Environmental Data Science (2026), 5: e38, 1–24
doi:10.1017/eds.2026.10065

CAMBRIDGE UNIVERSITY PRESS

METHODS PAPER

Advancing supraglacial lake mapping in Antarctica: A deep learning approach using Sentinel-2 imagery

Jonas Koehler[1,2], Celia Baumhoer[1], Bert Wouters[3], Stef Lhermitte[3,4] and Andreas Dietz[1]

[1]Landsurface Dynamics, German Remote Sensing Data Center, German Aerospace Center, DLR, Germany
[2]Department of Glaciology and Climate, Geological Survey of Denmark and Greenland (GEUS), Denmark
[3]Department of Geoscience and Engineering, Faculty of Civil Engineering and Geoscience, TU Delft, Netherlands
[4]Department of Earth and Environmental Sciences, KU Leuven, Belgium
Corresponding author: Celia Baumhoer; Email: celia.baumhoer@dlr.de

Received: 25 June 2025 UTC; **Revised:** 28 May 2026 UTC; **Accepted:** 05 August 2026 UTC

Keywords: Antarctica; deep learning; remote sensing; supraglacial lakes; U-Net

Abstract

With the ongoing effects of global warming, Antarctic supraglacial meltwater plays a critical role in ice-sheet instability. Spaceborne remote sensing provides a viable solution to detect surface melt in these vast and remote regions. Innovative deep learning image segmentation methods have the potential to improve supraglacial lake (SGL) mapping by adding spatial awareness to the mapping process, as currently established methods only consider the spectral variety of multispectral satellite data (Sentinel-1 and Landsat-8) on a per-pixel basis. In this study, we adapt a U-net image segmentation method to detect SGLs in Sentinel-2 data and quantify its performance in comparison to established mapping approaches in Antarctica. At a 90% probability threshold, the proposed workflow achieves a producer's accuracy, user's accuracy, and an F1 score of 0.915, 0.912, and 0.913, respectively. We compare the classification results to those of an established Random Forest classifier, and find that the U-net achieves better results in 13 of the 16 validation scenes. In a relative comparison with five established SGL products, the U-net approach maps, on average, the largest lake area, although there is disagreement between the datasets due to different data sources, cloud cover, and misclassification of some clouds as lakes. The main advantages of the developed approach are its capabilities to map lakes under thin clouds and ice, resulting in more continuous lake areas than conventional approaches, and a reduced number of input features required, which will be advantageous in future large-scale SGL mapping efforts.

Impact Statement

Detecting supraglacial lakes (SGLs) on Antarctica's ice sheets is important for glacier mass balance modeling and sea level rise projections. The presence of SGLs can have a negative influence on ice shelf stability by inducing hydrofracture and causing flexure, and reduces firn air content and surface albedo, which facilitates the formation of new surface water bodies. Due to the vastness and remoteness of the area, information about SGLs is commonly derived in automated workflows from large amounts of multispectral satellite imagery. While there are several established mapping approaches, the advent of DL for image segmentation offers a new methodology that enables the inclusion of spatial information into formerly pixel-wise image classification workflows. Our newly trained U-net DL model outperforms a traditional random forest-based mapping approach and maps by

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© DLR, 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

averaging the largest lake area in a relative comparison with five conventional methods. Although potential improvements still remain, once trained, the method is easily applicable on larger scales and thus lays the foundation for comprehensive circum-Antarctic SGL mapping efforts that can improve glacier mass balance modeling and sea level rise estimations in the context of global warming.

1. Introduction

Surface meltwater in Antarctica has been increasingly discussed as a determining factor in ice shelf instability and hence global sea level rise due to global warming (Lai et al., 2020; Gilbert and Kittel, 2021; Van Wessem et al., 2023; De Roda Husman et al., 2024). Meltwater ponding within crevasses can promote downward crack propagation, potentially leading to hydrofracture (Bell et al., 2018; Robel and Banwell, 2019; Lai et al., 2020). In addition, the filling and drainage of melt ponds, together with the redistribution of meltwater across the ice shelf, can induce flexure and thereby weaken the structural integrity of the shelf (Banwell et al., 2019). Both processes may result in through-ice fractures that can ultimately trigger the disintegration of entire ice shelves, as observed for the Larsen B Ice Shelf collapse (Banwell et al., 2013). Meanwhile, refreezing meltwater promotes the formation of ice lenses and firn densification. These processes inhibit water percolation and storage, increase surface runoff, and thereby enhance the formation of new surface water bodies (Hubbard et al., 2016; Machguth et al., 2016; Buzzard et al., 2018). This, in turn, has direct implications for the energy balance of ice shelves, as the resulting reduction in surface albedo further amplifies melt rates (Arthur et al., 2020). In Antarctica, only a few studies have observed a coupling between surface meltwater systems and the subglacial hydrological system of grounded ice, for example, via moulins (Bell et al., 2018; Tuckett et al., 2019). However, observations from Greenland have shown that increased meltwater input from supraglacial lakes (SGLs) into the subglacial drainage system can enhance basal sliding and accelerate glacier flow (Tedesco et al., 2013; Wehrlé et al., 2026). A similar scenario may become possible in Antarctica with the continued inland migration of SGLs (Dirscherl et al., 2021b). In combination with the reduced buttressing effect of already disintegrated ice shelves, this could further accelerate ice flow and exacerbate mass loss on connected glaciers (Royston and Gudmundsson, 2016; Rott et al., 2018). Projections suggest a doubling of current surface melt rates in Antarctica by 2050 (Trusel et al., 2015), with likely implications for global sea level rise.

The spatial extent, remoteness, and inaccessibility make it difficult to quantify surface meltwater extent and volume using in-situ measurements, particularly on Antarctica's vast ice sheets. Therefore, the use of spaceborne remote sensing data to map meltwater surface features such as lakes, streams, and water-saturated firn and snow ("slush") has become increasingly important in recent years (e.g., Stokes et al., 2019; Dirscherl et al., 2020; Moussavi et al., 2020b; Corr et al., 2022; Dell et al., 2024). Spaceborne remote sensing provides spatially continuous, large-scale measurements at resolutions on the order of tens of meters. In polar regions, the usual revisit frequency of several days is improved by overlapping swaths and can be further enhanced by combining different sensors. In addition, the long-term continuity of satellite missions allows the analysis of multiyear time series. Therefore, a variety of approaches have been developed to derive the extent and volume of surface meltwater in polar regions. Studies relying on multispectral satellite data often use a threshold-based SGL detection approach. Earlier studies mostly used single-band reflectance thresholds (Sneed and Hamilton, 2007; Liang et al., 2012; Moussavi et al., 2016; Williamson et al., 2018) or the ratio between the visible blue and red channels (Box and Ski, 2007; Pope et al., 2016; Banwell, 2017). Other thresholding approaches depend on one (Morriss et al., 2013; Yang and Smith, 2013; Stokes et al., 2019) or multiple (Burton-Johnson et al., 2016) spectral indices such as normalized differences between two spectral channels. Recent studies commonly utilize the approach developed by Moussavi et al. (2020b), which combines the normalized difference water index (NDWI), the normalized difference snow index (NDSI), and various thresholds on single spectral bands. This method exploits the light attenuation of SGLs and the resulting reflectance differences in the green and near-infrared, and blue and red spectral bands,

