Preprints.org

Article

Enhanced Glacier Segmentation Using Physics-Informed Cascaded Swin-Unet (PICSw-UNet) Model

Muhammad Y.S. Dasti , Fatima Yaqoob , Laeeq Aslam , Yushan Zhou , Akram Syed , Yunsheng Zhang [*]

Article

Enhanced Glacier Segmentation Using Physics-Informed Cascaded Swin-Unet (PICSw-UNet) Model

Muhammad Y. S. Dasti [1], **Fatima Yaqoob** [1], **Laeeq Aslam** [2], **Zhou Yushan** [1], **Akram Syed** [2] **and Yunsheng Zhang** [1,3,*]

1 School of Geoscience and Infophysics, Central South University, Changsha, Hunan, China
2 School of Automation, Central South University, Changsha, Hunan, China
3 Hunan Engineering Research Center of 3D Real Scene Construction and Application Technology
* Correspondence: Zhangys@csu.edu.cn

Abstract: The escalating impacts of climate change necessitate enhanced monitoring techniques for debris covered glaciers (DCG) , a critical yet challenging subject in environmental studies. This urgency underscores the importance of developing advanced methodologies that improve precision in satellite imagery analysis. Our study introduces the Physics-Informed cascaded Swin-Unet (PICSw-UNet) model, a substantial upgrade over traditional segmentation methods such as UNet-ResNet34. By integrating domain-specific physical knowledge with deep learning, our model refines segmentation tasks effectively. Leveraging the architecture of Swin-UNet cascades, this hybrid approach incorporates physical constraints into the training process, guiding the learning algorithm towards more realistic and accurate predictions. Experimental results demonstrate the superiority of our model with improvements in key performance metrics, including Intersection over Union (IoU) and Area Under the Curve (AUC). Specifically, the Physics-Informed cascaded Swin-Unet (PICSw-UNet) achieved an IoU of 91.65% and an AUC of 96.4%, compared to 81.24% and 87.6%, respectively, for the UNet-ResNet34. Moreover, our model also shows enhanced efficiency with a reduced inference time (IT) of 0.2795 seconds. These advancements indicate promising implications for future remote sensing applications and glaciological studies, offering a promising direction for advancing image analysis methodologies.

Keywords: debris covered glaciers; glacial ice segmentation; satellite imagery analysis; deep learning; Swin-UNet; physics-informed modeling; glacier monitoring; image segmentation techniques

---

1. Introduction

Glacier studies are essential for understanding Earth's climate and its impacts, with a specific focus on debris ice detection within glaciers. Debris ice, formed by accumulated debris on glacier surfaces, significantly alters glacier behavior, influencing flow dynamics and melting patterns. Detecting debris ice is crucial for assessing glacier health and stability, which in turn informs predictions regarding water resource availability, sea-level rise and natural hazard assessments in glacierized regions. Additionally, insights gained from studying debris ice contribute to refining climate models and enhancing understanding of glacier response to climate change. Thus, comprehensive glacier studies, including debris ice detection, are fundamental for informed decision-making and effective mitigation strategies in response to ongoing climate challenges.

Recent assessments of Himalayan glaciers project a loss of up to 75% of their ice by 2100 under current warming trends. This prospective reduction in glacier volume poses a dire threat to the availability of freshwater for nearly two billion people residing in, or dependent on, the waters flowing from the Himalayan region. Additionally, the Swiss glaciers have exhibited unprecedented rates of melt, with a reduction of 10% in volume within just the past two years, highlighting the rapid impacts of climate change [1].

The accelerating rate of glacier melt affects various critical aspects of human life and the environment. It contributes to rising sea levels, which threaten coastal communities and ecosystems. The diminishing glacier mass also jeopardizes the availability of freshwater resources essential for irrigation, drinking and hydropower generation, impacting millions who rely on glacier-fed rivers for their livelihood. Furthermore, changes in glacier dynamics can disrupt ecosystems and biodiversity, affecting flora, fauna and the services these ecosystems provide. The presence of debris significantly affects how glaciers respond to climate change, creating diverse patterns in glacier behavior [2–7]. Accurately determining the extent and thickness of supraglacial debris is crucial for glacio-hydrologic models [7–9]. Neglecting the presence of debris cover can result in errors in estimating glacial retreat rates, mass loss and the longevity of glacial water resources. Current global supraglacial debris databases, such as the Randolph Glacier Inventory, suffer from limitations due to their reliance on heterogeneous glacier inventories [10].

Recent efforts have significantly improved the understanding of DCG through the use of Synthetic Aperture Radar (SAR) and optical imaging techniques [11? ,12]. Traditional methods in glacier monitoring have historically relied on pixel-based and object-based image analysis techniques [13–17]. These methods have been instrumental in remote sensing image classification and entity extraction, providing foundational frameworks for glacier segmentation.

Pixel-based image analysis (PBIA) involves the evaluation and grouping of individual pixels based on statistical clustering of their values [13,14]. Object-based image analysis (OBIA) extends the segmentation process beyond individual pixels, grouping them into near-homogeneous objects based on spectral, spatial, contextual, hierarchical and textual attributes [15]. Commonly employed statistical classifiers, such as nearest neighbor or random forest and rule-based classification methods enhance the efficacy of OBIA [17].

Glacier image segmentation is a critical task in remote sensing applications for understanding glacier dynamics and monitoring changes over time. Traditional methods for glacier image segmentation have been based on various techniques, including threshold, clustering and edge detection [18–21]. While these methods have shown promise in glacier image segmentation, they often face challenges in accurately delineating glacier boundaries, especially in complex terrain and under varying environmental conditions.

In addition to optical methods, Synthetic Aperture Radar imaging plays a crucial role in glacier surface classification [22,23]. Recent studies advocate for integrating optical and SAR imaging to achieve a more comprehensive understanding of glacier dynamics [24–27]. Machine learning methods, particularly deep learning architectures, have shown promising results in various image segmentation tasks [24–27].

Existing glacier segmentation methods face several challenges that impede their accuracy and efficiency. One notable difficulty arises from the rugged and heterogeneous nature of glacier terrains, complicating the precise delineation of glacier boundaries, especially across varied environmental conditions. Moreover, the limited availability of labeled data presents a obstacle, particularly concerning machine learning approaches. Additionally, traditional methods often necessitate manual parameter adjustment and lack resilience to noise and artifacts in the imagery, further complicating segmentation. Furthermore, deploying these techniques on memory-constrained devices for on-site, real-time processing poses additional hurdles, emphasizing the necessity for more streamlined and scalable segmentation techniques to overcome these obstacles.

This work introduces several innovative contributions to remote sensing and image segmentation, with a particular focus on debris-covered glaciers (DCG). Specifically, it proposes a cascaded Swin-UNet model enhanced by physics-informed inputs. The approach incorporates thermal infrared data as an additional channel, thereby capturing the characteristic thermal properties of DCG. Although many physics-informed neural networks (PINNs) incorporate partial differential equation (PDE) constraints into their loss functions [28,29], it is also well established to embed domain knowledge by adding physically relevant features or channels to the input data [30–33]. Here, the thermal infrared band

serves as a physics-informed channel, furnishing temperature-based cues that assist the model in distinguishing debris-covered from clean ice without explicit PDE terms. This practice aligns with broader physics-guided strategies in remote sensing, whereby physically meaningful variables are incorporated directly into the model pipeline to improve predictive accuracy [?]. Consequently, the resulting hybrid method leverages fundamental glaciological insights to enhance segmentation reliability, thereby establishing a robust, physics-informed framework within deep learning for remote sensing and glaciological analysis.

1. **Physics-Informed Integration and Cascaded Swin-UNet Architecture**: This work incorporates thermal data as a physics-informed channel in the segmentation process to leverage the distinct thermal signatures of DCG, enhancing segmentation accuracy. This is a novel application in the context of Swin-UNet architectures. Furthermore, the method employs a cascaded architecture involving two Swin-UNet models. The first model generates initial segmentation masks, which are refined by a second Swin-UNet that processes inputs augmented with thermal imaging data. This cascaded approach boosts the precision and reliability of the segmentation outputs.

2. **SVD-Based Weight Pruning**: Furthermore, it optimize the Swin-UNet using Singular Value Decomposition (SVD) to prune the model weights. This technique reduces the model's complexity and computational demands, facilitating deployment on platforms with limited processing capabilities while maintaining high segmentation accuracy.

3. **Hybrid Dice-Focal Loss Function**: In addition the use of a hybrid loss function combining Dice loss and focal loss addresses class imbalance effectively. This dual approach is particularly adept at improving the model's ability to delineate under-represented classes and enhances the overlap accuracy between predicted and actual segmentation masks.

