dr. Bighnaraj Naik
(Supervisor)

Declaration

I hereby declare that the project titled **"Deep Learning–Based Glacier Semantic Segmentation Using Multispectral Satellite Imagery: Binary and Multi-Class Modeling"** submitted to Veer Surendra Sai University of Technology for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a result of original work carried out in this report. I also declare that the work has not been submitted, in whole or in part, to any other university as an exercise for a degree or any other qualification.

Date:
VSSUT, BURLA

Chinmaya kumar behera
Regd.No.:22020801078

Acknowledgement

I would like to thank my supervisor, **Dr. Bighnaraj Naik** for giving us the opportunity to undertake this project. I am extremely grateful for her insightful guidance, timely advice, continuous support, and motivation throughout my project work. I especially thank her for her kindness and excellent contributions.

I would like to thank the HOD, Department of Computer Science and Engineering for his support to carry out my project work. I would also like to thank all the faculty and staff members of the Department of Computer Science and Engineering at Veer Surendra Sai University of Technology, Burla for their help and support.

I am deeply thankful to the Vice-Chancellor for his moral support and for providing us with all the necessary facilities during our candidature.

I would like to thank our friends and family members for their unconditional love and support.

Abstract

Glaciers form an essential part of the Earth's natural systems, influencing climate, water availability, and the stability of high–altitude environments. In recent years, many glacier regions have been shrinking at an alarming rate, which has raised concerns about long–term water supplies and the growing risk of glacial lake outburst floods. Tracking these changes accurately is important, but traditional mapping methods depend heavily on manual effort and are often too slow to handle large or frequently updated satellite datasets.

This project explores the use of deep learning to automatically identify glacier regions from multispectral satellite images. Following the two phases of the GlacierHack 2025 challenge, the work begins with a **binary segmentation** model based on a U-Net architecture to separate glacier areas from non–glacier terrain. It then extends to a **four–class segmentation** model using a U-Net with a ResNet–34 encoder to distinguish background, clean ice, debris–covered glacier, and glacial lakes. The models were trained on 125 satellite tiles, and an 80:20 train–validation split was used. Since some classes, especially glacial lakes, were extremely rare, techniques such as weighted loss functions, patch sampling, and Dice/Tversky–based optimization were applied to improve learning.

To make the system practical and easy to use, both models were integrated into an **interactive dashboard** that allows users to upload TIFF images and view the predicted outputs directly. Rather than remaining a purely experimental study, the project offers a working tool that can support students, researchers, and environmental groups interested in glacier monitoring. The outcomes suggest that deep learning can play a meaningful role in speeding up glacier analysis and may help pave the way for future work involving larger datasets, time–based change detection, and cloud deployment.

Keywords: Glacier Segmentation; U-Net; ResNet-34; Multispectral Imagery; Semantic Segmentation; Glacial Lakes; Deep Learning; Dashboard.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	: Artificial Intelligence
ML	: Machine Learning
DB	: Database
ORM	: Object Relational Mapping
UI	: User Interface
API	: Application Programming Interface
CRUD	: Create, Read, Update, Delete
CSE	: Computer Science & Engineering

1. Introduction

Glaciers are among the most significant natural assets on Earth, acting as reservoirs of freshwater and serving as sensitive indicators of climate change. Over the past few decades, rapid temperature variations and environmental shifts have accelerated glacier melting, making it increasingly important to monitor glacier boundaries, surface conditions, and long-term changes. Reliable information about glacier extent is essential not only for scientific research, but also for water resource management, disaster prediction, and policy planning.

Traditionally, glacier mapping has relied on manual interpretation of satellite images or physically demanding field surveys. While valuable, these methods are often slow, subjective, and limited in coverage—especially in remote or high-risk terrains. As glaciers continue to evolve rapidly, there is a growing need for automated, scalable, and accurate techniques that can analyze large volumes of satellite data without human intervention.

The rise of multispectral satellite imagery has opened new possibilities by providing rich pixel-level information across different spectral bands. However, extracting meaningful insights from this data is challenging due to complex terrain conditions, debris-covered regions, varying illumination, and the subtle visual similarities between glacier surfaces and surrounding landscapes. Conventional image processing techniques often struggle to handle these complexities effectively.

Deep learning, particularly semantic segmentation, offers a powerful solution to this problem. Instead of simply identifying glaciers as a whole, semantic segmentation models classify every pixel in an image, enabling precise boundary detection and a deeper understanding of glacial features. Such pixel-level analysis is crucial for accurate mapping, change detection, and environmental monitoring.

This project explores deep learning-based glacier segmentation through two phases. Phase 1 focuses on binary classification—distinguishing glacier from non-glacier regions—while Phase 2 extends the problem to a more detailed four-class segmentation. By leveraging multispectral data and advanced model architectures, the project aims to create a scalable system that improves accuracy, reduces manual effort, and makes glacier analysis more accessible. Additionally, an interactive dashboard has been developed to allow users to upload satellite tiles and visualize model predictions in real time, supporting practical use and interpretation.

Overall, this work demonstrates how modern AI techniques can support climate research and environmental monitoring by providing faster, more detailed, and data-driven glacier segmentation capabilities.

1.1 Background and Context

Glaciers play a crucial role in maintaining the Earth's environmental balance. They store vast amounts of freshwater, regulate sea levels, and influence local and global climate systems. Because of their sensitivity to temperature changes, glaciers are often considered one of the clearest indicators of climate change. Even small shifts in temperature can lead to noticeable variations in glacier size and structure, making continuous monitoring essential for understanding long-term environmental trends.

Over recent decades, many glacier-rich regions around the world have experienced rapid melting and retreat. This has raised serious concerns for downstream communities that depend on glacier-fed water for agriculture, drinking water, and hydropower. In some areas, melting glaciers have also increased the risk of natural disasters such as floods and glacial lake outburst events. As a result, researchers, environmental agencies, and policymakers require reliable and timely information about glacier extent and behavior.

Traditionally, glacier mapping relied on manual field surveys or expert interpretation of satellite images. While these methods can provide accurate insights, they are often time-consuming, physically challenging, and limited in scope. Remote and hazardous terrains make ground surveys difficult, and manual image interpretation can introduce subjectivity and inconsistency—especially when dealing with large geographical areas.

With the increasing availability of satellite imagery, especially multispectral data, it has become possible to observe glaciers frequently and in greater detail. Multispectral images capture information across different spectral bands, revealing patterns that are not always visible to the human eye. However, the sheer volume and complexity of this data make manual analysis impractical. Traditional image processing methods also struggle to differentiate between visually similar surfaces, such as debris-covered ice and surrounding rocky terrain.

In recent years, deep learning has emerged as a powerful approach for image understanding. Semantic segmentation, in particular, enables pixel-by-pixel classification of satellite imagery, offering precise boundary detection and detailed surface characterization. This capability is especially valuable for glacier monitoring, where subtle differences in terrain can significantly impact environmental conclusions.

By combining multispectral satellite data with deep learning techniques, this project aims to support more accurate, scalable, and automated glacier segmentation. Such advancements can help researchers and decision-makers track glacier changes more effectively, respond to environmental risks, and gain a deeper understanding of how glacial landscapes are evolving over time.

1.2 Problem Statement – Phase 1 (PS1)

Monitoring glaciers at a large scale requires accurate identification of glacier-covered regions. However, most existing approaches depend heavily on manual interpretation of satellite images or basic remote sensing techniques. These methods are slow, labor-intensive, and often inconsistent, especially when different experts interpret the same data in varying ways. As glaciers continue to change rapidly due to climate fluctuations, relying solely on manual mapping makes it difficult to keep information updated and reliable.

Traditional image processing methods also face significant challenges when applied to complex terrains. Variations in lighting, shadows, snow cover, and rocky surfaces often cause confusion between glacier and non-glacier areas. In many cases, clean glacier ice blends visually with surrounding land, making it difficult to draw accurate boundaries using conventional techniques. As a result, glacier maps generated through these methods may lack precision and fail to capture subtle but important changes.

Phase 1 of this project focuses on a binary segmentation problem—classifying each pixel in a satellite image as either **glacier** or **non-glacier**. While this may appear straightforward, it presents several practical challenges:

- large variations in glacier appearance across different regions,
- inconsistent image conditions due to weather and seasonal changes,
- difficulty in separating glaciers from visually similar terrain, and
- the need to process a high volume of data efficiently.

Without an automated and scalable method, researchers and environmental agencies struggle to generate timely and accurate glacier maps. This limitation slows down climate analysis, risk assessment, and policy planning. Therefore, Phase 1 aims to develop a deep learning-based solution capable of performing reliable pixel-level binary segmentation, reducing manual effort and providing a foundation for more detailed glacier analysis in the subsequent phase.

1.3 Problem Statement – Phase 2 (PS2)

While binary segmentation provides a basic separation between glacier and non-glacier regions, it does not offer the level of detail needed for deeper environmental analysis. In reality, glaciers consist of multiple surface types, including clean ice, debris-covered ice, surrounding land, and glacial water bodies. Treating all glacier-related regions as a single class can lead to oversimplified maps and incomplete insights, especially when studying melting patterns, hazard risks, or long-term glacier behavior.

One of the major challenges in glacier research is the presence of debris-covered ice. Because debris can visually resemble nearby rocky terrain, it often gets misclassified as non-glacier in traditional methods. Similarly, glacial lakes—which can be early indicators of potential flood hazards—occupy very small areas and are often overlooked. As a result, many conventional approaches fail to identify these critical subregions accurately.

Phase 2 of this project extends the problem to a four-class semantic segmentation task, where each pixel must be classified into one of the following categories:

- **Class 0:** Background / Non-Glacier
- **Class 1:** Clean Glacier Ice
- **Class 2:** Debris-Covered Glacier Ice
- **Class 3:** Glacial Water Bodies

This multi-class setup introduces new difficulties:

- **High class imbalance**, especially for water bodies which occupy very few pixels.
- **Spectral similarity** between debris-covered ice and surrounding land.
- **Small and fragmented regions** that are easily missed during prediction.
- **Increased model complexity** requiring more robust training strategies.

Without a reliable multi-class segmentation system, researchers lack the detailed information needed to understand how different glacier components are changing over time. Therefore, Phase 2 aims to develop a deep learning model capable of accurate, fine-grained classification, enabling more meaningful interpretation of glacier dynamics and supporting climate-related decision-making.

1.4 Motivation and Need

Glaciers are not just frozen landscapes—they are vital components of the Earth's climate system and a primary source of freshwater for millions of people. As global temperatures continue to rise, glaciers are retreating at an alarming pace. This rapid decline has raised concerns about water scarcity, sea-level rise, changes in river flow, and the increased likelihood of glacial lake outburst floods. To manage these risks and plan for the future, reliable and frequent monitoring of glacier changes has become more important than ever.

Despite the availability of satellite imagery, many regions still rely on manual mapping and basic image interpretation. These approaches are slow, difficult to scale, and dependent on expert judgment. In fast-changing environments, such delays can lead to outdated assessments and missed warning signs. Remote and hazardous terrains further limit field-based surveys, making automated solutions not just convenient, but necessary.

Modern multispectral satellite data provides a rich source of information that, if analyzed correctly, can reveal subtle differences on glacier surfaces. However, the sheer volume and complexity of this data make manual analysis impractical. Conventional image processing techniques often fail to distinguish between visually similar features—such as debris-covered ice and rocky terrain—resulting in inaccurate boundaries and misleading conclusions.

Deep learning offers a promising way forward. By learning patterns directly from data, deep learning models can analyze imagery at a pixel level and adapt to complex variations in terrain, lighting, and surface texture. Automated segmentation not only reduces manual effort but also provides consistent and repeatable results, which are crucial for long-term monitoring and scientific comparison.

In addition to the technical need, there is also a practical motivation. Students, researchers, and environmental agencies require accessible tools that allow them to visualize, test, and interpret glacier segmentation results without extensive technical expertise. An interactive dashboard bridges this gap by making model predictions usable and interpretable, encouraging wider adoption of AI-driven glacier analysis.

Overall, the motivation behind this project is to create an accurate, scalable, and user-friendly system that supports real-time glacier monitoring and contributes to environmental research, disaster preparedness, and informed decision-making in a rapidly changing climate.