respectively. Using a physical model, it is also possible to calculate lake depth and volume. This approach has been applied with high classification accuracy to ice shelves in West Antarctica and the Antarctic Peninsula (Corr et al., 2022), East Antarctica (Spergel et al., 2021; Tuckett et al., 2021; Arthur et al., 2022; Trusel et al., 2022), and the entire continent (Shen et al., 2025). However, because optical data are prone to misclassification in the presence of slush, rocks, clouds, and shadows, careful post-processing of classification results is required (Corr et al., 2022). This, along with limited transferability of static thresholds to other regions, makes upscaling this method to larger spatial scales challenging (Tuckett et al., 2021; Dell et al., 2022).

This problem has recently been addressed by machine learning-based image classification and segmentation approaches. For the first time, Dirscherl et al. (2020) used a Random Forest classifier (RF) and demonstrated its usefulness for SGL mapping from Sentinel-1 data over a variety of regions in Antarctica. Hu et al. (2022) trained an RF with Sentinel-2-derived NDWI and textural features to map SGLs over the entire Greenland ice sheet. The good overall accuracy of the results over such a large area indicates good transferability of the method, although an application in Antarctica is still pending. The issue of distinguishing slush from ponded water in Landsat data is addressed by Dell et al. (2022, 2024), who trained an RF to separate these two spectrally similar classes in Antarctica.

While threshold-based and RF approaches heavily rely on the spectral information of optical datasets, some studies have also explored using spatial features by employing convolutional neural networks for image segmentation, more specifically modified versions of the original U-net by Ronneberger et al. (2015). For multispectral data, this neural network was applied for the first time by Lutz et al. (2023) to detect SGLs in Sentinel-2 imagery over two glaciers in Greenland. The U-net's focus on spatial information enabled the study to limit the provided spectral information for model training to only three bands (red, blue, and green) and nonetheless achieved a high classification accuracy. Meanwhile, the full spectral information of Sentinel-2 was used by Niu et al. (2023) for SGL mapping in six consecutive melting seasons over the Amery ice shelf in Antarctica. Different artificial neural network approaches were compared, with an attention-based U-net achieving the best results. Moreover, studies that attempted mapping SGLs from synthetic aperture radar (SAR) backscatter data have shown that a U-net is well suited to this task even on data with low spectral variety (Chen, 2021; Dirscherl et al., 2021a; Jiang et al., 2022). This indicates that including spatial features into the modeling reduces the dependence on spectral features and might reduce the amount of data needed in large-scale mapping efforts based on multispectral data. The combination of spectral information and spatial context makes the U-net an appealing classifier for optical data, where other classification approaches have difficulties in discriminating spectrally similar classes such as slush from SGLs. If provided with a wide variety of training scenes containing different surface features, this approach has the potential for pan-Antarctic SGL mapping. Therefore, we investigate how much a deep neural network enhances SGL mapping accuracy compared to traditional and machine learning (ML) methods.

In this article, we present an automated mapping workflow that uses a U-net to detect SGLs in multispectral Sentinel-2 data. In particular, we build on the work by Dirscherl et al. to advance their SGL mapping approach for optical data (Dirscherl et al., 2020) by adapting and evaluating a U-net architecture that was originally implemented for the same purpose for SAR data (Dirscherl et al., 2021a, 2021b). As demonstrated in this work, this can potentially reduce the amount of required input data and thus streamline the processing for large-scale mapping efforts. Currently, for multispectral data, 18 features (spectral bands plus derived indices) are required to train an RF model (Dirscherl et al., 2020). We argue that the U-net's awareness of the spatial context will reduce the amount of spectral information required. We also aim to explore potential improvements of our approach in discriminating spectrally similar classes, for example, shallow lakes and slush, and contextual features, such as ice cover, lake edges, or cloud shadows. To this end, we (i) generate high-precision training and validation labels of SGLs on Sentinel-2 imagery of a wide selection of Antarctica's marginal ice shelves and glaciers; (ii) train and evaluate different configurations of a Deep Learning (DL) model to detect SGLs in Sentinel-2 data;

(iii) comprehensively compare our results with an ML approach by Dirscherl et al. (2020) using manually labeled validation data; and (iv) compare the DL-based classification results to established SGL mapping datasets.

This article is organized as follows: Section 2 describes the datasets, training and test sites, model architecture, validation approaches, and comparison schemes used. Section 3 presents the results of the accuracy evaluation of the DL approach in direct comparison to its RF predecessor, as well as other SGL mapping products. Section 4 discusses the results in terms of improvements and drawbacks of the proposed approach compared to the original and other products. Section 5 summarizes the results and provides an outlook for future research.

2. Materials and methods

In this article, we develop a DL-based SGL detection approach for Sentinel-2 data (Figure 1). First, training and testing data are acquired, then the most suitable spectral features to detect SGLs are determined. Using these data, a U-net model is tuned, trained, and evaluated. The final model is then used for large-scale inference of SGLs in Antarctica. The materials and methods used in this workflow will be presented in detail below.

2.1. Datasets, training, and testing sites

To derive SGLs, we used data from the Sentinel-2 mission (European Space Agency, 2022). This constellation of three satellites carries a passive Multispectral Instrument (MSI) that records reflectance values in 13 spectral bands at a spatial resolution of up to 10 meters. Sentinel-2 data have been available since 2015 with a revisit time of 5 days, which is effectively higher at the poles due to overlapping swaths. For this study, we use Level-1C Top-of-Atmosphere reflectance data (European Space Agency, 2022).

For model training and testing, we selected a variety of testing sites distributed across Antarctica to reflect regional variability. We selected sites where SGL occurrence had been reported before in the literature. We aimed for scenes acquired during the melting season in austral summer (November to March) where SGL numbers and extents peak. However, we had to prioritize scenes with an abundance of lakes, so the selected data covered the months of December to February, with a strong focus on January. For the same reason, most of the training and testing data are from the strong melting seasons of 2019/2020 and 2020/2021, but the years 2017, 2018, 2023, and 2024 were also covered. Furthermore, the training and testing data reflect a variety of environmental settings, including diverse surface features next to lakes, such as slush, blue ice, rock outcrops, clouds and cloud shadows, moraines, and crevasses (Figure 2). It was also ensured that the large variety of SGL types was reflected, such as deep and shallow lakes, ice cover, oval and elongated lakes, as well as ice.