The novelty of our approach lies in the integration of physics-informed thermal infrared data into a cascaded Swin-UNet model, which enhances the model's ability to segment debris-covered glaciers (DCG). Unlike traditional deep learning approaches that rely solely on spectral data, our method leverages the physical signatures of surface temperature to inform segmentation. This infusion of physical insight allows the model to make more accurate predictions in complex environmental conditions, setting it apart from other techniques like UNet-ResNet34 and Swin-UNet that do not incorporate such domain-specific knowledge. This rest of the manuscript is organized as follows, introduces of the study area is given in Section 2. Section 3 details the dataset and pre-processing steps, followed by Section 4, which outlines the architecture and methodology of the PICSw-UNet model. Finally, Section 5 presents the experimental results, showcasing our model's performance in comparison to existing methods.

2. Location of Study Area

The study area within the Hindu Kush Himalaya (HKH) region spans geographical coordinates ranging from 27° to 38° N latitude and 67° to 98° E longitude, encompassing a vast expanse of critical ecological and hydrological importance. Renowned for its extensive glacier coverage, including both clean-ice and DCG, this region serves as a vital component of water supply, climate regulation and ecological stability across countries such as Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal and Pakistan.

Often referred[34] to as the "Third Pole," the HKH region harbors one of the largest reserves of freshwater outside the polar areas, estimated at approximately 6000 cubic kilometers of ice. The average temperature within this region varies significantly due to its vast elevation range and topographical diversity. Winter temperatures can plummet below -20°C in higher altitudes, while summer months may see temperatures rising above 15°C in lower elevations. This temperature variability, coupled with the influences of monsoon and westerly weather systems, plays a crucial role in determining glacier health and water availability.
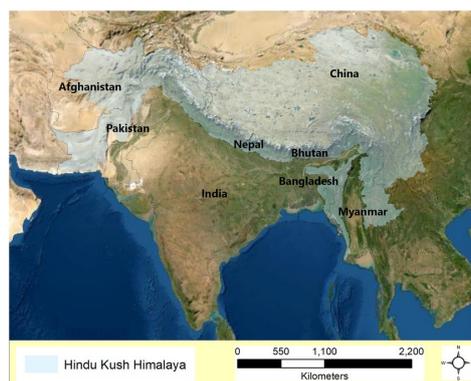
Figure 1. Study Area Map.

Debris-covered glaciers in the HKH region are particularly noteworthy due to their complex interactions with the climate. The debris layer, composed of rocks and sediment, affects the glacier's response to climatic changes by insulating the ice beneath and altering melt rates. This insulation can lead to differential melting patterns, complicating predictions of water release and glacier retreat.

The impact of global warming on the HKH glaciers is a pressing concern. Increased temperatures accelerate glacier melt, potentially resulting in the loss of a portion of the region's freshwater reserves. Such a scenario would have catastrophic implications for water availability, affecting agriculture, hydropower production and the livelihoods of millions reliant on these water resources. Furthermore, rapid glacier melting contributes to rising sea levels, posing risks to coastal communities worldwide.

The Hindu Kush Himalaya region, with its intricate mosaic of glaciers, is a critical component of the global climate system and a vital water source for millions. Understanding glacier dynamics, particularly in the face of global warming, is essential for predicting future water availability, managing resources sustainably and mitigating climate change impacts on this fragile ecosystem.

2.1. Dataset and Data Explanation

The foundational dataset utilized in this study originates from multispectral imagery taken by the Landsat 7 satellite, launched by NASA and the US Geological Survey on April 15, 1999. The Landsat 7 is equipped with the Enhanced Thematic Mapper Plus (ETM+) sensor, capturing data across various spectral bands important for environmental monitoring and research. These include visible, near-infrared, short-wave infrared and thermal infrared spectra.

The dataset specifically targets the Hindu Kush Himalaya (HKH) region, covering clean-ice and DCG. Glacier outlines were sourced from the International Centre for Integrated Mountain Development (ICIMOD) Regional Database System [35], with segmentation techniques developed by Aryal et al. [36].

2.2. Data Characteristics and Pre-Processing

The spectral bands of Landsat 7, particularly the Near-Infrared (Band 4) and Shortwave Infrared (Bands 5 and 7), are crucial for differentiating between clean and DCG. A summary of the spectral bands and their applications is provided in Table A1 of Appendix A. The pre-processing pipeline involves downloading images, upsampling thermal infrared bands, segmenting images into sub-images and normalizing pixel values to prepare for the segmentation task.

2.3. Dataset Details

The processed dataset consists of images with dimensions of $512 \times 512 \times 15$, including the following channels:

- LE7 Bands: B1 (Blue), B2 (Green), B3 (Red), B4 (Near Infrared), B5 (Shortwave Infrared 1), B6_VCID_1 (Low-gain Thermal Infrared), B6_VCID_2 (High-gain Thermal Infrared), B7 (Shortwave Infrared 2) and B8 (Panchromatic).

- Quality Bitmask (BQA) to identify the quality and usability of each pixel.
- Vegetation Index (NDVI), Snow Index (NDSI) and Water Index (NDWI) for specialized environmental analyses.
- SRTM 90 Elevation and Slope to provide topographical context which is crucial for understanding glacier dynamics.

Binary masks are generated for clean and DCG segmentation, with the black pixel representing clean ice pixels (value of 0) and the white pixel representing DCG pixels (value of 255).The sample images of 15 channels with the mask is shown in Figure 2.
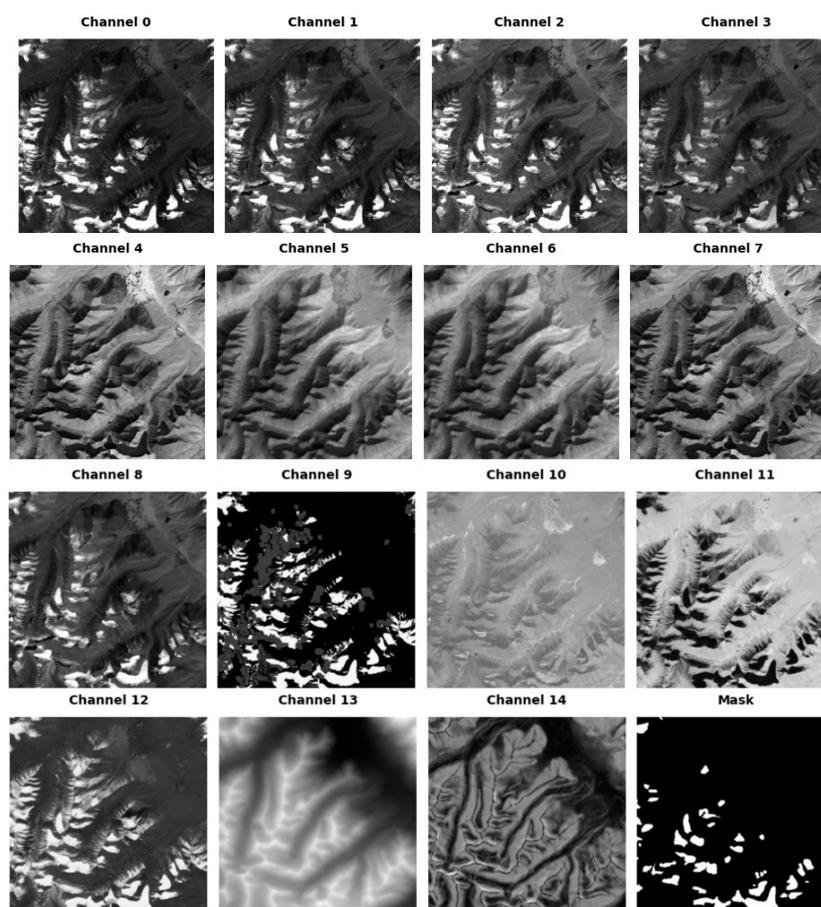

Figure 2. Sample images of 15 channels with the Mask.

3. Swin-UNet Model Architecture

Transformer models, first introduced by Vaswani et al. [37], have revolutionized natural language processing (NLP) and are increasingly applied in diverse domains. The architecture of Transformer models, characterized by self-attention mechanisms, enables parallel processing of sequences and effective handling of long-range dependencies [38]. In computer vision and image processing, the adaptation of Transformer models has led to advancements. Vision Transformers (ViT) [39] treat images as sequences of patches, allowing for comprehensive analysis of visual data. The Swin Transformer [40], further refines this approach with a hierarchical structure and shifted window self-attention, making it particularly adept at handling tasks like image segmentation. The Swin-UNet architecture represents an innovative fusion of the Swin Transformer and the traditional U-Net framework, tailored for biomedical image segmentation. This hybrid model leverages the hierarchical structure and the shifted window self-attention mechanism of the Swin Transformer as its backbone, enhancing U-Net's feature extraction and localization capabilities.