1.5 Objectives of the Project

The primary aim of this project is to develop an automated, accurate, and scalable system for glacier segmentation using multispectral satellite imagery. To achieve this goal, the project is structured into two progressive phases, each addressing a specific level of segmentation complexity. Beyond model development, the project also focuses on practical usability through an interactive dashboard that enables real-time testing and visualization.

The key objectives of the project are as follows:

- **To automate glacier identification** by eliminating the need for manual interpretation and reducing dependence on time-consuming field-based surveys.
- **To develop a binary deep learning model (Phase 1)** capable of accurately classifying each pixel as glacier or non-glacier.
- **To extend segmentation to a four-class model (Phase 2)** that differentiates between background land, clean glacier ice, debris-covered ice, and glacial water bodies.
- **To effectively utilize multispectral satellite data** by leveraging multiple spectral bands for improved segmentation accuracy and feature discrimination.
- **To address class imbalance challenges** in the multi-class model using techniques such as weighted loss functions, sampling strategies, and targeted patch extraction.

- **To evaluate model performance using robust metrics** such as Accuracy, IoU, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC).
- **To build an interactive dashboard** that allows users to upload .tif images, run inference, and visualize segmentation outputs in a user-friendly manner.
- **To create a scalable workflow** that can be adapted for future datasets, additional classes, or extended environmental applications.

Together, these objectives aim to bridge the gap between advanced deep learning research and practical, real-world glacier monitoring tools, enabling faster analysis, improved accuracy, and broader accessibility for researchers and environmental stakeholders.

1.6 Scope and Limitations

This project focuses on automating glacier segmentation using multispectral satellite imagery and deep learning techniques. The work is designed to demonstrate how AI can assist in large-scale glacier monitoring by providing pixel-level classification and reducing the dependence on manual mapping. The scope of the project is divided into two major phases—binary segmentation and multi-class segmentation—along with the development of an interactive dashboard for real-time testing and visualization.

Scope of the Project:

- Utilization of multispectral satellite tiles provided by the GlacierHack 2025 dataset.
- Development of a binary segmentation model (Phase 1) to classify glacier vs. non-glacier pixels.
- Development of a four-class segmentation model (Phase 2) to identify background, clean ice, debris-covered ice, and water bodies.
- Use of U-Net based deep learning architecture for both segmentation tasks.
- Training, validation, and performance evaluation using defined metrics such as IoU, Accuracy, F1-Score, and MCC.
- Creation of an interactive dashboard that allows users to upload .tif images and visualize model outputs.
- Focus on pixel-level segmentation rather than volume estimation or temporal change prediction.

Limitations of the Project:

- The dataset size is limited to the 125 tiles provided by the hackathon, which may not represent all glacier regions globally.
- Class imbalance in the four-class model—especially for water bodies—may affect precision and recall despite mitigation techniques.
- The model focuses on static imagery and does not analyze time-based glacier changes or movement patterns.
- Results depend on the spectral properties of the given dataset; performance may vary on other satellites or imaging conditions without retraining.
- Terrain factors such as shadows, cloud cover, and seasonal variations may still introduce misclassifications.
- The dashboard supports inference and visualization but does not include advanced analytics or predictive forecasting.
- The project does not estimate glacier thickness, volume, or hydrological impact; it is limited to surface-level segmentation.

Overall, the project establishes a practical and scalable foundation for automated glacier segmentation while acknowledging that further enhancements—such as larger datasets, multi-temporal analysis, and advanced post-processing—would be required for broader operational deployment.

2. Literature Review

Understanding glacier dynamics and their accurate mapping has been an active area of research for several decades. As glaciers continue to retreat due to climate change, the scientific community has placed increasing emphasis on reliable monitoring methods that can capture changes in glacier extent, surface characteristics, and associated environmental risks. A wide range of approaches—ranging from ground-based surveys to automated computational models—have been explored to address this need.

The introduction of satellite-based remote sensing marked a major turning point in glacier observation. Instead of relying solely on field measurements, researchers gained the ability to observe large and remote regions at regular intervals. Multispectral imagery, in particular, enabled more detailed surface analysis by capturing information across multiple wavelength bands. Over time, this led to the development of various techniques for glacier detection and classification.

Early glacier mapping methods were largely manual or based on simple thresholding techniques. Although useful, these approaches were limited by subjectivity, environmental variability, and the difficulty of distinguishing visually similar surfaces. As computational methods evolved, machine learning algorithms—such as Support Vector Machines and Random Forests—were introduced to improve classification accuracy by learning from handcrafted features.

In recent years, deep learning has emerged as a powerful alternative due to its ability to learn complex patterns directly from data. Semantic segmentation models, especially those based on convolutional neural networks, have demonstrated significant improvements in pixel-level image classification tasks across various remote sensing applications. Architectures like U-Net have played a key role in advancing glacier segmentation research by offering precise boundary detection and robust performance even with limited training data.

This chapter reviews the existing body of research related to glacier monitoring, remote sensing technologies, traditional and modern mapping techniques, and the evolution of machine learning and deep learning approaches. By examining these developments, the chapter highlights the strengths and limitations of current methods and identifies the research gaps that motivate the present project.

2.1 Glaciers and Climate Significance

Glaciers are large, persistent bodies of ice that form over long periods through the accumulation and compaction of snow. Although they cover only a small percentage of the Earth's surface, they hold nearly 70% of the planet's freshwater reserves. This makes glaciers one of the most important natural resources for sustaining rivers, lakes, and groundwater systems, particularly in mountainous and polar regions where communities depend on meltwater for drinking water, agriculture, and hydropower.

Beyond their role as freshwater reservoirs, glaciers also influence global climate systems. Their highly reflective surfaces help regulate Earth's temperature by reflecting solar radiation back into the atmosphere. As glaciers shrink, darker land or water surfaces are exposed, absorbing more heat and further accelerating warming—a feedback loop that contributes to climate change. This connection highlights the importance of glacier monitoring as part of broader climate research.

In recent decades, scientific observations have documented widespread glacier retreat across many regions, including the Himalayas, the Alps, Greenland, and Antarctica. Rising global temperatures have caused glaciers to melt more rapidly than they can accumulate new ice. This has contributed to rising sea levels and increased the risk of flooding in coastal and low-lying regions. In high-altitude areas, the formation of expanding glacial lakes has raised concerns about glacial lake outburst floods, which can cause sudden and severe damage downstream.

Glacier changes also have ecological impacts. Altered water availability affects freshwater habitats, disrupts seasonal river flow patterns, and influences biodiversity. For communities that rely on glacier-fed water, long-term retreat could lead to water shortages, reduced agricultural productivity, and economic hardship. These far-reaching consequences underscore the need for accurate and continuous glacier monitoring.

Because glaciers respond quickly to changes in temperature and precipitation, they serve as early indicators of environmental change. Tracking their size, shape, and behavior provides valuable insight into ongoing climate trends and helps scientists assess future risks. As a result, reliable glacier mapping and segmentation have become essential tools for climate research, environmental planning, and disaster preparedness.

2.2 Remote Sensing and Multispectral Imagery

Remote sensing refers to the process of collecting information about the Earth's surface without direct physical contact. Instead of field-based measurements, remote sensing relies on sensors mounted on satellites, aircraft, or drones to observe large and often inaccessible regions. This capability has transformed environmental monitoring by enabling continuous and large-scale data acquisition across diverse terrains and climatic zones.

Satellite-based remote sensing has played a particularly important role in glacier research. Glacial regions are frequently located in harsh, remote, and high-altitude locations, making ground surveys difficult, dangerous, and sometimes impossible. By capturing imagery at regular intervals, satellites allow researchers to monitor glacier changes over time, track seasonal variations, and identify emerging risks such as expanding glacial lakes.

Multispectral imagery is one of the most widely used forms of satellite data in glacier studies. Unlike standard RGB images that capture only three visible light channels, multispectral images record information across multiple wavelength bands, including visible, near-infrared, and shortwave infrared regions. Each band highlights different surface properties, allowing subtle variations in material composition, moisture content, and reflectance to be detected.

For glacier segmentation, multispectral imagery provides significant advantages. Clean ice, debris-covered ice, rock surfaces, vegetation, and water bodies each reflect light differently across spectral bands. While some of these differences may be visually indistinguishable in standard images, multispectral data captures patterns that can be used to differentiate between surface types more accurately. This is especially valuable in areas where debris-covered ice visually resembles surrounding terrain.

Another key strength of remote sensing is temporal coverage. Many modern satellites collect data at frequent intervals, enabling time-series analysis and long-term monitoring. This helps researchers study glacier retreat rates, seasonal melt cycles, and the formation of new glacial lakes. In addition, remote sensing provides consistent, standardized data across large spatial scales, making it possible to compare glacier behavior across different regions.

Despite its advantages, multispectral imagery also introduces challenges. Variations in lighting, shadows, cloud cover, and atmospheric conditions can affect image quality. The high dimensionality of multispectral data makes manual interpretation difficult, and traditional threshold-based techniques often struggle to capture complex surface patterns. These challenges have motivated the shift towards machine learning and deep learning approaches that can automatically learn and extract meaningful features from multispectral inputs.

2.3 Traditional Glacier Mapping Techniques

Before the rise of advanced computational methods, glacier mapping primarily relied on manual and rule-based approaches. These traditional techniques played an important role in early glaciological studies and laid the foundation for modern remote sensing research. However, they also carried significant limitations in terms of scalability, accuracy, and consistency, especially when applied to large or complex terrains.

One of the earliest approaches involved **field-based surveys**, where researchers physically visited glacier regions to record boundaries, measure ice thickness, and observe surface conditions. Although field surveys provided valuable ground truth data, they were physically demanding, time-consuming, and often restricted to small areas. Harsh weather, difficult terrain, and safety risks further limited the feasibility of conducting frequent on-site measurements.

As satellite imagery became available, glacier mapping shifted toward **manual visual interpretation**. Experts analyzed satellite images or aerial photographs and manually traced glacier outlines. While this method improved spatial coverage, it still depended heavily on human judgment. Different analysts could produce varying results for the same region, making reproducibility a challenge. Moreover, manual interpretation required substantial expertise and was impractical for large datasets or frequent monitoring.

In an effort to automate the process, researchers introduced **threshold-based and rule-based image processing techniques**. These methods typically relied on spectral indices, such as the

Normalized Difference Snow Index (NDSI), to distinguish between snow, ice, and other surfaces. Pixels above a certain threshold were classified as glacier, while those below were categorized as non-glacier. Although simple and computationally efficient, threshold-based methods were highly sensitive to illumination, shadows, seasonal variations, and debris cover. In many cases, debris-covered ice appeared similar to surrounding rock, leading to frequent misclassifications.

Other traditional approaches included edge detection, band ratioing, texture analysis, and supervised classification using manually selected features. However, these techniques struggled to adapt to the complex spectral and spatial variability present in glacial environments. They often required significant parameter tuning and expert intervention, limiting their effectiveness in large-scale or automated workflows.

Overall, while traditional mapping methods contributed valuable insights to early glacier research, they lacked the precision, scalability, and automation needed for modern climate monitoring. These limitations paved the way for machine learning and, later, deep learning approaches that could learn patterns directly from data and handle complex terrain conditions more effectively.

2.4 Machine Learning in Glacier Analysis

As the volume and availability of satellite imagery increased, researchers began exploring ways to move beyond manual and rule-based techniques. This shift marked the entry of **Machine Learning (ML)** into glacier analysis. Unlike traditional methods that relied on fixed thresholds or handcrafted rules, ML techniques sought to learn patterns directly from data, making them more adaptable to real-world variations.

Early machine learning approaches typically focused on **supervised classification**. Algorithms such as **Support Vector Machines (SVM)**, **Decision Trees**, and **k-Nearest Neighbors (k-NN)** were trained on labeled samples to distinguish glacier pixels from non-glacier regions. These models performed well in clean, snow-covered glacier areas where spectral signatures were clear. However, their performance dropped significantly when dealing with debris-covered ice, shadows, or mixed terrains.

To improve generalization, some studies adopted **unsupervised methods**, including **k-means clustering** and **ISODATA**, to group pixels based on similarities without requiring labeled data. While these clustering techniques helped automate initial segmentation, they still required manual interpretation to assign cluster meanings and often struggled with overlapping spectral characteristics.