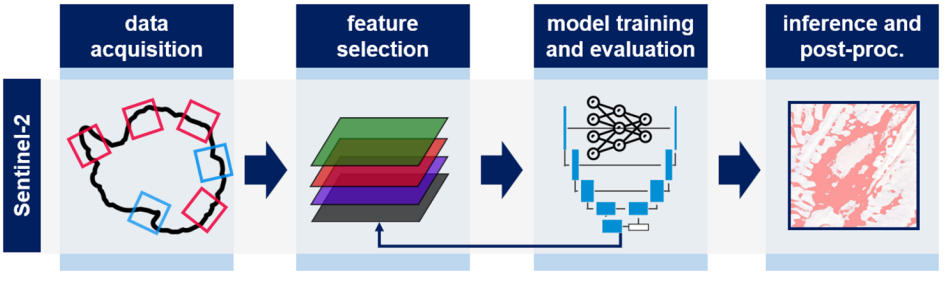


Figure 1. *Conceptual workflow of the proposed Sentinel-2-based supraglacial lake detection approach. Figure contains elements from https://www.freepik.com/.*

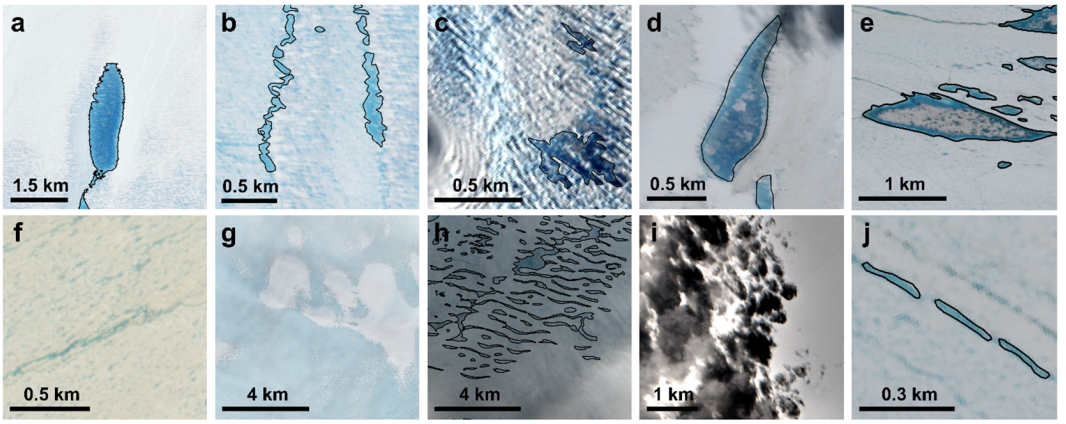


Figure 2. *Visual examples of various surface features included in the train and test datasets. (a) Oval lakes at Roi Baudouin, (b) shallow lakes at Roi Baudouin, (c) lake among crevasse fields at Shirase Glacier, (d) thin ice cover at Amundsen, (e) a thick ice lid at Amery, (f) slush at Riiser, (g) blue ice at Nivlisen, (h) lakes under clouds at George VI, (i) dark clouds over Amundsen, and (j) thin streams at Amery. Background: Copernicus Sentinel-2 Data.*

For model training, 16 Sentinel-2 scenes were acquired, in which 33 training areas were defined. The performance of the model was validated on 8 Sentinel-2 scenes containing 16 testing sites (Figure 3 and Supplementary Table S1). The cloud cover percentage of the training scenes ranges from 0 to 94.9, while the mean sun elevation angle varies between 20.2 and 36.1 degrees. In the testing scenes, the cloud cover percentages ranges from 0 to 54.6 and the mean sun elevation angle from 26.8 to 38.3 degrees (Supplementary Table S1).

Additionally, the 90–m TanDEM-X PolarDEM digital elevation model (DEM) was used to mask out steep slopes (> 5%), where lake formation is highly unlikely, and high altitudes during post-processing (Wessel et al., 2021). To mask out areas seaward of the Antarctic coastline, we use the IceLines product to determine the ice shelf front positions in coastal areas (Baumhoer et al., 2023). To compare the performance of the proposed U-net-based mapping approach, we use existing SGL maps generated by Dirscherl et al. (2020) where available or otherwise created them ourselves using the original RF-based workflow.

We selected five SGL datasets that use optical satellite imagery to map supraglacial lakes in Antarctica (Table 1), which we compared to the mapping results of our proposed approach (see Section 2.3). The dataset by Stokes et al. (2019) uses both Landsat 8 and Sentinel-2 data to map SGLs on the margins of the East Antarctic ice sheet (EAIS) for January 2017. Lakes were automatically detected by thresholding the NDWI, which calculates the normalized difference of the green and near-infrared (NIR) spectral bands. The SGL dataset by Moussavi et al. (2020b) covers 20 regions across Antarctica and the years 2013–2020. It is derived from Landsat 8 data by thresholding an NDWI variant that uses the blue and red spectral channels, as well as differences between green and red, and blue and green, respectively, to identify potential lake areas. Further thresholding using the NDSI (green and short-wave infrared bands) and a ratio between brightness temperature and the blue spectral band masks out clouds, rocks, and seawater. It is noteworthy that that even though the original mapping approach combines Landsat 8 and Sentinel-2 data (Moussavi et al., 2020b), the published dataset itself is based only on Landsat 8 (Moussavi et al., 2020a). This approach and its thresholds were adapted and modified by Corr et al. (2022) to include both above-mentioned NDWI variants. The resulting dataset covers the West Antarctic Ice Sheet (WAIS) and Antarctic Peninsula (API) for January 2017 and, furthermore, discriminates between lakes and channels. Finally, we included the SGL and slush dataset by Dell et al. (2022, 2024), which covers 57 ice