3.1. Integrating Swin Transformer as Backbone

The Swin Transformer processes input images in a hierarchical manner, allowing the Swin-UNet to handle features at multiple scales effectively. The architecture of Swin-UNet can be described as follows:

$$\text{Swin-UNet}(I) = \text{U-Net}(\text{Swin Transformer}(I)) \tag{1}$$

where I denotes the input image, Swin Transformer(I) represents the hierarchical feature extraction performed by the Swin Transformer and U-Net$(\cdot)$ indicates the segmentation processing of U-Net utilizing these features.
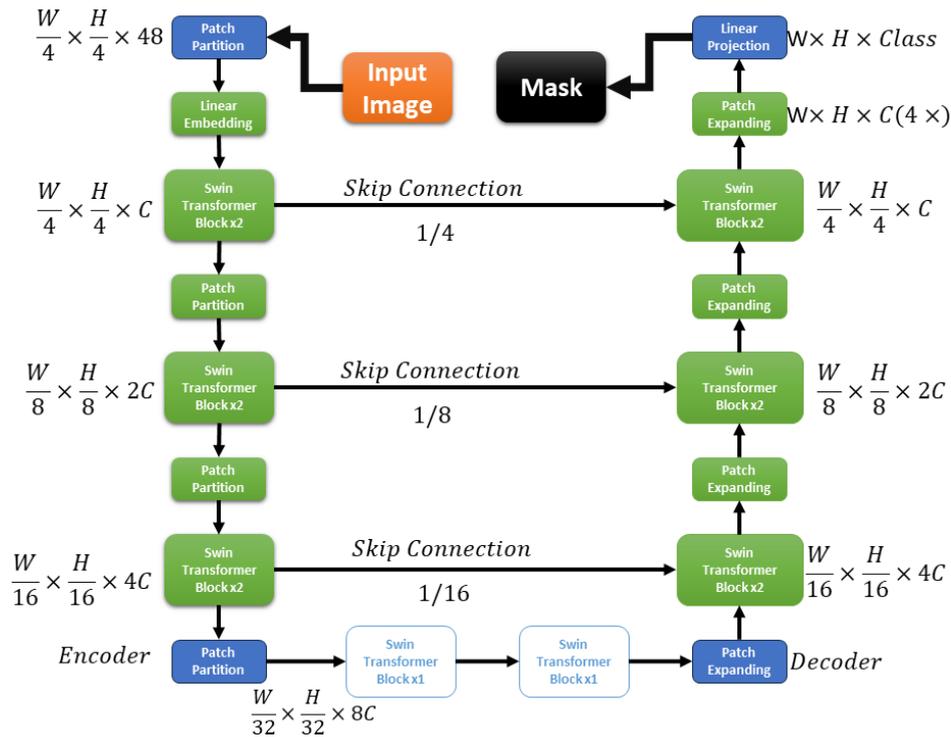

Figure 3. Architecture of Swin-UNet.

3.2. Hierarchical Feature Representation

The Swin Transformer's hierarchical design allows for efficient processing of high-resolution images. It captures features at multiple scales, facilitating accurate segmentation. The hierarchical feature representation enables Swin-UNet to understand both global context and fine details within the image, crucial for precise segmentation.

3.3. Shifted Window Self-Attention

The Swin Transformer incorporates shifted window self-attention, which limits attention computation to non-overlapping local windows. This mechanism significantly reduces computational complexity and allows the model to focus on relevant features within each window. The Swin-UNet benefits from this attention mechanism, enabling it to process large images efficiently while maintaining segmentation accuracy.

3.4. SVD-Based Optimization

Singular Value Decomposition (SVD) is a matrix factorization method that decomposes a matrix into three separate matrices: a left singular matrix, a diagonal matrix of singular values and a right singular matrix. In the context of neural networks, SVD can be used to compress model parameters by approximating the weight matrices with lower-rank matrices, thus reducing the number of parameters and model size while preserving important information.

The SVD-based optimization is a crucial step in enhancing the efficiency of the Swin-UNet model, particularly for deployment in environments with limited computational resources. By reducing the size of the weight matrices while maintaining accuracy, this technique forms an integral part of our overall segmentation strategy. With this optimization in place, we now present the proposed segmentation scheme, which builds on this foundation through a structured, multi-phase approach designed to improve the accuracy of debris-covered glacier (DCG) segmentation.

4. Proposed Segmentation Scheme

The proposed segmentation approach aims to improve DCG segmentation through a cascaded Swin-UNet model enhanced by physics-informed inputs. The proposed cascaded Swin-UNet architecture consists of two stages. The first Swin-UNet model generates an initial segmentation mask from the input data. Then, the second Swin-UNet model refines this mask by incorporating thermal infrared data, which provides additional physical insight into the temperature variations of debris-covered glaciers. This dual-phase approach improves segmentation accuracy by progressively enhancing the model's understanding of the glaciological context. To enhance the segmentation of debris-covered glaciers, we integrate thermal infrared data, which provides unique thermal signatures revealing the insulating effects of the debris. This additional information is critical for distinguishing between clean ice and debris-covered regions, improving the model's ability to accurately segment glacial features based on their distinct temperature profiles.

The methodology is structured into four sections. The first section discusses data pre-processing. In the second section, the Swin-UNet model is trained using focal loss and selected channels from the dataset. During the third phase, the trained model is optimized through singular value decomposition. In the final phase, a cascaded Swin-UNet model incorporates a physics-informed thermal band from the dataset to generate a predicted mask. This mask is subsequently compared with the actual mask to evaluate the results.The overall block diagram of the proposed model is shown in Figure 4.
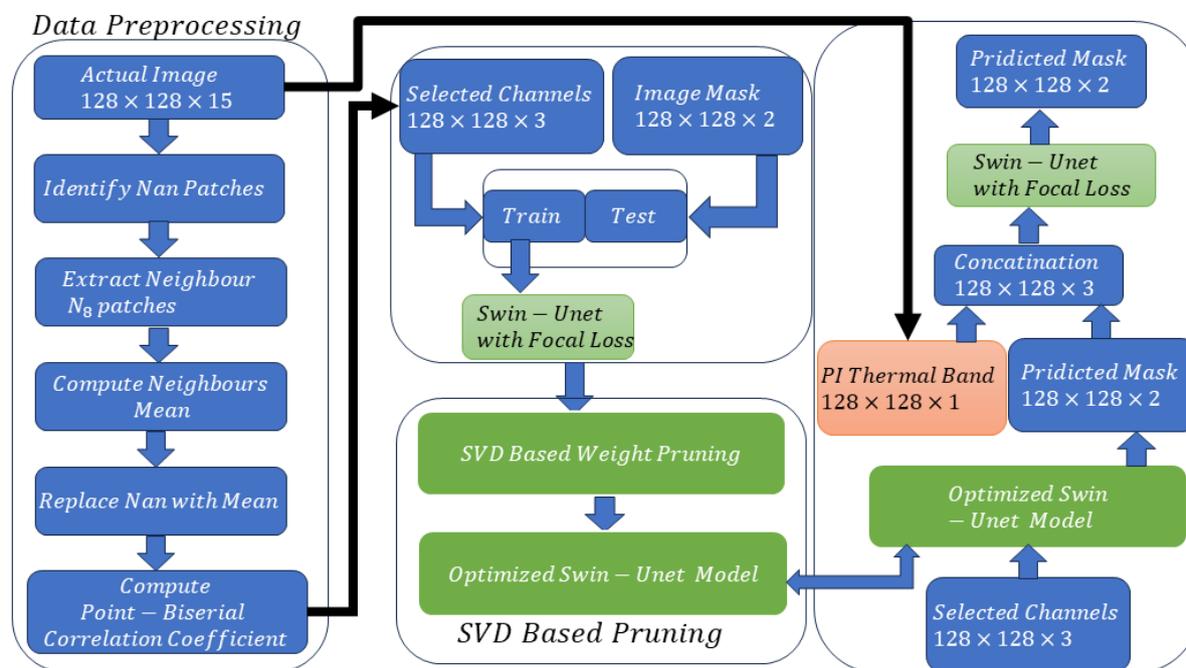

Figure 4. Block diagram of the proposed segmentation scheme.

4.1. Data Pre-Processing

In this phase, the original dataset, which has already been divided into non-overlapping window sizes of $128 \times 128 \times 15$, is passed through multiple steps to enhance data representation and identify the most significant channels in the dataset for DCG prediction. A detailed discussion of each step is provided in the following subsections.