Feature engineering played a central role in traditional ML-based glacier studies. Researchers designed handcrafted features such as spectral ratios, texture descriptors, and elevation data to help models separate glacier surfaces from surrounding terrain. Although feature-based methods provided better results than simple thresholding, they remained highly dependent on expert knowledge. Small changes in illumination or environmental conditions could still lead to inconsistent outcomes.

Moreover, most ML approaches treated glacier identification as a **pixel-wise classification problem** rather than a spatially aware task. They often ignored contextual relationships between neighboring pixels, which are crucial in landform mapping. As a result, outputs sometimes appeared noisy or fragmented, requiring additional post-processing to refine glacier boundaries.

Despite these challenges, machine learning marked an important transition toward automation and data-driven analysis. It demonstrated that computational models could extract meaningful patterns from satellite imagery and reduce manual effort. However, the limitations in handling complex surfaces, debris-covered ice, and large-scale segmentation highlighted the need for more powerful techniques—eventually leading to the rise of **deep learning** and fully automated semantic segmentation approaches.

2.5 Deep Learning and Semantic Segmentation

The limitations of traditional and classical machine learning methods led researchers to explore more advanced techniques capable of understanding complex patterns in glacier imagery. This shift introduced **Deep Learning (DL)**, a subset of artificial intelligence that uses multi-layer neural networks to learn hierarchical features directly from data. Unlike earlier methods that relied heavily on handcrafted rules or manually designed features, deep learning models automatically extract relevant information, making them far more robust in challenging real-world environments.

A major breakthrough came with **Convolutional Neural Networks (CNNs)**, which proved highly effective in processing visual data. CNNs learn spatial, textural, and spectral patterns from images, enabling them to differentiate between snow, ice, debris, water bodies, and surrounding terrain more accurately than traditional approaches. Their ability to learn from raw pixel data significantly reduced the need for expert-driven feature engineering.

As remote sensing applications grew more demanding, researchers moved beyond simple classification to more detailed pixel-level understanding. This led to the rise of **semantic segmentation**—a deep learning approach where every pixel in an image is assigned a class label. For glacier studies, semantic segmentation allows precise mapping of glacier boundaries, debris-covered areas, and meltwater lakes, offering a level of detail that earlier methods could not achieve.

Early CNN-based segmentation models, such as Fully Convolutional Networks (FCNs), showed promising results but often struggled with spatial resolution loss during downsampling. To address this, architectures like **U-Net**, **SegNet**, and **DeepLab** introduced encoder–decoder structures and skip connections, allowing models to retain fine-grained spatial information. These innovations proved especially valuable in glacier environments, where subtle variations in texture and shape can play a critical role in accurate classification.

Another key advantage of deep learning is its ability to handle **multispectral and high-dimensional data**. Models can be trained on multiple spectral bands—beyond the visible range—to capture deeper insights about snow, ice composition, and surface characteristics. This makes deep learning particularly suitable for satellite-based glacier monitoring.

Overall, deep learning and semantic segmentation have transformed glacier analysis by offering higher accuracy, better generalization, and fully automated workflows. These advancements enable more frequent and large-scale monitoring, supporting climate research, disaster preparedness, and long-term environmental planning. As datasets continue to grow and computational tools become more accessible, deep learning is poised to remain a central technology in future glacier mapping and climate analytics.

2.6 U-Net and Related Architectures

As semantic segmentation gained prominence in remote sensing and medical imaging, researchers recognized the need for models that could accurately identify objects at the pixel level while preserving fine spatial details. One of the most influential breakthroughs in this direction was the **U-Net** architecture, introduced by Ronneberger et al. in 2015. Although originally designed for biomedical image segmentation, U-Net quickly became a widely adopted model across numerous domains, including glacier mapping and environmental monitoring.

The strength of U-Net lies in its **encoder–decoder structure**. The encoder progressively downsamples the input to extract high-level features, while the decoder upsamples these features to reconstruct a full-resolution segmentation map. What sets U-Net apart is the use of **skip connections**, which directly transfer feature maps from the encoder layers to the corresponding decoder layers. This mechanism preserves important spatial information that might otherwise be lost during downsampling, enabling the network to detect small, thin, or irregular structures—such as glacier boundaries or debris-covered regions.

Over time, several enhanced variants of U-Net emerged to improve performance, efficiency, and adaptability. For example, **U-Net++** introduced nested skip pathways to better fuse multi-scale information, while **Attention U-Net** incorporated attention gates to help the model focus on the most relevant regions of an image. These modifications proved especially valuable in remote sensing, where subtle spectral or textural differences can determine whether a pixel belongs to clean ice, debris-covered ice, snow, or surrounding terrain.

Beyond the U-Net family, other architectures have also contributed to advancements in semantic segmentation. **SegNet** introduced an encoder–decoder framework optimized for efficient inference, making it suitable for large-scale satellite imagery. **DeepLab** and its later versions (DeepLabV2, V3, and V3+) leveraged atrous convolutions and conditional random fields to capture multi-scale context and refine segmentation outputs. Similarly, **Fully Convolutional Networks (FCNs)** pioneered the concept of end-to-end pixel-wise prediction, laying the foundation for all modern segmentation models.

Despite these alternatives, U-Net remains a preferred choice for glacier studies due to its **simplicity, strong performance on limited datasets, and ability to handle fine-grained structures**. In many real-world scenarios, glacier datasets are relatively small, imbalanced, or difficult to annotate, making U-Net's data efficiency and architectural flexibility particularly advantageous. Furthermore, its compatibility with multispectral inputs allows researchers to integrate additional spectral bands—beyond standard RGB imagery—to improve the distinction between glacier classes.

In summary, U-Net and its related architectures have played a central role in advancing glacier segmentation by enabling accurate, high-resolution mapping from satellite imagery. Their design principles continue to influence modern deep learning research and remain foundational to state-of-the-art approaches in environmental and climate-related image analysis.

2.7 Research Gaps Identified

Although significant progress has been made in glacier mapping and semantic segmentation, several critical gaps continue to limit the accuracy, scalability, and real-world applicability of existing approaches. These gaps highlight the need for more robust, data-efficient, and class-aware models tailored specifically for glacier environments.

1. Limited Pixel-Level Accuracy for Complex Glacier Surfaces:
Many traditional and early machine learning methods struggle to distinguish between visually similar surface types, such as debris-covered ice and surrounding rocky terrain. Even deep learning models often misclassify thin, fragmented, or irregular glacier boundaries, leading to inaccurate area estimates and misleading trend analyses.

2. Underrepresentation of Minority Classes:
In most glacier datasets, certain classes—especially **glacial lakes** or **thin debris-covered ice**—occupy only a tiny portion of the imagery. This imbalance causes models to favor majority classes while ignoring or incorrectly predicting smaller, yet critical, regions. As a result, minority classes receive low recall and precision, reducing the model's reliability in risk-sensitive applications such as lake-outburst flood analysis.

3. Limited Use of Multispectral Information:
While modern satellites provide multiple spectral bands, many studies continue to rely on RGB or grayscale imagery. This restricts the model's ability to leverage spectral differences that could improve class separations, particularly between snow, ice, water, and debris. The potential of multispectral fusion for glacier segmentation remains under-explored.

4. Insufficient Evaluation Metrics for Imbalanced Datasets:
A large number of studies depend primarily on overall accuracy or pixel-wise correctness, which can be misleading when one class dominates the dataset. Metrics such as Intersection over Union (IoU) or Matthews Correlation Coefficient (MCC) are rarely emphasized, despite being more suitable for assessing imbalanced class performance—especially for rare glacier features.

5. Lack of Practical, Deployable Solutions:
Most research stops at model development and does not extend to real-world deployment. There is a noticeable gap in interactive tools, dashboards, or automated systems that allow scientists, policymakers, or disaster management teams to visualize predictions, monitor glacier changes, or perform rapid assessments in the field.

6. Limited Adaptability Across Regions and Datasets:
Glacier characteristics vary significantly across geographic locations, seasons, and climatic conditions. Models trained on one dataset often fail to generalize to new regions without extensive retraining. This lack of transferability limits large-scale or global glacier monitoring initiatives.

In summary, the existing body of research reveals clear opportunities for advancement. There is a strong need for segmentation models that are capable of handling class imbalance, exploiting multispectral data, supporting reliable evaluation metrics, and ultimately offering deployable, user-friendly solutions for real-world glacier monitoring and climate research.

3. Proposed Method / Model

This chapter presents the methodology adopted to address the two problem statements defined in the GlacierHack 2025 challenge. The core objective of the project was to develop a deep learning–based system capable of accurately segmenting glaciers from multispectral satellite imagery, first in a **binary setting** (glacier vs. non-glacier) and later in a more challenging **multi-class setting** involving four distinct glacier-related classes.

To achieve this, two separate U-Net-based semantic segmentation models were designed, trained, evaluated, and compared. Both models were supported by a structured dataset containing 125 satellite tiles, provided as part of the hackathon. The first model, aligned with Phase 1 (PS1), focuses on binary segmentation, enabling clear identification of glacier regions. The second model, aligned with Phase 2 (PS2), extends this capability to classify each pixel into one of four classes, addressing class imbalance and spectral complexity.

In addition to model development, the project includes the creation of a **web-based dashboard** that allows users to upload TIFF images, run inference using the trained models, and visualize segmentation outputs in real time. This deployable interface bridges the gap between research and practical usability, making the system accessible to researchers, students, and field practitioners.

The methodology followed a systematic workflow that included dataset exploration, preprocessing, model design, training configuration, metric selection, and performance evaluation. The Matthews Correlation Coefficient (MCC) was selected as the primary evaluation metric due to its robustness in imbalanced class scenarios—a key challenge in glacier segmentation.

This chapter is organized into two major solution blocks:

- **Section 3.1: Solution to PS1 (Binary Segmentation)**
  Development of a U-Net model to classify pixels as glacier or non-glacier, including dataset preparation, preprocessing, architecture, training setup, MCC evaluation, and output analysis.
- **Section 3.2: Solution to PS2 (Four-Class Segmentation)**
  Extension of the approach to multi-class segmentation, addressing class imbalance, weighted loss functions, sampler strategies, patch-based augmentation, and performance improvements.

Together, these models form a complete solution that progresses from fundamental glacier detection to detailed glacier surface characterization, supporting more informed climate monitoring and environmental research.

3.1 Solution to PS1 – Binary Glacier Segmentation

3.1.1 Overview of PS1

Phase 1 (PS1) of the project focused on the most fundamental task in glacier analysis: identifying whether a pixel in a satellite image belongs to a glacier or not. This binary segmentation step serves as the foundation for more advanced classification and monitoring tasks, making it an essential starting point for automated glacier mapping.

The challenge arises from the fact that glacier surfaces often share visual similarities with surrounding terrain. Snow-covered rocks, frozen lakes, cloud patches, and bright non-glacial surfaces can appear almost identical in satellite imagery, making manual interpretation time-consuming and error-prone. PS1 aimed to eliminate this bottleneck by creating a model capable of automatically separating glacier regions from the rest of the landscape.

In this phase, the problem was formulated as a two-class semantic segmentation task with the following labels:

- **Class 0: Non-Glacier**
- **Class 1: Glacier**

Each pixel in the input image was assigned to one of these two classes, resulting in a binary mask that clearly highlights glacier boundaries. This simplified structure allowed the model to learn core spatial patterns, spectral characteristics, and edge transitions relevant to glacier detection.

PS1 served two major purposes in the overall project workflow:

- **Baseline Understanding:** Establishing a strong binary classifier to confirm whether deep learning could reliably detect glaciers in the provided multispectral imagery.
- **Foundation for PS2:** Providing insights into model behavior, data characteristics, and segmentation challenges that would later guide the development of the more complex four-class model in Phase 2.

By successfully completing PS1, the project validated that deep learning—specifically a U-Net based architecture—could accurately separate glacier regions, setting the stage for more detailed multi-class segmentation in the subsequent phase.

3.1.2 Dataset and Image Tiles

The dataset used for Phase 1 (PS1) was provided as part of the GlacierHack 2025 challenge. It consisted of a curated collection of satellite image tiles representing glacier-dominated regions.

In total, **125 image tiles** were supplied, each paired with a corresponding binary mask indicating glacier and non-glacier areas. This structured dataset formed the foundation for training and validating the binary segmentation model.