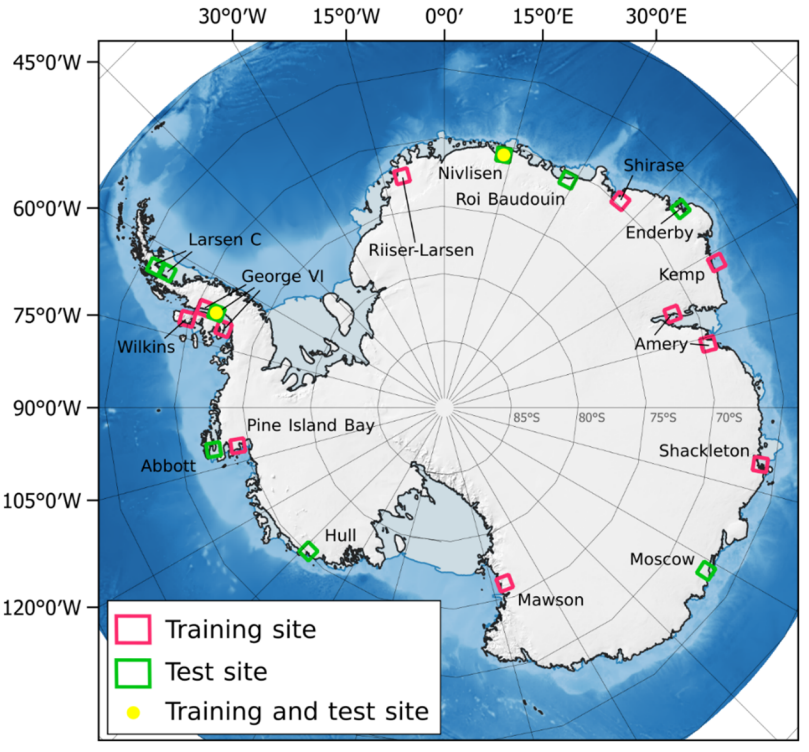


Figure 3. *Locations of the training and testing sites. The map is displayed in Polar Stereographic projection. Base maps provided by the Antarctic Digital Database (https://add.scar.org/).*

Table 1. *SGL datasets and their characteristics used in the relative comparison scheme*

Dataset	Spatial coverage	Method	Sensors
U-net	API, EAIS, WAIS	U-net	Sentinel–2
Stokes et al. (2019)	EAIS	NDWI thresholding	Landsat 8, Sentinel–2
Moussavi et al. (2020a)	API, EAIS, WAIS	NDWI/NDSI thresholding	Landsat 8, (Sentinel–2)
Dirscherl et al. (2020)	API, EAIS, WAIS	Random Forests	Sentinel–2
Corr et al. (2022)	WAIS, API	Modified Moussavi et al. (2020)	Landsat 8, Sentinel–2
Dell et al. (2024)	API, EAIS, WAIS	NDWI thresholding, Random Forests	Landsat 8

Note. The column spatial coverage indicates the Antarctic region (partially) covered by the dataset: API, Antarctic Peninsula; EAIS, East Antarctic Ice Sheet; WAIS, West Antarctic Ice Sheet.

shelves across Antarctica between 2013 and 2021 and is derived from Landsat 8 data. The method considers areas above a fixed NDWI (blue and red channels) threshold, which are then clustered in an unsupervised classification scheme. Among others, the classes “ponded water” and “slush” were determined, and a sample of pixels from each class was used to train an RF classifier. Finally, we used the method described by Dirscherl et al. (2020) to create SGL maps for the accuracy assessment (Section 2.2.4) and the relative dataset comparison (Section 2.3). In this approach, an RF classifier is trained on 18 spectral features such as spectral bands and indices, including both variants of the NDWI to detect SGLs.

Another SGL dataset worth mentioning is that by Arthur et al. (2022). It covers the EAIS between 2014 and 2020, and is derived from Landsat 8 data. However, since the classification workflow closely follows that by Moussavi et al. (2020b), we did not include this.

2.2. Supraglacial lake mapping

2.2.1. Label generation and data preprocessing

There is no universally accepted definition of a "supraglacial lake" in the existing literature, to our knowledge. As a basis for this study, we adopt the classification scheme of Dell et al. (2022) who distinguish stored surface meltwater on Antarctic ice shelves as either ponded water (i.e., lakes) or meltwater retained within firn pore spaces. This distinction implicitly separates supraglacial lakes from other surface hydrological features such as streams and slush. Building on this framework, we introduce several modifications and caveats to establish consistent training labels for the target class "supraglacial lake" across different scenes and geographical settings.

First, we only consider lakes that form on ice (Supplementary Table S2 [a]). For some ice-marginal lakes, it is not possible to determine unequivocally whether the water body is situated on bedrock or ice; however, if the lake is in direct contact with the ice surface, we classify it as an SGL (Supplementary Table S2 [p]). Second, we consider the morphology of water bodies within supraglacial hydrological networks. Narrow channels connecting wider ponded areas are excluded (Supplementary Table S2 [g]), whereas broad, stream-like features are included, as the distinction between elongated lakes and wide streams is often arbitrary and such features can retain substantial volumes of meltwater over extended periods (Supplementary Table S2 [h]). Third, we only classify lake areas that are visible in the imagery. This also includes lakes that can be identified through spatial context, such as lakes that are faintly visible beneath thin cloud cover (Supplementary Table S2 [o]) or ice (Supplementary Table S2 [b]–[d]). We argue that the spatial awareness of the tested segmentation model facilitates the detection of such features.

Based on these considerations, a concise rule set was developed in which labeling guidelines for a variety of scenarios (e.g., ice cover, minimum lake size, streams, lakes under clouds, etc.) were formulated. These rules are documented with examples in Supplementary Table S2. In experiments, we found that a reduced number of consistent high-precision labels generated following this rule set resulted in significantly higher classification accuracies than an increased amount of less consistent training data, for example, generated through automated thresholding.

Using this ruleset, we generated fully segmented training and testing scenes by manually digitizing SGLs in the selected training and testing regions. SGL areas, as per our definition above, were labeled "1" (the target class) and everything else "0". Next, the acquired Sentinel-2 L1C data were atmospherically corrected to L2A Bottom-Of-Atmosphere (BOA) reflectance using the Sen2Cor processor (Pignatale, 2022). Digital numbers (DNs) were converted to reflectance values using the appropriate scaling factors and offsets for the respective processing baseline of the product. Missing no data values (DN = 0) inside the image acquisition strip of the scene were spatially interpolated. For both the training and testing scenes, image stacks were generated that contained the spectral bands in the visible green and red (Sentinel-2 bands 3 and 4), as well as the Soil/Water Index (SWI) (Dirscherl et al., 2020) and the Automated Water Extraction Index ($AWEI_{nsh}$ [Feyisa et al., 2014]) (Table 2). These bands and indices were selected as they were previously found to have high predictive power for SGL detection (Dirscherl et al., 2020) and our empirical experiments showed that the $AWEI_{nsh}$ in particular reduced misclassifications in areas with topographic shadows (Supplementary Figure S1), while the SWI is useful for the discrimination of rocks and moraines from water bodies. The shortwave infrared bands (SWIR1 and SWIR2) were resampled for this purpose to match the 10 m resolution of bands 3 and 4.

The training data consists of binary label masks and corresponding imagery containing information on the spectral bands and indices. This dataset was used for model training and internal validation to iteratively optimize the model parameters.