4.1.1. Initial NaN Value Replacement

Firstly, images from the dataset, which contain multiple NaN values of various patch sizes across the 15 channels, are processed to ensure data integrity. If a NaN patch is detected, the mean values of the eight neighboring pixels (N8 neighbors) of the same patch size are computed and used to replace the NaN values. This step ensures that no valuable spatial information is lost due to missing values.

$$\text{replaced value} = \frac{1}{N} \sum_{i=1}^{N} \text{value}_i \qquad (2)$$

where N is the number of neighboring pixels used in the computation.

4.1.2. Channel Selection Using Point-Biserial Correlation

For selecting the most relevant channels, the Point-Biserial Correlation Coefficient is employed instead of the traditional Point-Biserial correlation coefficient. While Point-Biserial correlation is effective for measuring the linear relationship between two continuous variables and is often used for feature selection, it is not suitable for scenarios where the target variable is binary. In glacier segmentation, our target variable is binary (glacier vs. non-glacier), necessitating an alternative approach. The Point-Biserial Correlation Coefficient measures the strength and direction of the association between a continuous variable and a binary variable, making it more appropriate for this context:

$$r_{pb} = \frac{\overline{X}_1 - \overline{X}_0}{\sigma_X} \sqrt{\frac{n_1 n_0}{n^2}} \qquad (3)$$

where $\overline{X}_1$ and $\overline{X}_0$ are the means of the continuous variable for the binary categories, σ_X is the standard deviation of the continuous variable and n_1 and n_0 are the sample sizes for the two categories.

4.2. Swin-UNet Model Training

After selecting the most relevant channels—specifically, the top three channels that demonstrate the highest correlation with changes in the masks—the entire dataset with selected channels is partitioned into training and testing datasets. Eighty percent of the data is designated for training the model.

The Swin-UNet architecture, previously described, employs a hybrid Dice-Focal loss function to effectively manage the imbalance prevalent in classes of the dataset. This combined loss function integrates the Dice loss, which is crucial for optimizing the overlap between the predicted and actual masks, with the focal loss, which is tailored to enhance focus on difficult, misclassified examples. The focal component of the loss is mathematically defined as in Equation (4).

$$L_{focal}(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \qquad (4)$$

where p_t is calculated by the Equation (5).

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases} \qquad (5)$$

In these equations, p represents the probability of the class with label $y = 1$, α_t is a class weighting factor and γ is a focusing parameter. The Dice component is calculated using the formula for the Dice coefficient, enhancing the model's ability to improve the prediction accuracy by maximizing the overlap between the predicted and true segmentations.

Additionally, the incorporation of the Diace score further enhances the model's capability to handle class imbalance. The Diace score, similar to the Dice coefficient, measures the overlap between the predicted and actual masks but includes a factor to adjust for class imbalances. The Diace score is defined as follows:

$$\text{Diace} = \frac{2|P \cap T|}{|P| + |T| + \epsilon} \tag{6}$$

where P is the set of predicted positives, T is the set of true positives and ϵ is a small constant to avoid division by zero.

The combined Diace-Focal loss function, which integrates the Diace score into the focal loss, is defined as:

$$L_{diace\_focal}(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) + \beta(1 - \text{Diace}) \tag{7}$$

where β is a balancing factor that regulates the influence of the Diace score within the focal loss framework.

Training the model with this Diace-Focal loss function allows for an effective balance between precision in class overlap and addressing class imbalance, focusing specifically on the challenges of detecting and delineating DCG, a significantly under-represented class in the dataset. This approach ensures that both common and rare classes are accurately modeled, enhancing overall predictive accuracy. The model then progresses to the optimization phase, where further adjustments are made to enhance its performance.

4.3. SVD-Based Model Weight Pruning

A fundamental component of transformer models is the Multi-Head Attention (MHA) mechanism, which assigns attention scores to different segments of an image. This mechanism increases the size of a model, which can slow down inference time(IT) and pose challenges for deployment on edge devices where memory is limited. To address these issues, this work incorporates an SVD-based weight pruning mechanism designed to evaluate the importance of the weights in the attention matrix of the Swin Transformer blocks used in this network.

The attention mechanism in Swin Transformers can be represented by matrix A, which is derived from the query (Q), key (K) and value (V) matrices to the shifted window strategy. The attention scores in Swin Transformers are computed as follows:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \tag{8}$$

where d_k is the dimensionality of the keys and B is a relative position bias that is unique to each pair of windows in Swin Transformers, enhancing the model's ability to preserve local information across different window configurations.

The SVD-based pruning process begins by decomposing the attention score matrix A for each head in the MHA using Singular Value Decomposition, allowing for the expression of A as:

$$A = U\Sigma V^T \tag{9}$$

where U and V are orthogonal matrices representing the left and right singular vectors and Σ is a diagonal matrix containing the singular values. These singular values provide a measure of the importance of the connections in the attention matrix.

By retaining only the top k singular values—and their corresponding vectors in U and V—the complexity of A is reduced, focusing on the most significant interactions within the matrix. This reduced form, A_k, is given by:

$$A_k = U_k\Sigma_k V_k^T \tag{10}$$

where U_k, Σ_k and V_k^T are the truncated versions of U, Σ and V^T, containing only the top k components.

This pruning reduces the number of parameters and accelerates inference by focusing computational resources on the most impactful elements of the attention mechanism. This technique is especially valuable in environments with strict memory constraints, such as edge computing devices.

4.3.1. Proof of Optimality of Weights Extracted by SVD

The optimality of the truncated Singular Value Decomposition (SVD) in the context of model weight pruning can be demonstrated using the Eckart-Young-Mirsky theorem. This theorem provides a foundational understanding that the best rank-*k* approximation to a matrix, in terms of the least squares error, is given by its SVD truncated to the first *k* singular values.

The theorem states:

$$\min_{B:\text{rank}(B)=k} \|A - B\|_F^2 = \|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2 \tag{11}$$

where:

- $\|.\|_F$ denotes the Frobenius norm.
- $A_k = U_k \Sigma_k V_k^T$ is the approximation of A using the top k singular values.
- σ_i are the singular values of A.
- r is the rank of A.

Derivation: The Frobenius norm of the difference between A and its rank-*k* approximation A_k can be expanded as follows:

$$\begin{aligned}
\|A - A_k\|_F^2 &= \text{trace}((A - A_k)^T (A - A_k)) \\
&= \text{trace}(A^T A - A^T A_k - A_k^T A + A_k^T A_k) \\
&= \text{trace}(A^T A) - 2\text{trace}(A^T A_k) + \text{trace}(A_k^T A_k).
\end{aligned}$$

Since A_k is formed by keeping only the top k singular values (and corresponding singular vectors), it holds that $A^T A_k = A_k^T A_k$ and thus:

$$\begin{aligned}
\|A - A_k\|_F^2 &= \text{trace}(A^T A) - \text{trace}(A_k^T A_k) \\
&= \sum_{i=1}^{r} \sigma_i^2 - \sum_{i=1}^{k} \sigma_i^2 \\
&= \sum_{i=k+1}^{r} \sigma_i^2.
\end{aligned}$$

This derivation shows that A_k is indeed the optimal rank-*k* approximation of A under the Frobenius norm criterion, thereby justifying the use of SVD-based pruning to maintain model performance while reducing computational requirements.

4.4. Cascaded Swin-UNet Model

In this work the existing data is given to the already trained and optimized swin-unet model let's denote the output from the Swin-UNet model as O, which consists of two channels:

$$O = [O_{\text{clean}}, O_{\text{debris}}] \tag{12}$$

where O_{clean} represents the channel for clean ice and O_{debris} represents the channel for DCG. Each channel is a two-dimensional matrix corresponding to the spatial dimensions of the input images.

The thermal band, denoted as T, is a single-channel image of the same spatial dimensions as O. This thermal band is selected for its capability to reflect thermal variations indicative of different material properties on the ice surface, particularly highlighting areas with debris due to their typically higher temperatures from lower albedo.

The concatenation operation integrates the model outputs with the thermal data and is mathematically represented as:

$$\text{new image} = \text{concatenate}(O, T) \tag{13}$$

This operation merges the spatial and contextual information provided by the Swin-UNet with the thermal insights from the physics-informed thermal band. The resulting image, now enriched with both visual and thermal cues, significantly enhances the model's ability to generalize across different conditions and datasets. This composite image serves as the input for subsequent processing stages aimed at refining and enhancing the segmentation accuracy further. Algorithm 1.