Each image tile captured a spatial segment of the glacier landscape, allowing the model to learn patterns across multiple terrains, lighting conditions and surface textures. The binary masks served as ground truth labels, where glacier pixels were marked as **Class 1** and non-glacier pixels were marked as **Class 0**. This pixel-wise labeling enabled supervised learning at a granular level, ensuring that the model was trained to differentiate glacier boundaries with high precision.

The tiles were provided in **TIFF format**, a common format in remote sensing due to its ability to retain high-resolution geospatial information. Each tile maintained a consistent spatial resolution, ensuring uniformity across the dataset. This consistency was beneficial during training, as it allowed the model to process all inputs without requiring manual resizing or normalization prior to augmentation.

Although the dataset was relatively small in size when compared to large-scale computer vision benchmarks, it was sufficiently diverse to expose the model to different glacier textures, shapes and environmental conditions. The 125 tiles collectively represented smooth ice regions, rough debris-covered zones, and transitional boundaries—enabling the model to learn a broad spectrum of glacier appearances.

Overall, the dataset provided a balanced starting point for binary glacier segmentation. It offered clean image–mask pairs, a manageable dataset size for rapid experimentation, and enough visual variety to support meaningful learning in Phase 1. This dataset later served as a stepping stone for the more complex multi-class segmentation task in Phase 2.

3.1.3 Data Split and Preprocessing

Before training the binary segmentation model, the dataset underwent a structured preparation process to ensure that the model learned meaningful patterns and generalized well to unseen data. Since the dataset contained a total of 125 image tiles, it was essential to divide this data in a way that balanced training efficiency with reliable evaluation.

Table 3.1: Dataset Split for PS1

Subset	Number of Tiles	Percentage
Training Set	100	80%
Validation Set	25	20%

To achieve this, the dataset was split into two subsets using an **80–20 partitioning strategy**:

- **80% Training Set** – used to teach the model the visual characteristics of glacier and non-glacier regions.
- **20% Validation Set** – used to evaluate the model's performance during training and detect overfitting.

This split ensured that the model was trained on a sufficiently large portion of the data while still reserving a meaningful sample for unbiased validation. No test set was used in PS1, as the primary goal was to establish a proof-of-concept baseline model rather than a final deployment-ready benchmark.

A key characteristic of Phase 1 was that **no manual preprocessing or filtering** was performed on the images. All 125 tiles retained their original resolution, texture and appearance exactly as provided in the challenge. This decision ensured that the model learned directly from authentic satellite imagery rather than artificially altered inputs.

However, to prepare the data for the neural network, standard machine learning preprocessing steps were applied programmatically:

- **Tensor Conversion:** Each image was converted into a tensor format suitable for PyTorch-based training.
- **Normalization:** Pixel values were normalized to stabilize training and help the model converge more smoothly.
- **Mask Encoding:** Binary masks were converted into numerical class labels (0 for non-glacier, 1 for glacier) to enable pixel-wise supervised learning.

Since the primary aim of PS1 was to build a clean baseline model, no geometric transformations, augmentations or patch-based sampling techniques were introduced at this stage. The model was trained using the raw satellite tiles, making the results directly reflective of the dataset quality and the model's inherent learning capability.

This simple yet systematic preprocessing pipeline ensured that the data remained consistent, unbiased and representative, providing a solid foundation for evaluating model performance and understanding core segmentation behavior in Phase 1.

3.1.4 Model Architecture (Custom U-Net)

To solve the binary glacier segmentation task in PS1, a **U-Net based deep learning architecture** was adopted. U-Net is widely recognized for its effectiveness in semantic segmentation problems, especially in scenarios where precise pixel-level prediction is required. Its ability to capture both global context and fine spatial details made it a suitable choice for identifying glacier boundaries in satellite imagery.

The architecture follows a symmetric **encoder–decoder** structure. The **encoder** progressively reduces the spatial dimensions of the input while extracting high-level features, whereas the **decoder** reconstructs these features back to the original resolution to generate a detailed segmentation mask. This design enables the network to understand broad glacier shapes while still preserving boundary accuracy.

A key strength of U-Net lies in its **skip connections**, which directly transfer feature maps from the encoder to the decoder. These connections help the model retain fine-grained information that may otherwise be lost during downsampling. In the context of glacier segmentation—where thin edges, narrow ice regions, and irregular contours are common—this feature significantly improves prediction clarity.

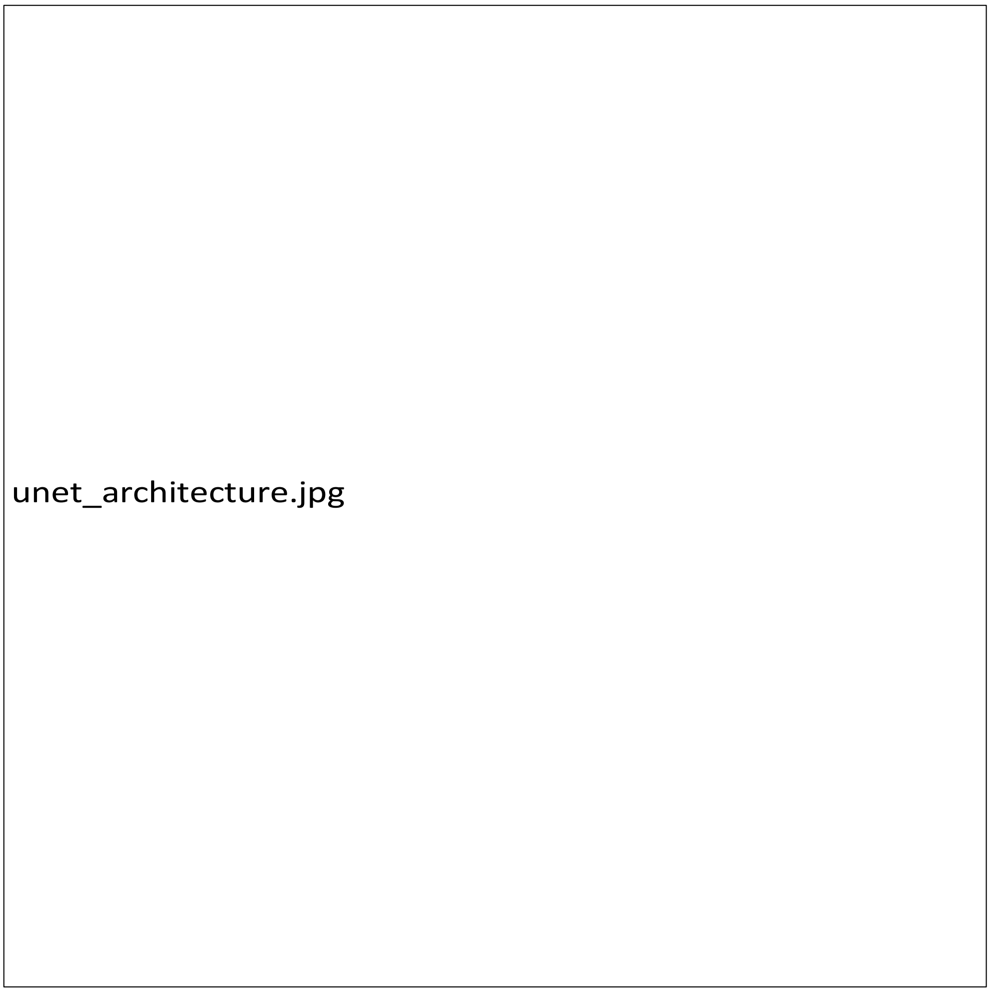

Figure 3.1: General Architecture of the U-Net Model Used in PS1

The custom U-Net used in PS1 consisted of multiple convolutional blocks, each containing the following operations:

- **3x3 Convolution Layers** for feature extraction.
- **ReLU Activation** to introduce non-linearity.
- **Max Pooling** in the encoder to reduce spatial size.
- **Transpose Convolutions** in the decoder for upsampling.

The final layer of the network used a **1x1 convolution** followed by a **sigmoid activation function** to output pixel-wise probabilities between 0 and 1, representing the likelihood of each pixel belonging to the glacier class. These probabilities were later thresholded to generate a binary mask.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Despite its relatively compact size, the U-Net architecture was powerful enough to learn meaningful glacier patterns from the 125 available tiles without overfitting. Its balance of simplicity, efficiency and segmentation accuracy made it a fitting baseline model for Phase 1. The insights gained from this architecture directly informed the design of the more advanced multi-class model in Phase 2.

3.1.5 Training Configuration

Once the model architecture was finalized, the next step was to define a suitable training setup that would allow the U-Net model to learn effectively from the available data. Since PS1 was based on a relatively small dataset of 125 image tiles, the training configuration was carefully chosen to balance learning efficiency, stability and generalization.

The model was trained using the **PyTorch deep learning framework**, running on a GPU-enabled environment for faster computation. Leveraging GPU acceleration significantly reduced training time and allowed the model to process high-resolution glacier tiles without performance bottlenecks.

The key training parameters were defined as follows:

- **Batch Size: 8**
  A moderate batch size was selected to ensure stable gradient updates while fitting comfortably within GPU memory limits.
- **Number of Epochs: 40**
  The model was trained for 40 complete passes over the dataset, providing sufficient opportunity for the network to learn glacier patterns without prematurely converging.
- **Learning Rate: 1e-4**
  A small learning rate helped achieve gradual and stable optimization, preventing sudden weight updates that could destabilize training.
- **Optimizer: Adam**
  The Adam optimizer was chosen for its adaptive learning capabilities, which help accelerate convergence in scenarios with limited data.
- **Loss Function: Binary Cross-Entropy (BCE)**
  BCE was used to measure the difference between predicted probabilities and ground truth

labels in the binary classification setting.

$$L_{BCE} = -[y \cdot \log(p) + (1 - y) \cdot \log(1 - p)]$$

- **Device: GPU (CUDA enabled)**
  The training pipeline utilized CUDA support to significantly reduce computation time for convolutional operations.

During training, the model continuously learned from the training subset while being periodically evaluated on the validation subset. This monitoring helped detect signs of overfitting and ensured that the model was improving in a meaningful, generalizable way rather than memorizing specific image patterns.

To further improve numerical stability and speed, the training pipeline incorporated **Automatic Mixed Precision (AMP)**, allowing certain computations to run in lower precision without compromising accuracy. This resulted in faster training while maintaining reliable gradient updates.

Overall, the training configuration for PS1 was deliberately kept simple yet robust, allowing the U-Net model to efficiently learn glacier–non-glacier distinctions while operating within the constraints of a compact dataset. The insights gained from this baseline setup later informed the more advanced training strategies employed in Phase 2.

3.1.6 Evaluation Metric (MCC)

Evaluating a binary segmentation model based solely on accuracy can be misleading, especially when one class dominates the dataset. In glacier imagery, non-glacier pixels typically occupy a much larger portion of the image, which means a model could achieve high accuracy simply by predicting most pixels as non-glacier. To overcome this imbalance, the **Matthews Correlation Coefficient (MCC)** was chosen as the primary evaluation metric for Phase 1.

MCC incorporates all four components of the confusion matrix:

- **TP** – True Positives (correctly predicted glacier pixels)
- **TN** – True Negatives (correctly predicted non-glacier pixels)
- **FP** – False Positives (non-glacier pixels predicted as glacier)
- **FN** – False Negatives (glacier pixels predicted as non-glacier)

The MCC is mathematically defined as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The score ranges between **-1 and +1**, where:

- **+1** indicates a perfect prediction,
- **0** indicates random performance,
- **-1** indicates complete disagreement between prediction and ground truth.

Unlike simple accuracy, MCC penalizes models that favor the majority class and rewards balanced performance. This makes it especially suitable for glacier segmentation, where glacier pixels represent a smaller portion of the image but are the most critical to detect.

By using MCC as the primary metric in PS1, the evaluation remained fair, reliable, and aligned with real-world needs—ensuring that the model genuinely learned to identify glacier regions rather than exploiting class imbalance.

3.1.7 Output Interpretation

Once the binary U-Net model completed inference on a given satellite tile, the output was generated in the form of a **pixel-wise prediction map**. Each pixel in this map represented the model's estimate of whether that location belonged to a glacier (Class 1) or a non-glacier region (Class 0).