First, the training data were tiled into 480 × 480 pixel tiles with an overlap of 100 pixels that could be fed into the model training pipeline. The resulting stack of 658 tiles was subdivided depending on the

Table 2. *Spectral features included in the training of the Sentinel-2-based SGL detection model*

No.	Band/index	Description/formula	Reference
1	B03	Visible green (560 nm)	https://sentiwiki.copernicus.eu/web/s2-mission
2	B04	Visible red (665 nm)	
3	SWI	(Blue–SWIR1)/(Blue+SWIR1)	Dirscherl et al. (2020)
4	$AWEI_{nsh}$	4 × (Green–SWIR1)–(0.25×NIR) + (2.75×SWIR2)	Feyisa et al. (2014)

number of labeled lakes visible within a tile, that is, in stacks containing no lakes, few lakes, and many lakes. The threshold between few and many lakes was drawn at the 80th percentile, that is, tiles with an amount of more than 15,041 lake pixels were considered "lake-rich." To increase the number of training samples and address the imbalance of the classes "lake" and "no lake," we applied an augmentation strategy to the training data (Dirscherl et al., 2021a). Tiles containing lakes were flipped, shifted, and rotated to various degrees, with tiles containing numerous lakes being augmented 12 times and tiles with fewer lakes 8 times. Additionally, tiles containing particularly difficult surface features such as clouds and dark topographic shadows were augmented 12 times regardless of whether they contained any lakes. The resulting tiles, along with a 3% random selection of lake-less tiles, comprised the final stack of training data containing a total of 3,420 tiles. For each of the four bands of the training imagery, we calculated statistics such as the minimum, maximum, mean, and standard deviation across all tiles. Then, each tile and band was normalized using *z*-score normalization. The final training data were then subdivided into 80% training data and 20% validation data for the internal model validation, resulting in stacks of 2,736 labels and images for training and 684 for validation. It is noteworthy that labels and imagery used in the accuracy assessment (Section 2.2.4) were not used during model training.

2.2.2. U-net model architecture and training

The aim of this article is to analyze how a deep-learning image segmentation approach can improve SGL detection.

The U-net, which has been used for image segmentation of various surface features on a variety of Earth observation datasets before, typically consists of an image encoding followed by a decoding pipeline and yields a probability map of the target class. Specifically, variants of this architecture have been used in local and regional studies for the detection of SGLs in multispectral (Lutz et al., 2023; Niu et al., 2023) and SAR data (Chen, 2021; Dirscherl et al., 2021a). For this study, we use the U-net architecture proposed by Dirscherl et al. (2021a). It uses residual skip connections and Atrous Spatial Pyramid Pooling (ASPP) and has significantly improved SGL mapping accuracies based on Sentinel-1 SAR data before (Dirscherl et al., 2021a). We decided to evaluate this specific architecture because (i) we were already familiar with its code and application; (ii) we expected it to adapt well to the multispectral data during the training process based on the aforementioned studies; and (iii) our initial tests already showed good results. The only modification made to the model architecture is that the input imagery contains four layers instead of one. These spectral features comprise the ones depicted in Table 2.

The evaluated U-net architecture consists of four downsampling (encoding) and four corresponding upsampling blocks (decoding), with the deepest blocks being connected by an ASPP module (Chen et al., 2018) (Figure 4). In each downsampling block, the input layer is convoluted two times using a 3×3 kernel, while the original image size is kept by padding. Additionally, the input layer is convoluted with a 1×1 kernel and added to the previously generated convoluted layer to establish a residual connection to keep the original input information in later steps of the model. A transformation using the Leaky nonlinear Rectified Linear Unit (LeakyReLU) activation function is introduced to enable nonlinear modeling. Then, the image size is halved by applying a 2 × 2 maximum pooling operation, and 30% of neurons are dropped out to avoid overfitting. The resulting image comprises the input to the next downsampling block.

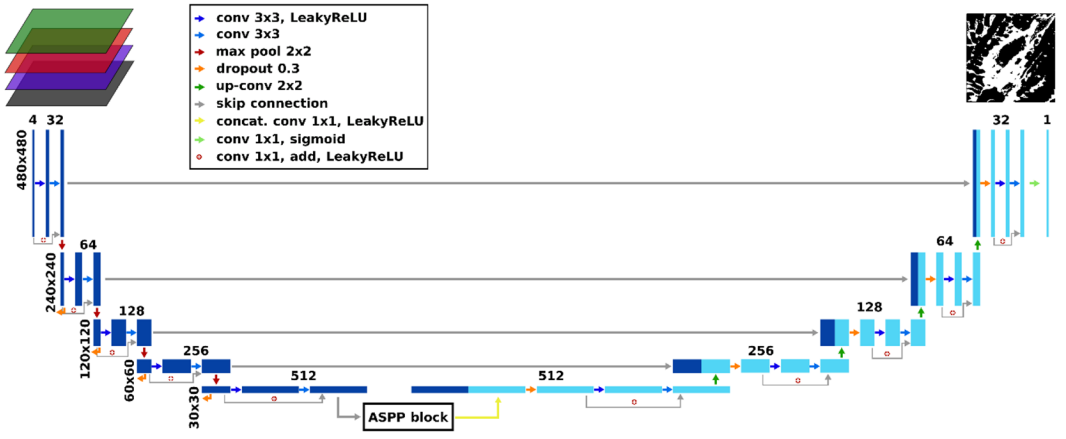


Figure 4. *Architecture of the U-net model used for Sentinel-1 and Sentinel-2-based supraglacial lake detection. Modified after Dirscherl et al. (2021a).*

This procedure is repeated until a depth of 512 layers is reached. The resulting image features are then put into an ASPP block, in which they are convoluted with dilation rates of 2, 4, 8, and 12, and the outputs are concatenated. This enhances the model's ability to learn and model lakes of different shapes and sizes. To restore the original image size, features are then given into four upsampling blocks that mirror the convolution strategy of the downsampling blocks. Additionally, via a skip connection, each upsampled feature is concatenated with the output of the corresponding downsampling block. This enables the reconstruction of the high-resolution image information from the lower-resolution features while maintaining the abstraction level of the convoluted features. In the final upsampling block, the output image size is reconstructed with an image depth of 1 and passed into a sigmoid activation function that returns a probability map of SGL occurrence. For model training, the Adamax optimizer with an initial learning rate of 0.0001 was selected, which was reduced by a factor of 10 if the learning rate did not improve over three epochs. Learning was finished if a learning rate of 0.0000001 was reached. "Binary cross-entropy" was used as the loss function.

The model was trained for 50 epochs, but convergence in the internal model validation was reached much earlier (Figure 5). The best model run was achieved at epoch 16 at a loss rate of 0.06 and an accuracy of 0.97. Since there was some fluctuation at later epochs, the model from epoch 25 was selected for validation and inference, as the model loss and accuracy fully stabilized at 0.06 and 0.97, respectively.