Algorithm 1 Proposed Model for Debris Ice Segmentation

Input: Dataset D with images I having 15 channels, each with NaN values
Output: Segmentation masks M for clean and DCG
procedure PREPROCESSDATA(D)
    for each image $I \in D$ **do**
        for each channel C_k in I **do**
            if NaN patches in C_k **then**
                $C_k \leftarrow \text{mean}(N8 \text{ neighbors})$
            end if
        end for
    end for
end procedure
procedure SELECTCHANNELS(D)
    for each image $I \in D$ **do**
        $\rho_k \leftarrow \text{corr}(C_k, M)$ ▷ Point-Biserial correlation
        Select C_{top3} with highest ρ_k
    end for
end procedure
procedure INITIALSEGMENTATION(C_{top3})
    $O \leftarrow \text{Swin-UNet}(C_{\text{top3}})$ ▷ 2 channel output
end procedure
procedure INTEGRATETHERMALBAND(O)
    $N_{\text{new}} \leftarrow \text{concat}(O, \text{thermal band})$
end procedure
procedure REFINESEGMENTATION(N_{new})
    $M \leftarrow \text{Swin-UNet}(N_{\text{new}})$
end procedure

The newly formed images from Equation (13) serve as input to another Swin-UNet model. This second model is specifically trained on these images, learning to refine the segmentation masks further. The combination of the thermal band with outputs from the first model substantively improves segmentation accuracy, as it provides additional contextual information relevant to the physical properties of the ice.The pseudo code of proposed scheme is shown in Table 1.

$$\text{refined output} = \text{Swin-UNet}(\text{new image}) \tag{14}$$

Table 1. Distribution of the dataset into training, testing and validation sets.

Data Split	Number of Images
Training	564,823
Testing	105,942
Validation	35,314

5. Experiment

This section details the results of the proposed segmentation scheme. Initially, the comparison metrics are defined. Then the results of the first phase of the proposed scheme are then explored, followed by the results of the proposed cascaded Swin-UNet model are presented. Then the results of the proposed scheme are compared with other models.

5.1. Comparison Metrics

The performance of the segmentation model is evaluated using several metrics designed to provide a comprehensive assessment of its effectiveness. These metrics include Accuracy, F1 Score, Precision, Recall, Intersection over Union (IoU) and Area Under the Curve (AUC). Below, each metric is defined and explained:

(1) **Accuracy:** This metric measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. It is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

(2) **F1 Score:** The harmonic mean of Precision and Recall, the F1 Score provides a balance between them, particularly useful when dealing with an uneven class distribution. It is defined as:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

(3) **Precision:** Also known as positive predictive value, Precision measures the accuracy of positive predictions made by the model. It is computed as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

(4) **Recall:** Also known as sensitivity, Recall indicates the ability of the model to identify all relevant instances. It is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

(5) **Intersection over Union (IoU):** Also known as the Jaccard index, IoU measures the overlap between the predicted segmentation and the ground truth. It is defined as:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{19}$$

(6) **Area Under the Curve (AUC):** Refers to the area under the Receiver Operating Characteristic (ROC) curve. It quantifies the overall ability of the model to discriminate between the classes across all thresholds. A higher AUC value indicates better model performance.

5.2. Experimental Setup

A sliding window operation augments the dataset by generating overlapping sub-images of size 128×128 by shifting the window by 12 rows or columns for each step, resulting in a total of 1089 sub-images. The dataset is divided into training, testing and validation sets with distributions of 80%, 15% and 5%, respectively.

This setup ensures effective training and evaluation, utilizing Landsat 7's spectral capabilities and segmentation techniques. With this foundation in place, next section disscuss focus to the Transformer models that drive our segmentation approach.

5.3. Pre-Processing Results

In the implemented scheme, the dataset initially contains NaN values. These are addressed by replacing them with the mean of adjacent pixels. Subsequently, a correlation test is conducted to determine the most significant channels within the dataset for debris ice segmentation. The analysis identifies key channels that notably impact glacier segmentation: NDWI (water index), NDSI(Normalized Difference Snow Index) and LE7 B1 (Blue) channels. These channels are then chosen for constructing the input images for the model, effectively reducing the input dimensionality to $128 \times 128 \times 3$. The results of the correlation tests are depicted in Figure 5.
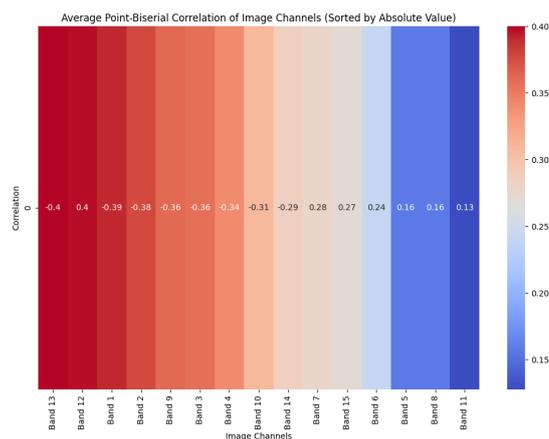

Figure 5. Point-Biserial Correlation Matrix

5.4. Results of Proposed Model

This section presents the performance results of the PICSw-UNet model, highlighting its effectiveness in debris ice segmentation tasks. The model's metrics are visualized in a radar chart shown in Figure 7, which allows for an immediate and intuitive comparison across various critical performance measures.

The model's **Accuracy (95.67%)** indicates a high level of overall correctness across all classes, which is essential for reliable predictions in real-world applications. High accuracy signifies that the model is capable of correctly classifying most of the pixels in the input images, whether they belong to DCG or clean ice. This is particularly important in environmental monitoring, where accurate segmentation can inform critical decisions and interventions.

The **F1 Score (95.37%)** balances precision and recall, providing a comprehensive view of the model's ability to accurately classify debris without mis classifying clean ice. The F1 score is crucial in scenarios where both false positives (mis classifying clean ice as debris) and false negatives (missing actual debris) can have significant implications. A high F1 score demonstrates that the model maintains a good balance between precision and recall, effectively minimizing both types of errors.

Precision (94.23%) highlights the model's capability to correctly label actual debris, which is crucial in scenarios where false positives are costly. Precision is particularly important in environmental monitoring and disaster response, where incorrectly identifying clean ice as debris could lead to unnecessary actions or missed opportunities for intervention. High precision ensures that when the model identifies debris, it is likely to be correct.

Recall (96.53%) demonstrates the model's effectiveness in identifying all actual debris instances, vital for environmental monitoring to ensure no DCG is missed. Recall is critical in ensuring that all areas of concern are accurately detected. High recall indicates that the model successfully identifies most, if not all, of the DCG areas, which is essential for comprehensive monitoring and analysis.

The **IoU (Intersection over Union) (91.65%)** represents the accuracy of overlap between predicted and actual debris regions, which is important for precise segmentation. IoU is a measure of the model's ability to predict the exact boundaries of debris-covered areas. A high IoU indicates that the

predicted segmentation closely matches the ground truth, which is vital for applications requiring precise delineation of DCG.

AUC (Area Under the Curve) (96.4%) shows the model's discriminative power, indicative of robustness against varying operational thresholds. A high AUC value indicates that the model is capable of distinguishing between DCG and clean ice across a range of threshold settings. This robustness is important for adapting the model to different operational conditions and ensuring consistent performance.

Lastly, the **IT (0.2795 seconds)** is critical for evaluating the model's efficiency, particularly in time-sensitive applications. Low IT ensures that the model can process and analyze images quickly, which is crucial for real-time monitoring and decision-making scenarios. Efficient IT allows the model to be deployed in operational environments where timely information is essential.