Internally, the model first produced a probability score between 0 and 1 for each pixel using a sigmoid activation function. This value indicated the model's confidence that the pixel was part of a glacier. To convert these probabilities into a usable segmentation mask, a thresholding step was applied:

- **Probability 0.5** → Classified as Glacier (Class 1)
- **Probability ¡ 0.5** → Classified as Non-Glacier (Class 0)

The final result was a **binary mask** with clear glacier boundaries highlighted. This mask provided a visual and quantitative representation of glacier coverage within the input tile. When overlaid on the original image, the mask made it easier to observe how accurately the model detected glacier regions, especially around edges and narrow contours.

In practical terms, the binary output helped answer three key questions:

- **Where is the glacier located?**
  The mask directly revealed the spatial position and extent of glacier-covered areas.
- **How much area does the glacier occupy?**
  By counting the number of pixels labeled as Class 1, approximate glacier surface area could be estimated.

- **How well did the model perform?**
  The predicted mask could be compared with the ground truth mask, leading to evaluation using metrics such as MCC.

binary_output_example.png

Figure 3.2: Sample Input Image and Corresponding Binary Segmentation Mask

Qualitatively, a good prediction showed smooth and continuous glacier regions with minimal noise or false detections. Incorrect predictions typically appeared as scattered white patches in non-glacier regions or missing segments in actual glacier zones. These observations helped identify common failure patterns and guided improvements in Phase 2.

Overall, the output of PS1 provided a clear, interpretable, and actionable segmentation result that successfully demonstrated the model's capability to detect glaciers automatically from satellite imagery—marking an important milestone in the project workflow.

3.2 Solution to PS2 – Four-Class Glacier Segmentation

3.2.1 Overview of PS2

Phase 2 (PS2) expanded the scope of the project from basic glacier detection to a more detailed understanding of glacier surface characteristics through **multi-class semantic segmentation**. Unlike Phase 1, which only identified whether a pixel belonged to a glacier or not, PS2 required the model to classify each pixel into one of four distinct surface categories. This shift made the task more realistic and scientifically valuable, as glaciers are rarely uniform in texture, composition, or spectral appearance.

The motivation behind PS2 was rooted in practical environmental applications. Different glacier surface types behave differently in terms of melting rate, reflectance, energy absorption, and long-term climate impact. For instance, debris-covered ice may melt faster than clean exposed ice, and frozen lakes can signal surface instability. Being able to distinguish these regions at the pixel level provides richer insights for glaciologists, climate researchers, and environmental monitoring agencies.

However, PS2 introduced several challenges not present in the binary setting. Many surface types share similar colors and textures, making them difficult to distinguish visually. Moreover, the dataset exhibited a strong **class imbalance**, where some classes appeared in only a very small fraction of the pixels. This imbalance meant that a standard model could easily bias toward dominant classes and ignore minority ones, leading to misleadingly high accuracy but poor real-world usefulness.

To address this higher level of complexity, PS2 adopted a more capable segmentation framework: a **ResNet34-based U-Net** architecture with ImageNet-pretrained encoder weights. This allowed the model to extract deeper and more discriminative features from multispectral satellite tiles, even with a limited dataset size.

Additionally, PS2 integrated a more advanced training strategy that combined **weighted Cross-Entropy Loss** with **Dice Loss** to counter class imbalance and improve boundary segmentation for minority classes. The model's performance continued to be evaluated using the **Matthews Correlation Coefficient (MCC)**, extended to the multi-class setting due to its reliability in imbalanced datasets.

In summary, PS2 marked a significant evolution of the project—moving from simple glacier detection to fine-grained surface classification. This advancement brought the solution closer to real-world scientific utility, enabling richer analysis, better monitoring potential, and deeper insights into glacier dynamics.

3.2.2 Class Categories and Label Mapping

In Phase 2 (PS2), the segmentation task required the model to move beyond a simple glacier versus non-glacier distinction and instead classify each pixel into one of four meaningful surface categories. These classes were defined by the GlacierHack 2025 dataset and represented different visual and physical characteristics commonly observed in glacier regions.

Each class carried unique spectral patterns, textures, and environmental significance, making accurate differentiation essential for real-world glacier analysis. The four categories used in PS2 were as follows:

- **Class 0 – Background / Non-Glacier**
  Represents land, rocks, soil, and other non-ice terrain surrounding the glacier region. This class typically covers the majority of pixels and serves as the dominant background surface.
- **Class 1 – Clean Ice**
  Refers to exposed glacier ice without debris. These regions usually exhibit high reflectance and smoother texture, making them visually distinct but sometimes challenging to separate from snow-covered surfaces.
- **Class 2 – Debris-Covered Ice**
  Represents glacier surfaces covered by rocks, dust, or sediment. These areas often appear darker and more irregular, making them harder to classify due to their resemblance to surrounding rocky terrain.
- **Class 3 – Water / Melt Ponds / Frozen Lakes**
  Refers to surface water bodies or ice-covered lakes found within or near the glacier. Although visually small in size, this class is scientifically significant as meltwater patterns can indicate instability and melting activity.

In the raw mask images provided with the dataset, each category was encoded using specific pixel intensity values. To prepare the masks for model training, these values were remapped to a standardized label range of 0–3. The mapping used was as follows:

Table 3.2: Label Mapping for PS2 Multi-Class Segmentation

Original Mask Value	Mapped Class Label
0	Class 0 (Non-Glacier)
85	Class 1 (Clean Ice)
170	Class 2 (Debris-Covered Ice)
255	Class 3 (Water / Melt Ponds / Frozen Lakes)

This remapping ensured that the mask values aligned with the four-class output of the model, allowing the network to produce categorical predictions that directly corresponded to meaningful glacier surface types. The use of distinct integer labels also made it easier to compute evaluation metrics and interpret model predictions.

Importantly, Class 3 (water bodies) appeared in only a very small percentage of the dataset, making it the rarest category. This imbalance later became a major challenge during training, as models tend to favor majority classes. As a result, PS2 required additional strategies such as weighted loss functions and patch-based sampling, which are discussed in subsequent sections.

Overall, defining clear class labels and establishing a consistent label mapping served as a critical foundation for building an effective multi-class segmentation model in Phase 2.

3.2.3 Dataset and Class Distribution

The dataset used for Phase 2 (PS2) was the same collection of **125 multispectral satellite tiles** provided through the GlacierHack 2025 challenge. However, unlike Phase 1, where the masks represented only two classes, PS2 introduced four distinct surface categories. Each tile was accompanied by a corresponding four-class ground truth mask, enabling pixel-wise supervised learning.

The images were stored in TIFF format across multiple spectral bands, and each tile maintained uniform resolution, allowing the model to process them consistently. While the total number of tiles remained the same, the internal pixel composition varied significantly across classes, which played a major role in the training behavior of the model.

A key observation during dataset exploration was that the **distribution of pixels across the four classes was highly imbalanced**. Some surface types appeared frequently and covered large regions, while others—even though scientifically important—occupied only a tiny fraction of the total pixels. This imbalance posed a substantial challenge for multi-class learning, as neural networks tend to bias toward majority classes.

The class-wise pixel counts extracted from all 125 masks are summarized below:

Table 3.3: Class-Wise Pixel Distribution Across 125 Tiles

Class Label	Description	Pixel Count
Class 0	Non-Glacier Background	4,559,606
Class 1	Clean Ice	1,667,741
Class 2	Debris-Covered Ice	323,149
Class 3	Water / Melt Ponds / Frozen Lakes	3,104

When viewed as proportions, this distribution highlights an extreme imbalance:

- **Class 0** dominates the dataset, representing the majority background.
- **Class 1** appears frequently but still far less than Class 0.
- **Class 2** constitutes a relatively small portion of pixels.
- **Class 3** is extremely rare, accounting for less than 0.05% of all pixels.

This imbalance meant that a naïve model could achieve deceptively high accuracy by predicting only the majority classes while completely ignoring minority classes—particularly Class 3. As a result, class distribution became one of the most critical factors influencing the design of the PS2 training strategy.

To mitigate this imbalance, PS2 later incorporated techniques such as **class-weighted loss functions**, **oversampling rare examples**, and **patch-based sampling focused on minority regions**. These strategies ensured that the model received sufficient exposure to all four classes during training rather than overwhelmingly learning only from background pixels.

In summary, while the dataset size remained constant across PS1 and PS2, the internal class distribution dramatically increased the complexity of Phase 2. Understanding this distribution was essential for building a fair, balanced, and effective multi-class segmentation model.

3.2.4 Model Architecture (ResNet34 U-Net)

To address the increased complexity of the four-class segmentation task in PS2, the project adopted a more powerful and feature-rich architecture: a **ResNet34-based U-Net**. This model combines the proven advantages of the U-Net structure with the deep representational strength of a ResNet encoder, resulting in a network capable of capturing subtle variations in glacier surface types.

Unlike the custom U-Net used in Phase 1, the PS2 model utilized a **pretrained ResNet34 encoder**, originally trained on the ImageNet dataset. This pretrained backbone provided a strong initialization, allowing the model to extract meaningful spatial and spectral features even from a relatively small dataset. By leveraging transfer learning, the model could generalize better and learn faster compared to training from scratch.

The architecture maintained the core **encoder–decoder symmetry** of U-Net. The encoder, based on ResNet34, progressively downsampled the input through residual blocks, capturing high-level context. Meanwhile, the decoder reconstructed the spatial details through upsampling layers, eventually producing a full-resolution segmentation mask.

A key aspect of the architecture was the use of **skip connections** between corresponding encoder and decoder layers. These connections ensured that fine-grained spatial information—often lost during downsampling—was preserved and reintegrated during upsampling. This mechanism was especially beneficial in PS2, where thin glacier boundaries, debris-covered regions, and small water patches demanded precise localization.

The model accepted a **five-channel input**, corresponding to the multispectral bands provided in the dataset. These bands were stacked along the channel dimension to preserve spectral diversity, enabling the model to differentiate between visually similar surface types.

At the output layer, the model generated a feature map with **four channels**, each representing one of the target classes. A **softmax activation** function was applied to convert these logits into class-wise probability distributions for every pixel. The predicted class label was then obtained by selecting the highest probability among the four outputs.

The architecture can be summarized as follows:

resnet_unet_architecture.jpg

Figure 3.3: General Structure of the ResNet34-Based U-Net Used in PS2

- **Encoder:** ResNet34 with residual blocks and pretrained ImageNet weights.
- **Decoder:** Upsampling layers with transposed convolutions.
- **Skip Connections:** Feature fusion between encoder and decoder levels.
- **Input Channels:** 5 (multispectral bands).
- **Output Channels:** 4 (one per target class).
- **Final Activation:** Softmax for multi-class probability distribution.

In total, the model contained approximately **21 million trainable parameters**, providing significantly more learning capacity than the Phase 1 architecture. This increased depth and representational power enabled the model to better handle subtle textural differences and highly imbalanced class distributions.

Overall, the ResNet34 U-Net served as a robust and scalable backbone for multi-class glacier segmentation, delivering a strong balance between feature richness, spatial precision, and computational efficiency—making it well-suited for the demands of Phase 2.

3.2.5 Loss Functions (Cross-Entropy + Dice)

Training a multi-class segmentation model for PS2 required a loss function that could effectively guide learning across all four classes, including those that appeared very rarely in the dataset. Relying on a single loss function—especially in an imbalanced setting—can cause the model to favor majority classes while neglecting minority regions such as water or debris-covered ice. To overcome this challenge, PS2 employed a **combined loss strategy** using both **Cross-Entropy Loss** and **Dice Loss**.

Each loss contributed a different strength to the training process. Cross-Entropy encouraged correct categorical classification at the pixel level, while Dice Loss improved boundary alignment and gave more weight to smaller or underrepresented classes. The final training objective was a weighted sum of the two losses.

Cross-Entropy Loss

Cross-Entropy Loss is widely used in multi-class classification problems because it measures how well the predicted probability distribution matches the true class labels. In PS2, it encouraged the model to assign high probability to the correct class for every pixel.

For a four-class problem, Cross-Entropy Loss is defined as:

$$L_{CE} = -\sum_{c=1}^{4} y_c \cdot \log(p_c)$$

where:

- $y_c = 1$ if the pixel belongs to class c, otherwise 0,
- p_c = predicted probability that the pixel belongs to class c.