2.2.3. Post-processing

The output of each U-net run is a 480 × 480 pixel probability map of SGL occurrence. All tiles of a scene are reassembled to reconstruct the original images *x*- and *y*-dimensions. For overlapping areas introduced during the tiling process, the mean probability is calculated. Next, topographic masking of areas with slopes steeper than 5% and altitudes higher than 1,500 m was applied using the TanDEM-X PolarDEM. Furthermore, areas seaward of the coastline as derived by the IceLines product (Baumhoer et al., 2023) were excluded. Next, thresholds of 90% and 99% probability were applied to generate binary lake/no-lake maps. Lake areas smaller than three pixels (~300m^2) were sieved out to reduce image noise. It is common practice in other SGL mapping studies (Stokes et al., 2019; Dirscherl et al., 2020, 2021a; Moussavi et al., 2020b; Corr et al., 2022) to filter out small lakes with sizes equal and lower than 2–45 pixels, depending on the spatial resolution of the input data. In a systematic assessment, Stokes et al. (2019) report the lake area omitted by filtering to be insubstantial compared to the overall detected lake area.

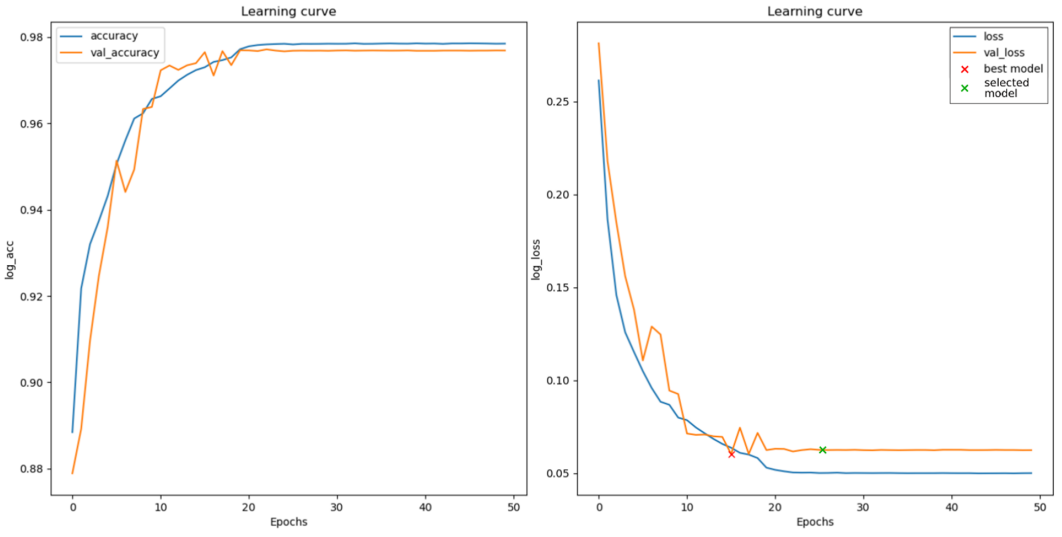


Figure 5. *Internal accuracy and loss rate of the trained U-net model.*

2.2.4. Accuracy assessment

The SGL maps generated for the testing sites (see Section 2.1) were then compared to manually labeled data that was not used during model training (see Section 2.2.1). In the original SGL mapping approach by Dirscherl et al. (2020), a point-based validation approach was used. We decided to perform the validation on fully segmented labels instead, since the development of the consistent labelling rule set required a rework of the testing data labels and, more importantly, a comparison to fully segmented data better reflects the ability of the model to map spatially consolidated lake areas. In our approach, the post-processed SGL maps are compared pixel-wise to the testing labels, and the producer's accuracy, user's accuracy, and F1 score are calculated. The latter was selected over the overall accuracy, as the ratio between SGLs and other surface features is highly imbalanced. For each testing scene, we compared the 90% and 99% thresholded outputs of the U-net and the output by the original RF SGL mapping approach as described by Dirscherl et al. (2020), including the improvements implemented in a follow-up study. These include a greater representation of shadow in the training data, fine-tuning of variable thresholds during post-classification, and slush elimination using probability maps (Dirscherl et al., 2021b). Additionally, we calculate the accuracy metrics on all scenes combined as a total metric.

2.3. Comparison with existing SGL datasets

While the accuracy assessment allows us to assess the performance of the proposed DL image segmentation algorithm relative to the RF classification approach by Dirscherl et al. (2020), we also compare our mapping results with those of other SGL datasets. It is noteworthy that this analysis only compares the output products and, since the comparison datasets are based on different input data sometimes acquired at different points in time (Sentinel-2 and/or Landsat), is not suited to compare the performance of the classification/segmentation algorithms themselves. This is because there are no independent ground-truth labels available for the intercomparison regions. The comparison evaluates the relative consistency and agreement of the different SGL datasets with the results generated by our U-net. This approach enables an assessment of the robustness of the mapped spatial SGL patterns and can help identify regions of systematic uncertainty and features that are sensitive to spatial resolution and acquisition timing. For this, we selected the most established large-scale mapping algorithms with open-access datasets for EAIS and WAIS. We selected two highly ponded test regions during January 2017: Bach ice shelf for WAIS and Amery ice shelf for EAIS. We acknowledge that two ice shelves may not be a statistically representative

example for the entirety of Antarctica's ice shelves and glaciers. However, these regions cover a wide variety of surface features, such as lake networks and isolated lakes, lakes of various morphologies (deep, shallow, oval, elongated, ice-covered, and ice-free), and blue ice and slush areas, and are located in different geographical sectors of the continent. Thereby, they can provide an informative evaluation of model agreement under realistic Antarctic melt conditions given the limited spatial and temporal overlap of the comparison datasets and the requirement that enough lake area be present in the data.