The visual results of the proposed scheme are shown in Figure 6, providing a clear representation of the model's segmentation capabilities. These visualizations help in understanding how well the model performs in distinguishing between debris-covered and clean ice.
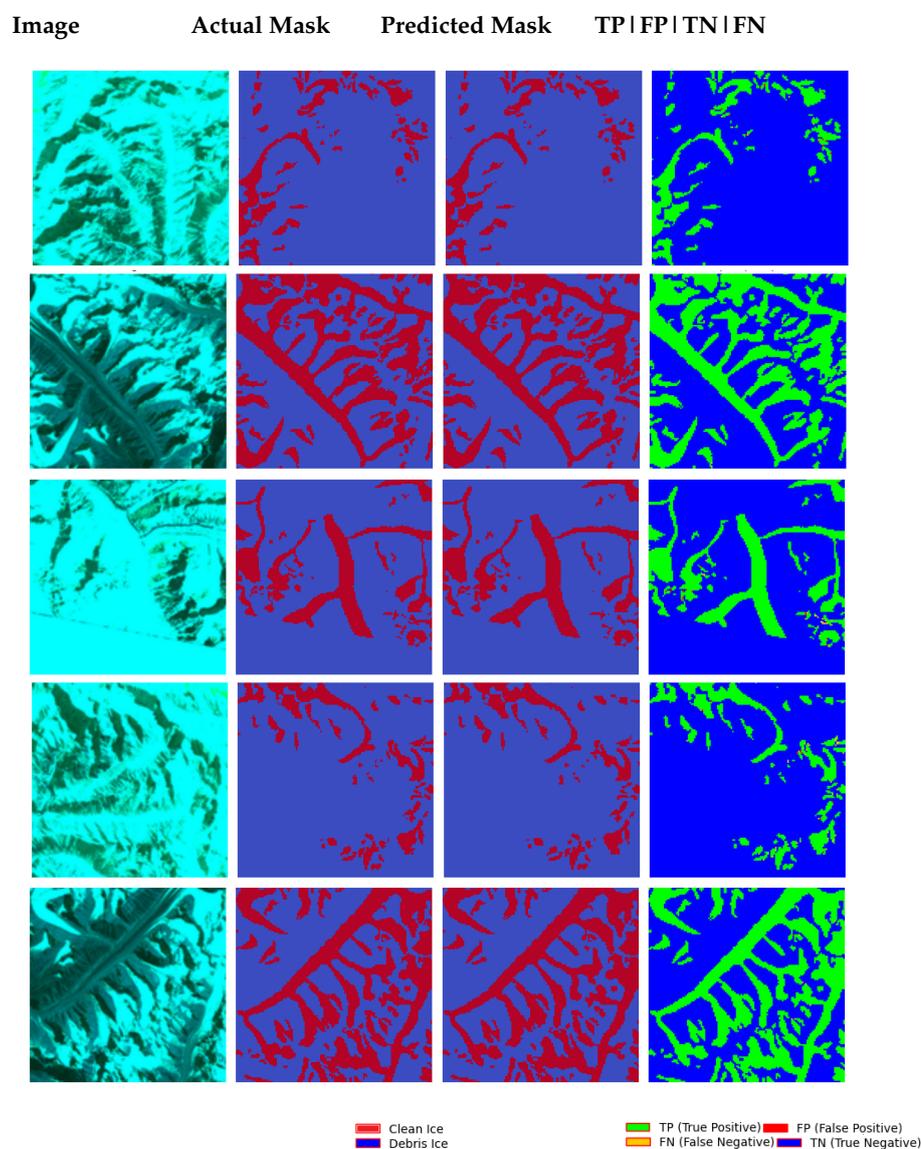
Figure 6. Overview of the SPSW-UNet(SPSW-UNet) results. Each row, from top to bottom, showcases the input image, the original mask, the prediction by SPSW-UNet and the analysis of correct classifications.
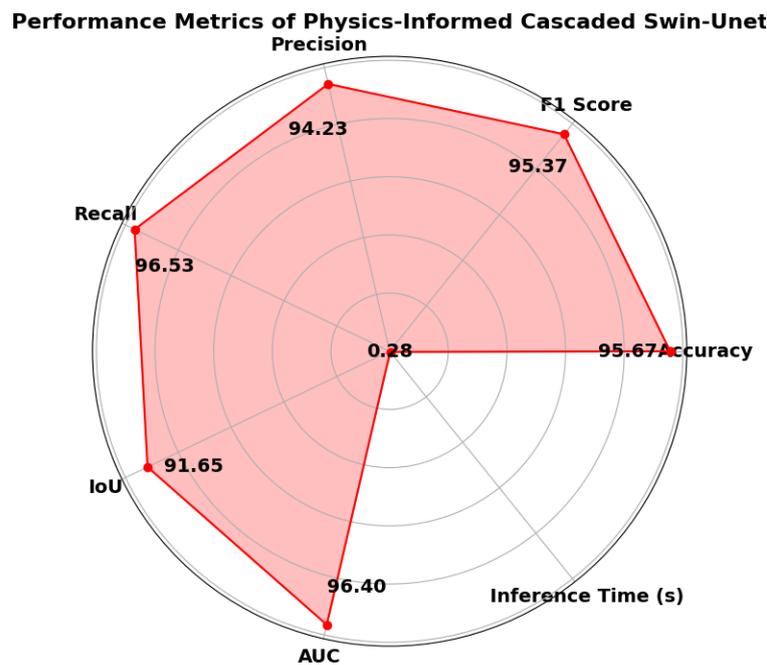
Figure 7. Radar Chart of the PICSw-UNet Performance Metrics

The visual results shown in Figure 6 illustrate the practical effectiveness of the PICSw-UNet model. Each row, from top to bottom, showcases the input image, the original mask, the prediction by the model and the analysis of correct classifications. The clear separation between debris-covered and clean ice regions demonstrates the model's high precision and recall. The legend helps in interpreting the different classifications, providing a comprehensive view of the model's performance in real-world scenarios. In some instances, false positives (FP) occurred in regions where debris features closely resembled clean ice, particularly in areas with light haze or shadow. Similarly, false negatives (FN) were observed in regions with sparse debris coverage or minimal thermal contrast, where the model struggled to distinguish between clean ice and small debris patches. These errors highlight the need for further refinement in model sensitivity to subtle environmental features. In summary, the PICSw-UNet model demonstrates significant improvements in segmentation accuracy and efficiency. The high values of accuracy, F1 score, precision, recall, IoU and AUC, combined with a low IT, highlight the model's effectiveness and practicality for real-world applications. These results underscore the model's potential to enhance DCG segmentation, providing more accurate and reliable data for environmental monitoring and analysis.

6. Results and Discussion

The performance of the PICSw-UNet is compared against three other models: Swin-UNet, svd-pruned swin-unet (SPSW-UNet) and UNet-ResNet34, as shown in Table 2. This comparative analysis aims to highlight the enhancements achieved by integrating physics-informed strategies into the cascaded model. The PICSw-UNet model demonstrates substantial improvements across several key performance metrics when compared to conventional models such as Swin-UNet, SPSW-UNet and UNet-ResNet34.

The model's **accuracy** of 95.67% is a significant improvement over the Swin-UNet (94.12%), SPSW-UNet (92.34%) and UNet-ResNet34 (89.9%). This high level of accuracy indicates that the PICSw-UNet model can correctly classify a large proportion of the pixels in the input images, reducing the overall error rate. High accuracy is crucial in environmental monitoring where precise data interpretation is necessary for effective decision-making.

The **F1 score** of 95.37% for the PICSw-UNet surpasses that of the Swin-UNet (93.65%), SPSW-UNet (93.35%) and UNet-ResNet34 (92.05%). The F1 score balances precision and recall, making it a reliable metric for assessing the model's performance, especially in scenarios with imbalanced class distributions. This high F1 score demonstrates that the proposed model can accurately identify DCG while minimizing both false positives and false negatives.

Precision measures the model's ability to correctly identify debris without misclassifying clean ice. The PICSw-UNet achieves a precision of 94.23%, which is higher than the Swin-UNet (93.09%), SPSW-UNet (92.92%) and UNet-ResNet34 (92.2%). High precision is essential in applications where false positives can lead to unnecessary interventions or actions and this improvement underscores the model's capability to make accurate predictions regarding DCG.

Recall for the PICSw-UNet is 96.53%, outperforming the Swin-UNet (94.21%), SPSW-UNet (93.80%) and UNet-ResNet34 (91.91%). Recall is crucial for ensuring that the model detects as many true positive instances of DCG as possible. High recall indicates that the model is effective in identifying all relevant instances, reducing the risk of overlooking areas that require monitoring or intervention.

Intersection over Union (IoU) is a metric used to evaluate the accuracy of segmentation by measuring the overlap between the predicted and actual debris regions. The PICSw-UNet achieves an IoU of 91.65%, significantly higher than the Swin-UNet (86.73%), SPSW-UNet (84.89%) and UNet-ResNet34 (81.24%). A higher IoU indicates better segmentation accuracy, which is crucial for precise delineation of DCG.

The **Area Under the Curve (AUC)** of the PICSw-UNet is 96.4%, higher than the Swin-UNet (93.93%), SPSW-UNet (91.20%) and UNet-ResNet34 (87.6%). A higher AUC indicates better discriminative power of the model across various threshold settings. This robustness is important for adapting the model to different operational conditions and ensuring consistent performance.

The **IT** for the PICSw-UNet is 0.2795 seconds, which is slower compared to Swin-UNet (0.1865 seconds) and SPSW-UNet (0.1247 seconds) but faster than UNet-ResNet34 (0.3059 seconds). While the increased IT is a drawback, it is a reasonable trade-off given the substantial improvements in other performance metrics. The slight increase in IT is justified by the model's enhanced accuracy and robustness, making it suitable for applications where precision is more critical than speed.

Table 2. Comparative Analysis of the Cascaded Model Performance.