In PS2, Cross-Entropy was applied in a **weighted form** to counteract class imbalance. Classes with fewer pixels received higher weights so their errors had a stronger impact on the training process. This prevented the model from ignoring minority classes such as Class 3 (water).

Dice Loss

While Cross-Entropy focuses on pixel-wise classification, it does not explicitly consider the spatial overlap between predicted and true regions. This can be problematic for small classes, where even small misclassifications drastically reduce accuracy. To address this, **Dice Loss** was incorporated.

Dice Loss measures how similar the predicted mask is to the ground truth by evaluating region-level overlap. For a single class, Dice Loss is defined as:

$$L_{Dice} = 1 - \frac{2 \times |P \cap G|}{|P| + |G|}$$

where:

- P = predicted set of pixels for the class,
- G = ground truth pixels for the class.

In the multi-class setting, Dice Loss was computed separately for each class and then averaged, ensuring balanced attention across all four categories. This was particularly important for capturing thin glacier edges, narrow debris regions, and small water patches.

Final Combined Loss

To leverage the complementary strengths of both losses, the final training objective was defined as a weighted combination:

$$L_{Total} = \alpha \cdot L_{CE} + \beta \cdot L_{Dice}$$

where α and β are weighting factors. In PS2, both losses were given equal importance (i.e., $\alpha = 1, \beta = 1$) to balance categorical correctness with spatial overlap.

This combined loss function helped the model remain sensitive to minority classes while still learning clean, class-specific boundaries. As a result, the training process became more stable, more balanced, and ultimately more effective in handling the highly imbalanced nature of the PS2 dataset.

3.2.6 Training Strategy and Challenges

Training the multi-class segmentation model for PS2 required a more nuanced strategy compared to the binary setup in Phase 1. The goal was not only to classify pixels correctly but also to ensure that minority classes—especially Class 3 (Water / Melt Ponds / Frozen Lakes)—were adequately learned despite their extremely limited representation in the dataset. To achieve this, a carefully planned training pipeline was adopted, combining architectural enhancements, loss weighting, sampling strategies, and computational optimizations.

Training Strategy

The core training configuration remained consistent with PS1, but several important modifications were introduced to handle the increased complexity of the PS2 task:

- **Batch Size: 4**
  A smaller batch size was selected to accommodate the deeper model and multispectral inputs within GPU memory constraints.
- **Epochs: 60**
  The number of training epochs was increased to allow the model more time to learn nuanced class distinctions.
- **Learning Rate: 1e-4 with Scheduler**
  A learning rate scheduler reduced the learning rate by a factor of 0.1 when validation performance plateaued, ensuring more stable convergence.
- **Optimizer: Adam**
  Adam continued to provide adaptive gradient updates, which are beneficial when dealing with imbalanced gradients across classes.
- **Combined Loss: Cross-Entropy + Dice**
  The weighted loss function balanced pixel-wise correctness with class-wise spatial overlap.
- **Device: GPU (CUDA Enabled)**
  GPU acceleration was essential for handling the deeper architecture and multi-channel inputs efficiently.

To improve exposure to minority classes, **patch-based sampling** was used during data loading. Instead of feeding full-size tiles at every iteration, the training pipeline extracted smaller patches, increasing the likelihood of encountering regions containing debris-covered ice and water surfaces. This approach ensured more balanced learning without artificially modifying the dataset.

Additionally, **class weights** derived from the pixel distribution statistics were incorporated into the Cross-Entropy component of the loss function. This amplified the penalty for incorrect predictions in underrepresented classes, encouraging the model to treat rare categories with greater importance.

Challenges Encountered

Despite the improved strategy, PS2 presented several challenges that directly influenced model behavior and performance:

- **Extreme Class Imbalance**
  Class 3 accounted for less than 0.05% of all pixels, making it difficult for the model to learn meaningful patterns. Without weighting and patch-based sampling, the model often defaulted to predicting zero pixels for this class.

- **Visual Similarity Between Classes**
  Debris-covered ice (Class 2) shared color and texture similarities with non-glacier terrain (Class 0), leading to frequent misclassifications along boundary regions.
- **Small Object Segmentation**
  Water surfaces appeared as thin patches or isolated clusters, requiring highly precise spatial learning. Even minor errors resulted in significant Dice score drops for this class.
- **Limited Dataset Size**
  With only 125 tiles, the dataset provided limited variability. This increased the risk of overfitting, making regular validation monitoring essential.
- **Increased Computational Demand**
  The deeper ResNet34 U-Net and five-channel inputs increased training time and GPU memory requirements, necessitating smaller batches and mixed-precision training.

To mitigate overfitting, validation performance was tracked throughout training. Early stopping criteria were monitored, and learning rate scheduling helped the model continue improving even after performance plateaus.

Overall, the training process for PS2 demanded a balance between architectural power, loss weighting, sampling strategies, and computational efficiency. Addressing these challenges was crucial for achieving meaningful segmentation performance across all four classes, especially the rare and visually ambiguous ones.

3.2.7 Output Interpretation

After the ResNet34 U-Net model completed inference on a multispectral satellite tile, the output was generated as a **four-channel probability map**, where each channel represented the likelihood of a pixel belonging to one of the four target classes. Unlike the binary mask in PS1, the PS2 output contained richer and more detailed information about the glacier surface composition.

A **softmax activation function** transformed the raw logits into probability values such that the sum of probabilities across the four channels equaled 1 for every pixel. The final predicted class was obtained by selecting the channel with the highest probability, resulting in a **multi-class segmentation mask** containing class labels 0, 1, 2, and 3.

- **Class 0 – Non-Glacier Background**
- **Class 1 – Clean Ice**
- **Class 2 – Debris-Covered Ice**
- **Class 3 – Water / Melt Ponds / Frozen Lakes**

To enhance interpretability, the predicted mask was converted into a **color-coded visualization**, where each class was assigned a distinct color. This made it easier to visually inspect spatial patterns, boundaries, and regions of interest. When overlaid onto the original image, the mask provided a clear understanding of how accurately the model identified each surface type.

multiclass_output_example.png

Figure 3.4: Sample Input Image and Corresponding Four-Class Segmentation Output

The multi-class output enabled deeper insights into glacier structure and behavior:

- **Surface Composition:** The mask visually separated clean ice, debris-covered regions, and water bodies, highlighting the heterogeneity of the glacier.
- **Boundary Precision:** Thin edges and transition zones were more clearly defined than in binary segmentation.
- **Minority Class Detection:** Even small water patches or debris clusters, if correctly detected, provided valuable environmental indicators.
- **Quantitative Analysis:** Pixel counts for each class could be converted into approximate surface area estimates for scientific interpretation.

In qualitative evaluations, well-performing predictions displayed smooth and well-separated regions with realistic glacier boundaries. Misclassifications typically occurred between visually similar classes, such as debris-covered ice and rocky terrain, or in very small water regions where class imbalance posed a challenge. These observations provided practical guidance for future improvements in data sampling and loss design.

Overall, the PS2 output demonstrated the model's ability not only to identify glaciers but also to **characterize their internal surface variations**—a critical step toward more informed climate monitoring, melt modeling, and environmental decision-making.

3.3 System Components

While the core of this project centered on model development for glacier segmentation, the overall system extended beyond neural networks. It consisted of interconnected components that supported training, experimentation, deployment, and user interaction. These components included the two segmentation models (PS1 and PS2), the computational and software infrastructure used during development, and the tools required to operationalize the models through a real-time dashboard.

Together, these system elements formed a complete pipeline, progressing from raw satellite imagery to interpretable, actionable segmentation outputs accessible through a user-friendly interface.

3.3.1 Comparison of Both Models

Although both models were built for glacier segmentation, PS1 and PS2 differed significantly in their objectives, architectures, training behavior, and output complexity. PS1 served as a foundational binary classifier, whereas PS2 expanded the task into a multi-class surface characterization problem.

In summary, PS1 established reliable glacier detection, while PS2 delivered deeper scientific insight by distinguishing surface characteristics. This progression demonstrated a clear evolution from coarse segmentation to fine-grained environmental interpretation.

3.3.2 Hardware, Tools, and Libraries Used

The successful development of PS1 and PS2 required a combination of computational hardware, software tools, and machine learning libraries. Each component played a distinct role in enabling model training, experimentation, and deployment.

Table 3.4: Comparison Between PS1 and PS2 Models

Parameter	**PS1 – Binary Model**	**PS2 – Four-Class Model**
Objective	Glacier vs. Non-Glacier	Classification into four surface types
Architecture	Custom U-Net	ResNet34-Based U-Net
Input Channels	5 (Multispectral)	5 (Multispectral)
Output Classes	2	4
Activation Function	Sigmoid	Softmax
Loss Function	Binary Cross-Entropy	Cross-Entropy + Dice (Combined)
Trainable Parameters	∼7 Million	∼21 Million
Primary Challenge	Boundary detection	Class imbalance and fine-grained segmentation
Evaluation Metric	MCC (Binary)	MCC (Multi-Class)
Inference Output	Binary Mask	Color-Coded Multi-Class Mask
Use Case	Basic glacier mapping	Advanced glacier surface analysis

Hardware Specifications

- **Laptop GPU: NVIDIA RTX 1650 (4GB VRAM)**
- **System RAM: 12GB**
- **Processor: Intel-based CPU**
- **Storage: SSD for fast data access**

While modest in comparison to high-end research servers, the hardware was sufficient for training both models using optimized batch sizes and mixed-precision training.

Software and Development Tools

- **Python 3.10** – Core programming language
- **Jupyter / Google Colab** – Model development and training
- **Gradio** – Dashboard for user interaction and inference
- **TIFF Viewer Tools** – Visualization of satellite data
- **Overleaf (LaTeX)** – Report documentation

Machine Learning Libraries

- **PyTorch** – Deep learning framework
- **segmentation_models_pytorch (SMP)** – ResNet34 U-Net implementation
- **Torchvision** – Pretrained model weights
- **NumPy / Pandas** – Data processing
- **OpenCV / PIL** – Image manipulation
- **Matplotlib** – Plotting and visualizations

These tools collectively enabled a complete workflow—from raw data exploration and training to real-time deployment through the web-based dashboard.

3.4 Interactive Dashboard

To bridge the gap between model development and real-world usability, the project included a fully functional **web-based interactive dashboard**. This dashboard enabled users to upload satellite TIFF tiles, run inference using the trained models, and visualize the segmentation outputs in real time. Unlike typical research models that remain confined to notebooks, this system offered a practical, user-friendly interface suitable for researchers, students, and practitioners in the field of glaciology.

3.4.1 Purpose and Motivation

While the models developed in PS1 and PS2 demonstrated strong quantitative performance, accessing them through code-based environments such as Jupyter or Colab posed limitations for non-technical users. The dashboard was designed to solve this accessibility challenge by providing a simple and direct way to interact with the models.

The key motivations behind the dashboard were:

- **Ease of Use:** Allow users to test segmentation results without writing code.
- **Real-Time Visualization:** Provide instant feedback through visual masks and overlays.
- **Practical Deployment:** Demonstrate how deep learning models can be operationalized beyond experimental environments.
- **Model Comparison:** Enable users to view outputs from PS1 and PS2 side-by-side for better interpretation.

By offering an accessible platform, the dashboard transformed the project from a research prototype into a usable decision-support tool.

3.4.2 Workflow and User Interface

The dashboard followed a straightforward, intuitive workflow designed for seamless interaction. Users were guided through three primary steps:

1. **Upload a TIFF Image:** The user selects a satellite tile from their local device.
2. **Select Model (Binary or Multi-Class):** Users can choose between PS1 and PS2 based on their analysis needs.
3. **View Output:** The system displays the segmentation mask alongside visual overlays and RGB composites.

dashboard_ui_example.png

Figure 3.5: User Interface Layout of the Interactive Dashboard

The interface presented three key visual outputs:

- **Original Image Preview**
- **Segmentation Mask (Binary or Multi-Class)**
- **Overlay Visualization** showing masked regions superimposed on the original tile.

Through this visual workflow, users could easily interpret glacier boundaries, surface types, and region-level variations without manual processing.