Over the Bach ice shelf, we compared the mapping results of our newly developed approach with the datasets published by Dirscherl et al. (2020), Corr et al. (2022), and Dell et al. (2024). Unfortunately, the dataset by Moussavi et al. (2020a) lacked data for this region in January. Over the Amery ice shelf, we compared the U-net approach with the datasets published by Stokes et al. (2019), Dirscherl et al. (2020), Moussavi et al. (2020a), and Dell et al. (2024). Please note that Stokes et al. (2019) and Corr et al. (2022) only cover the WAIS/API and EAIS, respectively. For the purposes of comparison, the datasets were transformed into vector data where necessary. Since the dataset by Moussavi et al. (2020a) was provided on a per-scene basis, we generated monthly maximum extent maps from all available SGL maps over the regions of interest. In all cases, only classes representing surface water were considered. In the Dell et al. (2024) dataset, we considered only the ponded water class and omitted areas classified as slush from the comparison, because our approach is focused on ponded surface water. In the Corr et al. (2022) dataset, we considered both the "lakes" and "channels" classes, as wider channels are also recognized as SGLs in our approach. The datasets were clipped to the shelf area seaward of the grounding line (Gerrish et al., 2024). This is because the Dell et al. (2024) dataset only considers ungrounded lakes. Regarding the Amery ice shelf, only the area south of 71.45°S was considered, as no data are available in the Stokes et al. (2019) dataset further north. The area is still deemed large enough to enable a meaningful analysis. To conduct a relative comparison, we calculated the absolute lake areas of each dataset in the respective region and compared them to the results of our U-net approach. Next, we quantified the spatial overlap between the U-net and each comparison dataset. Additionally, we calculated the difference between the areas detected by our approach and those detected by the comparison datasets to determine which areas were detected by the U-net but not the comparison datasets, and vice versa. Finally, we calculated the intersection over union (IoU) as a convenient metric of agreement between the U-net and the comparison datasets. It is calculated as

$$IoU = \frac{a_i}{a_u}, \tag{1}$$

where a_i denotes the area of intersection between the U-net mapping results and the comparison dataset, and a_u denotes the total area covered by the results of the U-net and the comparison dataset combined. The result is a value between 0 and 1, where 0 means no agreement and 1 means full agreement between the compared datasets.

3. Results

3.1. Comparison of U-net and random forests SGL mapping

Across all Sentinel-2 validation scenes, the U-net classifications thresholded at 90% probability achieved the highest F1 score, followed by the 99% U-net and the RF (Table 3). Increasing the probability threshold results in fewer misclassifications and thus higher precision but also increases the rate of false negatives, leading to the omission of lake areas. In comparison to the U-net, RF has lower rates of misclassifications, but the low producer's accuracy indicates that a substantial amount of lake area is not detected in an approach relying only on spectral image information.

A more detailed accuracy analysis based on validation regions allows us to assess the influence of different surface water types and surface features on the classification accuracy of each model (Figure 6). In 12 of the 16 validation sites, the 90% U-net produces the most accurate classification, while in three cases the RF is superior to the U-net. In only one case, the 99% U-net outperformed the 90% threshold. A full table containing all accuracy metrics for all validation sites is provided in the Supplementary Material (Supplementary Table S3).

Table 3. *Overall accuracy assessment of the Sentinel-2 U-net-based supraglacial lake detection approach and comparison to the existing RF-based approach*

Classification approach	Producer's accuracy (recall)	User's accuracy (precision)	F1 score
U-net 90%	0.915	0.912	0.913
U-net 99%	0.794	0.982	0.878
Random forests	0.750	0.945	0.837

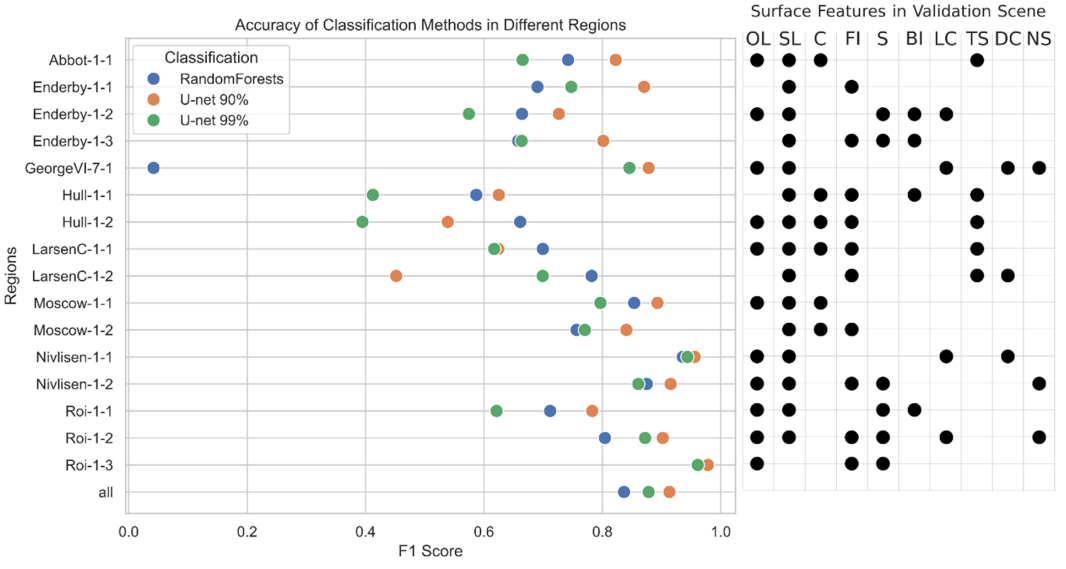


Figure 6. *Accuracy assessment of the Sentinel-2 U-net-based supraglacial lake detection approach and comparison to the existing RF-based approach for each validation scene. The table on the right shows which surface features are present in the scene: OL = oval lakes; SL = shallow lakes; C = crevasses; FI = floating ice (i.e., ice cover); S = slush; BI = blue ice; LC = lakes visible under clouds; TS = topographical shadows; DC = dark clouds; NS = narrow streams. See Figure 3 and Supplementary Table S2 for visual examples of these surface features.*

Figure 7 shows examples from the validation scenes. On the GeorgeVI-7-1 validation site, RF achieved a rather low classification accuracy (F1 score = 0.04) while the U-net achieved 0.88 and 0.95 at the 90% and 99% probability thresholds, respectively. In contrast to RF, the U-net is capable of detecting lakes reliably under a thin layer of dark clouds. Only very shallow lakes and some streams are omitted by the classifier. On Hull-1-2, all classifiers achieve low accuracy between 0.4 and 0.66. In this scene, most of the lakes are very shallow lakes and some are covered by ice. Particularly, the lakes in the south fall below the 90% probability threshold of the U-net classifier. In contrast, the 90% U-net fully classifies the lake areas covered by ice on the Enderby 1–1 validation site, while the detected lake areas by the RF remain patchy and incomplete. This highlights the spatial awareness of the U-net approach that detects lake shapes even if the spectral signature might be less clear due to the ice coverage. In general, the lower probability threshold of the U-net yields better results since, at 99%, the SGL area tends to be underestimated, resulting in a lower recall (Table 3). LarsenC-1-2 is the only example where the 99% U-net achieves a better F1 score (0.7) than the 90% version (0.45). The reason for this is dark clouds over lake-free areas, which the U-net tends to classify as lakes. Applying a high probability threshold reduces