Model	Acc.	F1	Pre.	Rec.	IoU	AUC	IT (s)
Swin-UNet	94.12%	93.65%	93.09%	94.21%	86.73%	93.93%	0.1865
SPSW-UNet	92.34%	93.35%	92.92%	93.80%	84.89%	91.20%	0.1247
UNet-ResNet34	89.9%	92.05%	92.2%	91.91%	81.24%	87.6%	0.3059
PICSw-UNet	95.67%	95.37%	94.23%	96.53%	91.65%	96.4%	0.2795

The PICSw-UNet model exhibits significant improvements over traditional and SVD-pruned models. From the Swin-UNet, there is an improvement of +1.65% in accuracy, +1.84% in F1 score, +1.2% in precision, +2.46% in recall, +5.67% in IoU, +2.63% in AUC and a slower inference by 49.87%. Comparatively, against the SPSW-UNet, it shows +3.61% better accuracy, +2.16% in F1 score, +1.41% in precision, +2.91% in recall, +7.96% in IoU, +5.70% in AUC and a significant increase in IT by 124.14%. Moreover, improvements over the UNet-ResNet34 are even more pronounced with +6.42% in accuracy, +3.61% in F1 score, +2.20% in precision, +5.03% in recall, +12.81% in IoU, +10.05% in AUC and a reduction in IT by 8.63%, indicating a notable enhancement in both performance and efficiency. The inclusion of the thermal infrared channel significantly contributed to these improvements, as it allows the model to better capture the distinct thermal signatures of debris-covered glaciers (DCG). This thermal information helped the model distinguish between clean ice and debris-covered regions with greater accuracy. For instance, the model's Intersection over Union (IoU) improved by 5.67%, and Area Under the Curve (AUC) increased by 2.63%, highlighting the contribution of the thermal data to the enhanced segmentation precision. These results indicate that incorporating thermal data as

a physics-informed feature plays a critical role in refining the model's predictions, particularly in environments where temperature variations are key to class differentiation.

The advancements underscore the model's robustness and the efficacy of incorporating physics-informed insights to substantially elevate segmentation accuracy and operational efficiency. The pronounced improvements in key performance metrics such as Intersection over Union (IoU) and Area Under the Curve (AUC) specifically illustrate the model's enhanced ability to distinguish between different classes more distinctly and handle complex segmentation tasks with greater precision. This indicates a significant step forward in the model's design, where the integration of physical principles directly contributes to its computational intelligence, enabling more nuanced and accurate predictions.

The PICSw-UNet model significantly surpasses both traditional and SPSW-UNet models across all evaluated metrics, with notable advancements in IoU and AUC, metrics that are critical for assessing the precision and efficacy of segmentation tasks. These improvements are crucial for applications that rely on the accurate categorization and segmentation of images, such as environmental monitoring and glaciological studies, where precision in interpreting satellite imagery can directly impact the quality of subsequent analyses and decisions. The model demonstrates remarkable gains in accuracy, which reflects the overall correctness of the model across all test scenarios. The F1 score, which balances the precision and recall, underscores the model's ability to correctly classify and recall instances of debris and clean ice without significant errors, providing a more reliable and consistent performance. These metrics highlight not only the model's accuracy but also its practical effectiveness in real-world applications.

While the increase in IT, compared to both the original and the SVD-pruned models, is noticeable, it is reasonable given the substantial enhancements in model performance. This trade-off is considered acceptable in many advanced applications where the stakes of decision-making and the cost of errors are high. Furthermore, the PICSw-UNet model shows improved IT compared to the UNet-ResNet34 model, indicating its enhanced efficiency in processing despite the higher complexity. This is particularly advantageous in resource-limited scenarios where computational efficiency is as critical as accuracy.

This detailed comparative analysis confirms the benefits of integrating physics-informed modifications into deep learning frameworks. It highlights the model's potential to transform the field of environmental monitoring. By incorporating physical laws into the training process, the model gains an intrinsic understanding of natural phenomena, enabling it to perform with high reliability under varied and challenging conditions. This approach is essential in the field of remote sensing, where accurate data interpretation is pivotal for environmental conservation and disaster management strategies.

In short, these enhancements contribute to a robust, reliable and efficient model capable of performing complex segmentation tasks crucial for scientific research and practical applications in various fields, particularly those involving environmental studies and climate monitoring. The PICSw-UNet model represents a significant advancement in the use of machine learning to address and interpret natural phenomena, setting a new benchmark for future developments in the field.
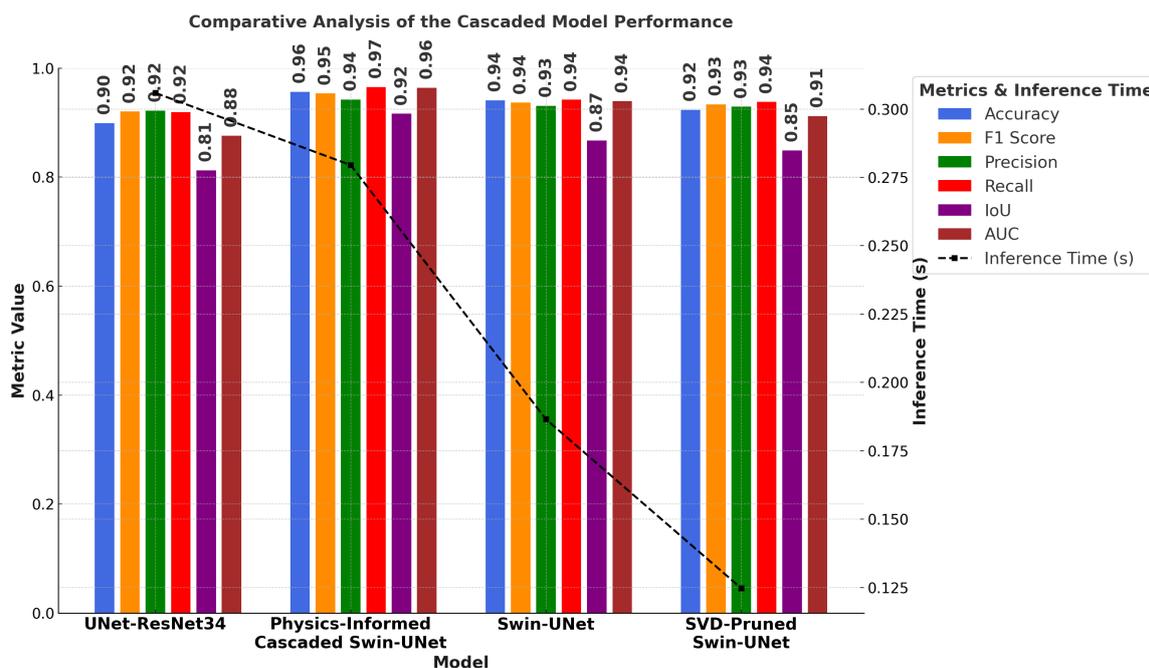
Figure 8. Graphical Representation of Comparative Analysis

7. Conclusions

The presented study introduces the PICSw-UNet model, advancing glacier segmentation techniques. By integrating domain-specific physical knowledge with a deep learning framework, our model achieves better performance metrics in segmenting DCG from satellite imagery. The evaluation of our model against traditional segmentation methods, such as Swin-UNet, SPSW-UNet and UNet-ResNet34, shows its effectiveness and robustness. Key performance metrics, including accuracy (95.67%), F1 score (95.37%), precision (94.23%), recall (96.53%), Intersection over Union (91.65%) and Area Under the Curve (96.4%), show improvements over the baseline models. Specifically, the PICSw-UNet shows an IoU and AUC, indicating its ability to delineate and accurately classify DCG. Although the model incurs a slightly higher IT (0.2795 seconds), this is a reasonable trade-off considering the gains in segmentation precision and accuracy. This study has many contributions, advancing glacier segmentation and offering a scalable solution for environmental monitoring and glaciological studies. The precision and reliability of the PICSw-UNet model provide insights into glacier dynamics, crucial for understanding climate change impacts and informing mitigation strategies. In conclusion, this study shows the potential of integrating physics-informed modeling with deep learning architectures to achieve improvements in complex environmental segmentation tasks. The findings highlight the importance of interdisciplinary approaches in enhancing the accuracy and efficiency of remote sensing applications, paving the way for future research and development in this area.