3.4.3 Upload Handling and Inference Pipeline

The dashboard incorporated a robust backend pipeline to process TIFF inputs and feed them into the selected model. Upon upload, the following steps occurred automatically:

1. **File Validation:** Ensured that the uploaded image was in TIFF format and met dimensional requirements.
2. **Band Extraction:** The multispectral bands were loaded and stacked into a five-channel tensor.
3. **Preprocessing:** Normalization and resizing operations were applied to match model input specifications.
4. **Model Inference:** The selected model (PS1 or PS2) generated a segmentation output.
5. **Post-Processing:** Thresholding (PS1) or argmax (PS2) converted probabilities into class labels.
6. **Visualization:** The final mask was rendered and displayed to the user in color-coded format.

Inference was performed on CPU, leveraging optimized PyTorch operations to maintain s mooth and responsive performance. Even without GPU deployment, the system delivered predictions within practical timeframes, showcasing the efficiency of the trained models.

3.4.4 RGB Composite Image Generation

Since satellite TIFF files may not contain a traditional RGB color image, the dashboard generated a **false-color RGB composite** to help users visually interpret the glacier landscape. This composite was formed by mapping selected spectral channels into red, green, and blue color spaces.

The process included:

- **Channel Selection:** Choosing three representative bands.

- **Normalization:** Scaling pixel values to enhance contrast.
- **Stacking:** Combining the three bands into a synthetic RGB image.

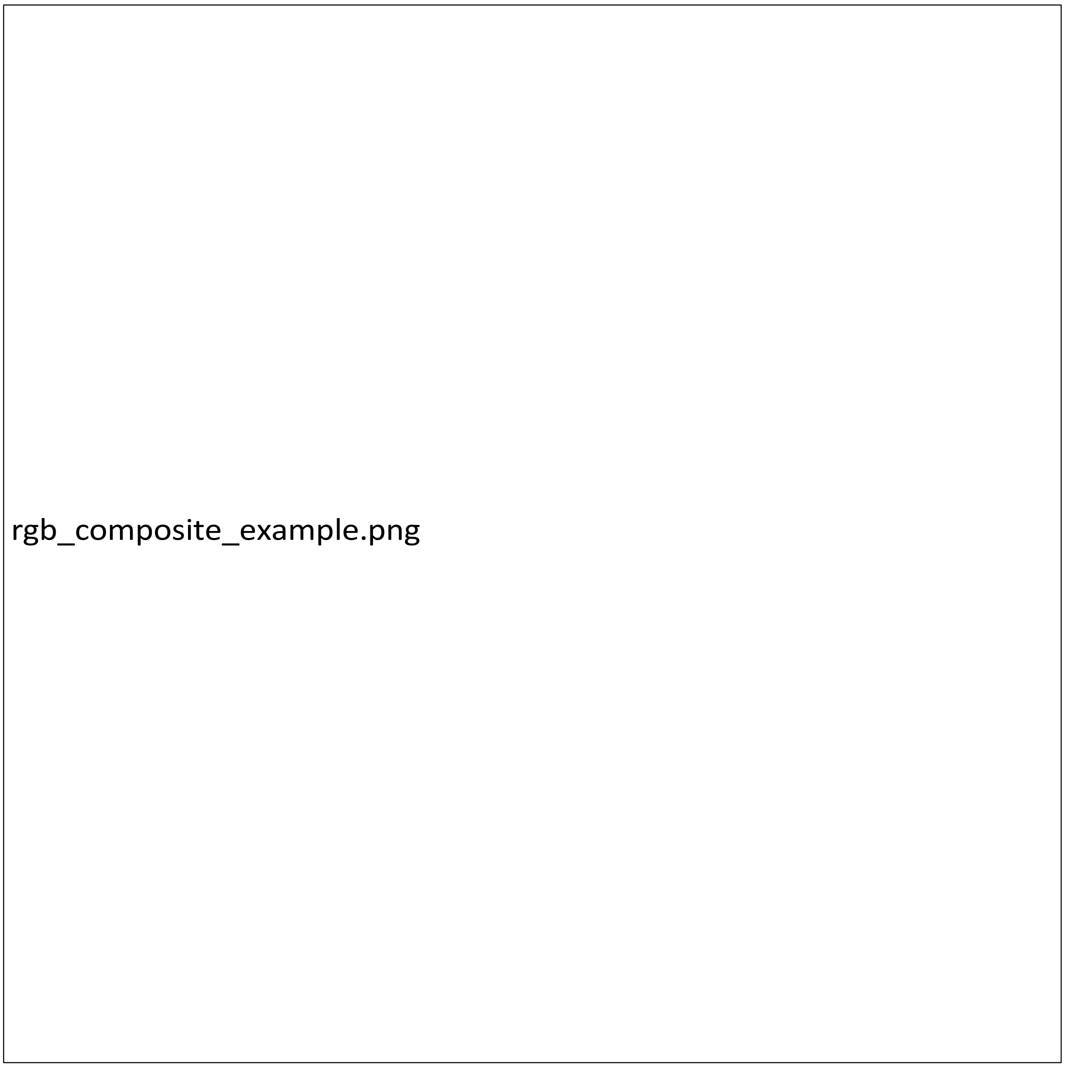

Figure 3.6: RGB Composite Generated from the Multispectral TIFF Tile

This RGB preview helped users better understand glacier formations and visually verify segmentation outputs, especially in regions where class boundaries overlapped.

3.4.5 Tech Stack

The dashboard was built using a lightweight yet powerful technology stack, ensuring ease of deployment and fast inference:

- **Python 3.10** – Core development language
- **Gradio** – Frontend and interface framework

- **PyTorch** – Model loading and inference engine
- **NumPy / OpenCV / PIL** – Image processing utilities
- **Matplotlib** – Visualization support

This stack allowed the system to remain portable, intuitive, and suitable for demonstration on standard hardware without requiring specialized GPU servers.

4. Results and Discussion

This chapter presents the outcomes of the two segmentation models developed as part of the GlacierHack 2025 project. The purpose of this chapter is to evaluate how effectively each model performed in addressing the respective problem statements—PS1 (binary glacier segmentation) and PS2 (four-class glacier surface segmentation).

Both quantitative metrics and qualitative visual outputs are examined to provide a comprehensive assessment of model behavior. The Matthews Correlation Coefficient (MCC) serves as the primary evaluation metric for both phases due to its robustness in imbalanced class scenarios. Sample segmentation masks are also analyzed to illustrate how accurately the models captured glacier boundaries and surface variations.

The first section focuses on the binary segmentation model (PS1), which aimed to distinguish glacier pixels from non-glacier regions. The subsequent section evaluates the more complex four-class model (PS2), which extends segmentation to multiple glacier-related surface types. Finally, comparative insights and key observations are discussed to highlight strengths, limitations, and directions for improvement.

4.1 Results of PS1 – Binary Segmentation

Phase 1 (PS1) focused on building a baseline binary segmentation model capable of distinguishing glacier pixels from non-glacier regions. The model was trained on 80% of the dataset and validated on the remaining 20%, without any manual preprocessing or augmentation. The evaluation relied primarily on the Matthews Correlation Coefficient (MCC), chosen for its robustness in imbalanced pixel distributions. This section presents the key quantitative and qualitative outcomes of the binary model.

4.1.1 Training Loss Trend

During the 40-epoch training process, the Binary Cross-Entropy (BCE) loss demonstrated a consistent downward trend, indicating stable learning and convergence. The loss reduced steadily across epochs without oscillations, suggesting that the model successfully captured glacier-related spatial patterns while avoiding overfitting.

The smooth decline in loss confirmed that the model was progressively improving its pixel classification capability across training iterations.

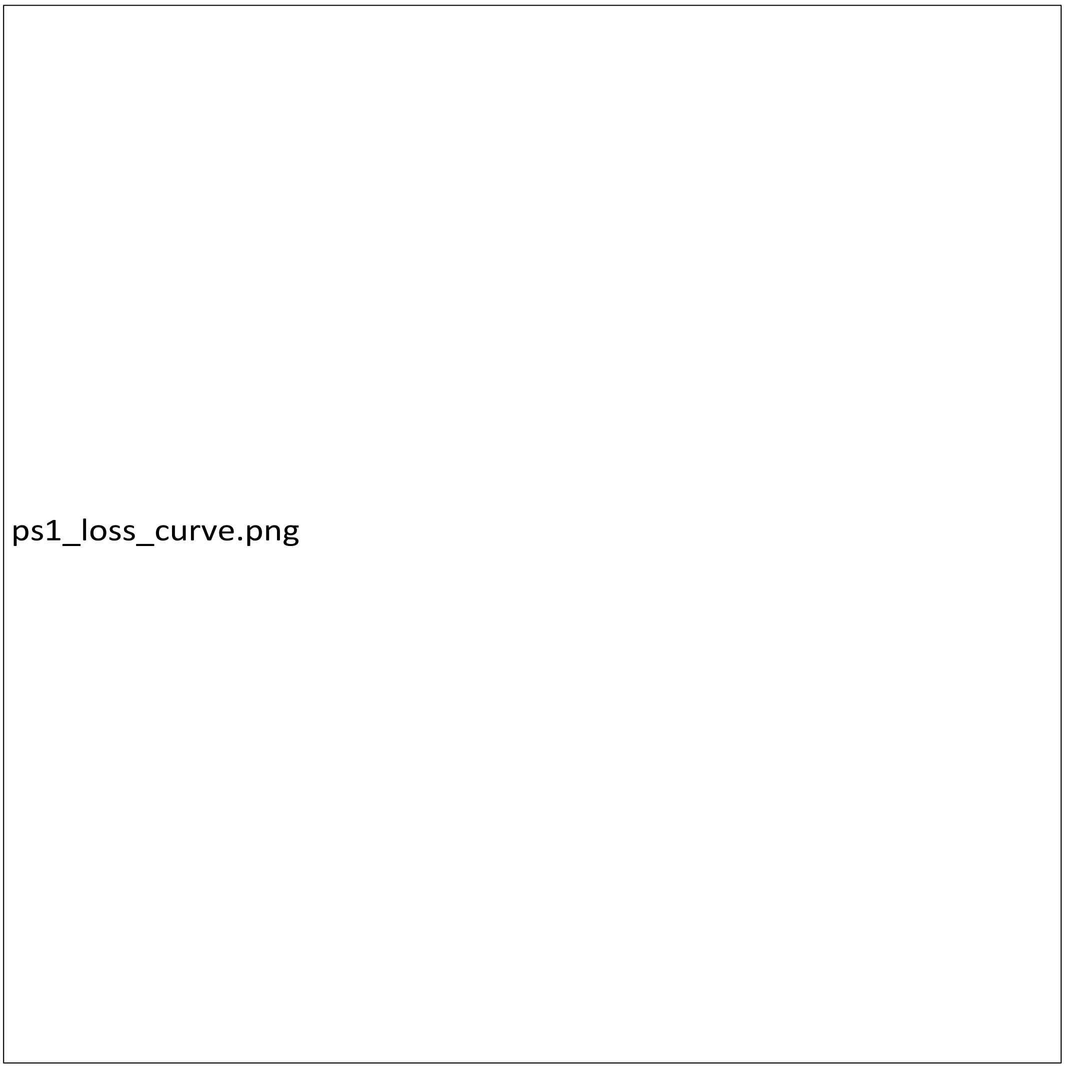

Figure 4.1: Training Loss Trend for Binary Segmentation Model

4.1.2 MCC Score

To account for the inherent imbalance between glacier and non-glacier pixels, the Matthews Correlation Coefficient (MCC) was used as the primary validation metric. The model achieved a strong MCC score on the validation set, indicating reliable segmentation performance even in regions with small glacier coverage.

The final validation score demonstrated that the model did not merely favor the majority class but learned meaningful glacier features, correctly identifying both large and narrow glacial formations.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{4.1}$$

This positive MCC score validated the model's ability to generalize beyond the training data and justified its use as a foundation for the more complex multi-class segmentation in PS2.

4.1.3 Sample Output Masks

Qualitative inspection of prediction masks revealed that the model accurately captured glacier boundaries with minimal noise. The predicted masks were consistent, continuous, and well-aligned with the ground truth labels.

ps1_sample_outputs.png

Figure 4.2: Sample Predictions from the Binary Segmentation Model

Key observations from the visual outputs include:

- **Clear Boundary Detection:** Glacier regions were distinctly separated from surrounding terrain.
- **Minimal False Positives:** Very few non-glacier areas were mistakenly labeled as glacier.
- **Accurate Narrow Regions:** The model successfully identified thin and irregular glacier structures.