Appendix A

Table A1. Spectral Bands of Landsat 7 and Their Applications in Glacier Segmentation

Band	Wavelength (μm)	Resolution (m)	Primary Use
1	0.45 - 0.52	30	Visible (Blue)
2	0.52 - 0.60	30	Visible (Green)
3	0.63 - 0.69	30	Visible (Red)
4	0.76 - 0.90	30	Near-Infrared
5	1.55 - 1.75	30	Shortwave Infrared 1
7	2.08 - 2.35	30	Shortwave Infrared 2
6	10.40 - 12.50	60	Thermal Infrared
8	0.52 - 0.90	15	Panchromatic

References

1. Deutsche Welle. Swiss Glaciers Record Unprecedented Melt Rates. https://www.dw.com/en/swiss-glaciers-melt/a-57483958, 2023. Accessed: 2024-04-24.

2. Sakai, A.; Fujita, K. Contrasting glacier responses to recent climate change in high-mountain Asia. *Scientific Reports* **2017**, *7*, 13717. https://doi.org/10.1038/s41598-017-14256-5.

3. Quincey, D.J.; Glasser, N.F. Morphological and ice-dynamical changes on the tasman glacier, New Zealand, 1990–2007. *Glob. Planet. Change* **2009**, *68*, 185–197. https://doi.org/10.1016/j.gloplacha.2009.05.003.

4. Jiang, S.; Nie, Y.; Liu, Q.; Wang, J.; Liu, L.; Hassan, J.; et al. Glacier change, supraglacial debris expansion and glacial lake evolution in the gyirong river basin, central Himalayas, between 1988 and 2015. *Remote Sens.* **2018**, *10*, 986. https://doi.org/10.3390/rs10070986.

5. Mölg, N.; Bolch, T.; Walter, A.; Vieli, A. Unravelling the evolution of zmuttgletscher and its debris cover since the end of the little ice age. *Cryosphere* **2019**, *13*, 1889–1909. https://doi.org/10.5194/tc-13-1889-2019.

6. Tielidze, L.G.; Bolch, T.; Wheate, R.D.; Kutuzov, S.S.; Lavrentiev, I.I.; Zemp, M. Supra-glacial debris cover changes in the Greater Caucasus from 1986 to 2014. *Cryosphere* **2020**, *14*, 585–598. https://doi.org/10.5194/tc-14-585-2020.

7. Herreid, S.; Pellicciotti, F. The state of rock debris covering Earth's glaciers. *Nat. Geosci* **2020**, *13*, 621–627. https://doi.org/10.1038/s41561-020-0615-0.

8. Ragettli, S.; Immerzeel, W.W.; Pellicciotti, F. Contrasting climate change impact on river flows from high-altitude catchments in the Himalayan and Andes Mountains. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9222–9227. https://doi.org/10.1073/pnas.1606526113.

9. Rounce, D.R.; King, O.; Mccarthy, M.; Shean, D.E.; Salerno, F. Quantifying debris thickness of debris-covered glaciers in the everest region of Nepal through inversion of a subdebris melt model. *J. Geophys. Res. Earth Surf.* **2018**, *123*, 1094–1115. https://doi.org/10.1029/2017JF004395.

10. Racoviteanu, A.; Nicholson, L.; Glasser, N. Surface composition of debris-covered glaciers across the Himalaya using linear spectral unmixing of Landsat 8 OLI imagery. *Cryosphere* **2021**, *15*, 4557–4588. https://doi.org/10.5194/tc-15-4557-2021.

11. Kulkarni, S.C.; Rege, P.P. Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion* **2020**, *59*, 13–29.

12. Begam, S.; Sen, D. Mapping of moraine dammed glacial lakes and assessment of their areal changes in the central and eastern Himalayas using satellite data. *Journal of Mountain Science* **2019**, *16*, 77–94.

13. Ardelean, F.; Ginzler, C.; Gartner-Roer, I.; Götz, J. Glacier classification and mapping by object-based image analysis (OBIA), using optical data from different sensors: A case study from the Suntar-Khayata Range, Russian Siberia. *Remote Sensing* **2011**, *3*, 782–800.

14. Bajracharya, S.R.; Shrestha, B.R. Automated mapping of debris-covered glaciers using object-based image analysis of Landsat data: A case study of Yala Glacier in Langtang Valley, Nepal. *Geocarto International* **2015**, *30*, 636–651.

15. Karimi, F.; Bahremand, A.; Roodposhti, M.S.; et al.. Detection of supraglacial lakes on debris-covered glaciers in the Himalaya/Karakoram using Landsat imagery. *Journal of Applied Remote Sensing* **2015**, *9*, 096098.

16. Nie, Y.; Sheng, Y.; Wang, Y.; et al.. A semi-automated object-based approach for glacier mapping. *Journal of Geographical Sciences* **2010**, *20*, 415–428.

17. Rastner, P.; Bolch, T.; Notarnicola, C.; Paul, F. A comparison of pixel- and object-based glacier classification with optical satellite images. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **2014**, *7*, 853–862. https://doi.org/10.1109/JSTARS.2013.2274668.

18. Smith, J.D.; Johnson, E.F. Glacier segmentation from satellite imagery. *Remote Sensing Letters* **2007**, *3*, 307–316.

19. Jones, M.A.; Smith, J.D. Comparative analysis of clustering algorithms for glacier segmentation. *International Journal of Remote Sensing* **2010**, *31*, 5417–5434.

20. Johnson, E.F.; Wang, L.; Liu, W. Automated glacier boundary extraction from aerial photographs using edge detection algorithms. *Photogrammetric Engineering & Remote Sensing* **2015**, *81*, 159–169.

21. Wang, L.; Liu, W. Deep glacier segmentation using convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56*, 218–233.

22. Nagai, H.; Abe, T.; Ohki, M. SAR-based flood monitoring for flatland with frequently fluctuating water surfaces: Proposal for the normalized backscatter amplitude difference index (NoBADI). *Remote Sensing* **2021**, *13*, 4136.

23. Muhammad, S.; Thapa, A. Daily Terra–Aqua MODIS cloud-free snow and Randolph Glacier Inventory 6.0 combined product (M* D10A1GL06) for high-mountain Asia between 2002 and 2019. *Earth System Science Data* **2021**, *13*, 767–776.

24. Bhardwaj, A.; et al. Synthesizing optical and SAR imagery from land cover maps and auxiliary raster data. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1–12.

25. Singh, S.; et al. A sensitivity analysis on the spectral signatures of low-backscattering sea areas in Sentinel-1 SAR images. *Remote Sensing* **2021**, *13*, 1183.

26. Corcione, V.; Buono, A.; Nunziata, F.; Migliaccio, M. A sensitivity analysis on the spectral signatures of low-backscattering sea areas in Sentinel-1 SAR images. *Remote Sensing* **2021**, *13*, 1183.

27. Stonevicius, E.; Uselis, G.; Grendaite, D. Ice Detection with Sentinel-1 SAR Backscatter Threshold in Long Sections of Temperate Climate Rivers. *Remote Sensing* **2022**, *14*, 1627.

28. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **2019**, *378*, 686–707.

29. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nature Reviews Physics* **2021**, *3*, 422–440.

30. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.F.; Kumar, V. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering* **2017**, *29*, 2318–2331.

31. Karpatne, A.; Watkins, W.; Read, J.; Kumar, V. Physics-guided neural networks (PGNN): An application in lake temperature modeling. *arXiv preprint arXiv:1710.11431* **2018**.

32. Aslam, L.; Zou, R.; Awan, E.; Butt, S.A. Integrating Physics-Informed Vectors for Improved Wind Speed Forecasting with Neural Networks. In Proceedings of the 2024 14th Asian Control Conference (ASCC). IEEE, 2024, pp. 1902–1907.

33. Morales, R.; Vidal, J. A physics-informed neural network for wind field retrieval from SAR images. *Remote Sensing* **2020**, *12*, 2507.

34. Talukder, B.; Matthew, R.; Bunch, M.J.; Hipel, K.W.; Orbinski, J.; et al. Melting of Himalayan glaciers and planetary health. *Current Opinion in Environmental Sustainability* **2021**, *50*, 98–108.

35. Bajracharya, S.R.; Shrestha, B.R. The Status of Glaciers in the Hindu Kush-Himalayan Region **2011**. International Centre for Integrated Mountain Development (ICIMOD), https://doi.org/10.53055/ICIMOD.55.

36. Aryal, B.; Miles, K.E.; Vargas Zesati, S.A.; Fuentes, O. Boundary Aware U-Net for Glacier Segmentation. *Journal of Remote Sensing or Conference Proceedings Name* **2023**, *Volume Number*, Page Range. https://doi.org/10.7557/18.6789.

37. Vaswani, A.; et al.. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30, 2017.

38. Devlin, J.; et al.. BERT: Bidirectional Encoder Representations from Transformers. In Proceedings of the Proceedings of NAACL-HLT 2019, 2018.

39. Dosovitskiy, A.; et al.. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929* **2020**.

40. Liu, Z.; et al.. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv preprint arXiv:2103.14030* **2021**.