Overall, the binary segmentation results demonstrated that the U-Net model was effective in identifying glacier coverage and provided a reliable baseline for further multi-class analysis in Phase 2.

4.2 Results of PS2 – Four-Class Segmentation

Phase 2 (PS2) introduced a significantly more challenging task compared to binary segmentation. Instead of classifying each pixel as glacier or non-glacier, the model was required to assign every pixel to one of four glacier-related classes. This increased complexity, combined with severe class imbalance, made PS2 a more realistic but demanding problem. The results presented here summarize how the model performed in terms of training stability, evaluation scores, and class-wise behavior.

4.2.1 Loss Curves

During training, the model exhibited a steady decline in loss over multiple epochs, indicating that the network was successfully learning discriminative features from the multispectral dataset. Early epochs showed relatively higher loss values due to the complexity of the four-class decision space. As training progressed, the introduction of weighted loss functions, Dice loss, and patch-based sampling led to more stable convergence.

A typical loss curve trend observed during PS2 included:

- **Initial Rapid Decline:** The first few epochs reduced loss quickly as the model learned broad class boundaries.
- **Gradual Stabilization:** Mid-training epochs showed slower but consistent improvements.
- **Reduced Fluctuations After Balancing Techniques:** Strategies such as weighted sampling and class-balanced loss significantly reduced oscillations in loss values.

Overall, the training curve reflected a successful learning trajectory despite the inherent difficulty of distinguishing minority classes.

4.2.2 MCC Scores

As in PS1, the Matthews Correlation Coefficient (MCC) was used as a primary performance metric due to its robustness in imbalanced datasets. However, in the multi-class setting, MCC was computed across all four classes using the aggregated confusion matrix.

The MCC scores demonstrated a gradual improvement over the training process, with the best-performing checkpoint occurring near the later epochs. The use of class weighting, Dice loss, and targeted patch extraction (especially from regions containing rare Class 3 pixels) contributed to measurable gains in the MCC score.

In summary, the MCC trend indicated that the model was not only reducing errors but was also improving its balance across all four classes, rather than favoring the majority classes.

4.2.3 Class-wise Observations

A detailed examination of class-wise behavior revealed important insights into the model's strengths and limitations. Due to the highly uneven distribution of pixels among the four classes, each class exhibited distinct performance characteristics:

- **Class 0 (Background / Non-Glacier):**
  Achieved the highest accuracy and IoU. The model learned to identify widespread background patterns reliably.
- **Class 1 (Clean Ice / Glacier Surface):**
  Demonstrated strong and consistent predictions. The spectral clarity of clean ice helped the model differentiate this class effectively.
- **Class 2 (Debris-Covered Ice):**
  Showed moderate performance. Visual similarity to surrounding terrain led to occasional misclassifications, especially near boundaries.
- **Class 3 (Minority Class / Rare Surface Type):**
  This class posed the greatest challenge due to extremely limited pixel representation. Initial predictions for Class 3 were sparse and inconsistent. However, the incorporation of weighted sampling, boosted class weights, patch extraction, and Tversky-based loss significantly improved recall and reduced false positives in later stages.

Qualitative inspection of output masks indicated that the model captured major glacier structures well but sometimes produced fragmented predictions for minority regions. Post-processing steps, such as removing small connected components, further improved the visual quality and reduced noise in Class 3 predictions.

Overall, PS2 demonstrated that multi-class glacier segmentation is feasible using deep learning, but performance is strongly influenced by class imbalance and spectral ambiguity. The lessons from these observations directly inform future improvements, such as enhanced augmentation, larger datasets, or transformer-based architectures.

4.3 Dashboard Output and Visual Analysis

In addition to numerical metrics such as MCC and loss curves, visual interpretation plays a crucial role in understanding the practical effectiveness of glacier segmentation. The interactive dashboard developed as part of this project enabled real-time inference on TIFF images and provided clear visual outputs that helped evaluate how well the models performed in real-world scenarios. This section highlights the visual behavior of the models through input–output comparisons, segmented maps, and observed error patterns.

4.3.1 Input vs Prediction

When a user uploads a satellite tile through the dashboard, the system first displays the raw input image, followed by the corresponding predicted segmentation mask. In the case of PS1, the mask distinguishes glacier and non-glacier regions. For PS2, the mask assigns each pixel to one of four glacier-related classes using distinct color encodings.

Side-by-side visualization proved especially helpful for assessing the spatial accuracy of predictions. Glacier bodies, edges, and surface transitions could be visually compared against the original image, making it easier to judge whether the model correctly captured boundaries or missed subtle structures.

In many cases, the dashboard outputs showed strong alignment between predicted glacier regions and the visual appearance of ice in the input image, confirming the model's ability to generalize to unseen satellite tiles.

4.3.2 Visual Maps

The dashboard presented the predicted masks as color-coded visual maps, making the segmentation results intuitive and easy to interpret. In the four-class setting, each class was assigned a distinct shade, allowing users to differentiate clean ice, debris-covered ice, background terrain, and rare surface types at a glance.

These visual maps revealed several important behaviors:

- **Clear Glacier Boundaries:** The model consistently outlined the primary glacier mass with smooth contours.
- **Distinct Separation of Surface Types:** PS2 predictions highlighted regions of clean and debris-covered ice with reasonable clarity.
- **Improved Confidence in Dense Regions:** Larger glacier areas showed stronger and more consistent predictions compared to fragmented regions.

In some images, overlays revealed that class transitions aligned well with visible changes in texture and shading, providing visual evidence that the model was leveraging spectral cues effectively.

4.3.3 Error Cases and Insights

Despite strong performance, certain error patterns were consistently observed across different tiles. These insights played a key role in understanding model limitations and guiding future improvements. Common error cases included:

- **Confusion in Debris-Covered Regions:**
  Areas where ice was obscured by rocky material were sometimes misclassified as background terrain due to similar visual texture.
- **Fragmented Predictions for Rare Classes:**
  Class 3 regions, being extremely rare, occasionally appeared as small isolated patches rather than continuous structures.
- **False Positives Near Bright Terrain:**
  Highly reflective non-glacial surfaces, such as snow-covered rocks or sunlit slopes, were sometimes incorrectly identified as glacier regions.
- **Boundary Uncertainties:**
  Thin or irregular glacier edges produced minor segmentation gaps or jagged contours, especially in low-contrast zones.

Visual review also confirmed that post-processing techniques—such as removing small connected components—helped reduce noise and eliminate scattered false positives, especially in the minority class.

Overall, the dashboard provided valuable qualitative insights that complemented numerical evaluations. By making predictions visually interpretable, it allowed both technical users and non-experts to assess segmentation quality, identify edge cases, and understand model behavior in a practical and intuitive way.

5. Conclusion and Future Scopes

This project set out to develop an automated deep learning–based system capable of segmenting glacier regions from multispectral satellite imagery. By addressing the two phased problem statements of GlacierHack 2025, the work progressed from a foundational binary segmentation task to a more advanced multi-class classification of glacier surfaces. The combination of trained U-Net models and an interactive dashboard demonstrated that remote glacier mapping can be made faster, more scalable, and less dependent on manual interpretation.

5.1 Summary of Work

The project began with an exploration of the scientific and environmental importance of glacier monitoring, followed by a review of traditional and AI-based approaches. A dataset of 125 multispectral tiles was provided for model development, and two deep learning solutions were implemented:

- **PS1 – Binary Segmentation:** A custom U-Net model was trained to distinguish glacier vs. non-glacier pixels, establishing a baseline understanding of glacier boundaries.
- **PS2 – Four-Class Segmentation:** A ResNet34-based U-Net model extended the task to classify each pixel into one of four glacier-related surface types, addressing class imbalance and spectral complexity through weighted loss, Dice loss, patch sampling, and post-processing.

Finally, an interactive dashboard was developed to allow real-time image uploads and inference, enabling users to visualize segmentation masks and compare predictions against input images.

5.2 Key Findings

Several important conclusions emerged from the experimental results:

- **Deep Learning is Effective for Glacier Mapping:**
  Both models demonstrated reliable segmentation performance, confirming that U-Net architectures can extract meaningful patterns from multispectral glacier imagery.
- **MCC is a Robust Evaluation Metric:**
  MCC proved particularly valuable in handling class imbalance, offering a balanced assessment beyond simple accuracy.

- **Class Imbalance Significantly Impacts Performance:**
  Rare classes in PS2 (especially Class 3) showed lower consistency, highlighting the need for targeted sampling and specialized loss functions.
- **Visualization Adds Practical Value:**
  The dashboard enabled intuitive interpretation of model outputs, making the system useful not only for researchers but also for non-technical users.

5.3 Limitations

While the project achieved its primary objectives, several limitations were identified:

- **Limited Dataset Size:**
  With only 125 tiles, the diversity of glacier conditions was restricted, which may limit generalization to new geographic regions.
- **Severe Class Imbalance in PS2:**
  The extremely small pixel count for minority classes affected the stability of model predictions despite advanced balancing techniques.
- **No Temporal Analysis:**
  The models operated only on static imagery and did not account for seasonal or multi-year glacier changes.
- **Computational Constraints:**
  Training was performed on limited hardware, restricting the use of larger models or extensive hyperparameter tuning.

5.4 Future Enhancements

The findings and challenges encountered in this project open several meaningful opportunities for future work:

- **Larger and More Diverse Datasets:**
  Incorporating images from multiple regions, seasons, and sensors could improve generalization and robustness.
- **Advanced Architectures:**
  Transformer-based or hybrid CNN–ViT models may capture long-range spatial dependencies more effectively than classical U-Net designs.

- **Temporal Change Detection:**
  Extending the system to track glacier retreat, growth, or fragmentation over time would support climate research and environmental policy.
- **Automated Post-Processing:**
  Integrating learned morphological filtering could reduce noise and improve minority-class continuity.
- **Cloud Deployment:**
  Hosting the dashboard as a cloud service would enable large-scale usage by researchers, universities, and environmental agencies.

In conclusion, this project demonstrates that deep learning can serve as a powerful tool for automated glacier segmentation. With continued advancements in data availability, model architectures, and deployment methods, such systems hold strong potential to support future climate monitoring and scientific exploration.

References

1. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

2. B. W. Matthews, “Comparison of the predicted and observed secondary structure of T4 phage lysozyme,” *Biochim. Biophys. Acta*, vol. 405, no. 2, pp. 442–451, 1975.

3. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

4. F. Paul, C. Huggel, and A. Kääb, “Combining satellite multispectral image data and a digital elevation model for mapping debris-covered glaciers,” *Remote Sens. Environ.*, vol. 89, no. 4, pp. 510–518, 2004.

5. D. G. Vaughan et al., “Observations: Cryosphere,” in *Climate Change 2013: The Physical Science Basis*, IPCC Working Group I Contribution, Cambridge Univ. Press, 2013.

6. N. Abraham and N. M. Khan, “A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation,” in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2019, pp. 683–687.

7. P. Yakubovskiy, “Segmentation Models Pytorch,” GitHub, 2020. Available: https://github.com/qub Accessed: 2025.

8. A. Paszke et al., “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 8026–8037.

9. A. Abid, A. Ali, and J. Zou, “Gradio: Hassle-free Sharing and Testing of Machine Learning Models,” *arXiv preprint arXiv:1906.02569*, 2019.

10. J. Gorodkin, “Comparing two K-category assignments by a K-category correlation coefficient,” *Comput. Biol. Chem.*, vol. 28, pp. 367–374, 2004.

11. J. B. Campbell and R. H. Wynne, *Introduction to Remote Sensing*, 5th ed. New York, NY, USA: Guilford Press, 2011.

12. J. Oerlemans, *Glaciers and Climate Change*. Lisse, The Netherlands: A.A. Balkema, 2001.

13. P. D. Kraaijenbrink, J. M. Shea, and F. Pellicciotti, “Automated mapping of supraglacial lakes using deep neural networks,” *Remote Sens. Environ.*, vol. 201, pp. 345–357, 2017.

14. A. Buslaev et al., “Albumentations: Fast and Flexible Image Augmentations,” *Information*, vol. 11, no. 2, p. 125, 2020.