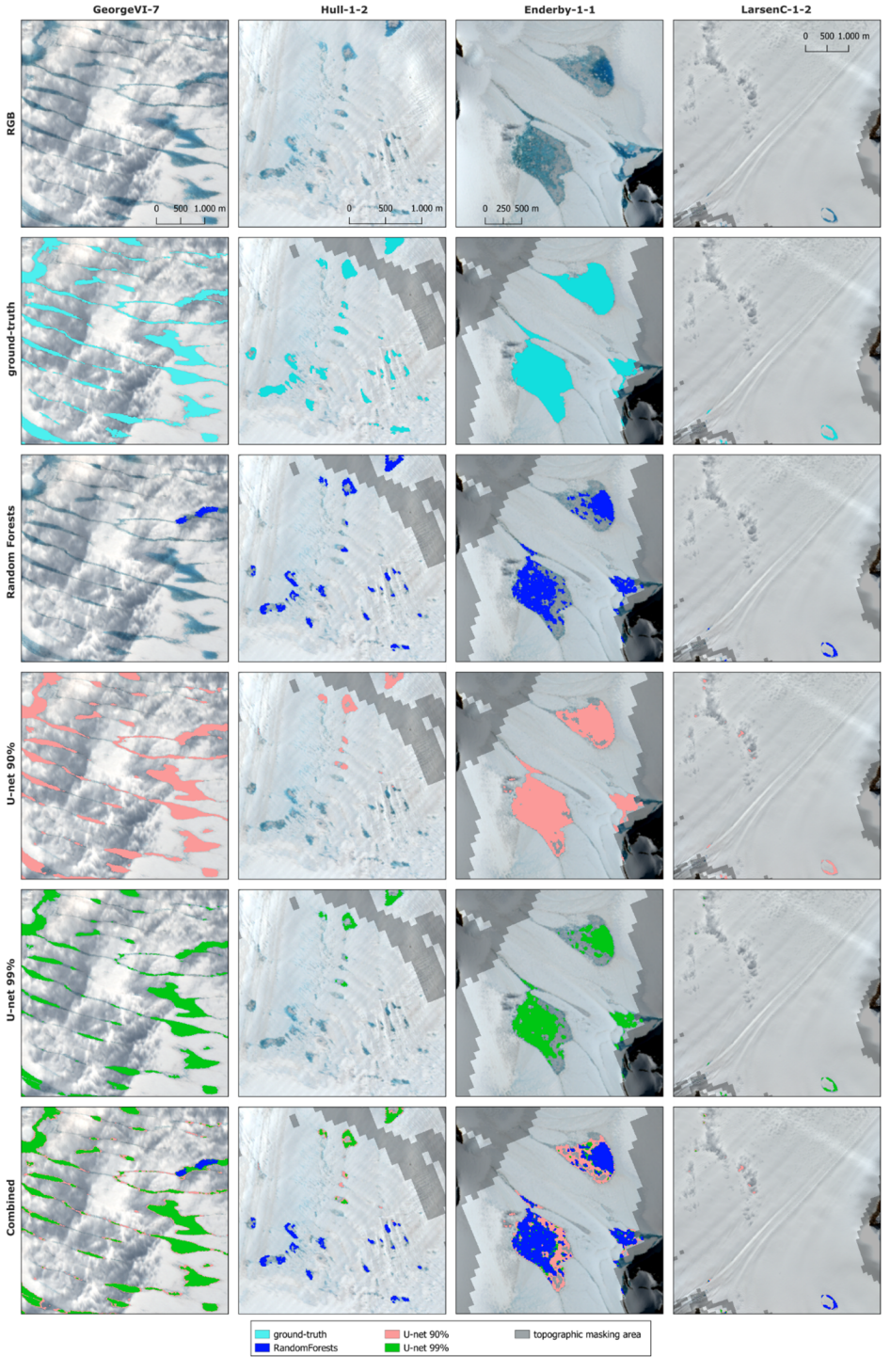


Figure 7. *Examples of SGL predictions above different surface features and comparison of the Sentinel-2 U-net and RF-based classification approaches. Columns show subsets of four different validation sites. Rows show the original RGB (true color) Sentinel-2 image, the testing SGL labels ("ground-truth"), the results of the evaluated classifiers, and a combination of all classification results.*

the number of misclassifications, thus increasing both precision and F1 score. A similar issue was not observed in the RF classification, resulting in a comparably higher accuracy (F1 score = 0.78).

3.2. Comparison of mapping results with other SGL data products

As elaborated in Section 2.3, the relative comparison of the mapping results of the U-net with those of the comparison datasets can only be an indicator for the general robustness of the mapped spatial SGL patterns. Due to the lack of independent ground-truth data and the fact that the datasets are based on different input data, there is no way to exactly quantify which product yields the most complete results or whether an algorithm outperforms another.

On the Amery ice shelf, the surface water product by Dell et al. (2024) and the U-net map the highest absolute lake area (Table 4), despite being based on different datasets (Landsat 8 and Sentinel-2, respectively). The dataset by Moussavi et al. (2020a), which is based on Landsat 8, follows closely behind. Despite combining both optical sensors, the dataset by Stokes et al. (2019) maps only two-thirds of the lake area mapped by the U-net or Dell et al. (2024). Even though these two approaches map a similar amount of lake area, the IoU between them is quite low at 0.56 (Figure 8a). This means that there is comparatively little overlap between the mapped lakes, while at the same time each dataset maps a considerable amount of lake area that the other does not (Table 4). While Moussavi et al. (2020a) and Dirscherl et al. (2020) map a lower amount of overall lake area compared to the U-net's results, their agreement with it is much higher (0.69 and 0.66, respectively). The highest agreement with other datasets is achieved by Moussavi et al. (2020a) in comparison with Stokes et al. (2019) (0.7), the U-net (0.69), and Dirscherl et al. (2020) (0.68). Meanwhile, Dirscherl et al. (2020) achieve the highest mean IoU with all other datasets with 0.65. The average agreement of our results with the comparison datasets is only slightly lower (0.63). Meanwhile, Dell et al. (2024) achieve the lowest mean IoU (0.55) and have particularly low agreement with Stokes et al. (2019) (0.48) and Dirscherl et al. (2020) (0.52). It is noteworthy, however, that the dataset by Dell et al. (2024) is the only one to consider slush as a separate class, while there is no sharp discrimination between slush and ponded water in the other datasets. The "slush" class was not considered in this comparison.

A visual comparison of the SGL products on a subset of the Amery ice shelf (Figure 8b–g) shows that lake areas tend to be more consolidated in the U-net's mapping results. At the same time, smaller lakes appear to be mapped with larger areas. These two observations can explain the larger overall lake area mapped by the U-net, particularly in direct comparison to Dirscherl et al. (2020), which is also based solely on Sentinel-2 data.

On the Bach ice shelf, the differences between the compared datasets are relatively bigger (Table 5) but note that the overall smaller area considered causes smaller deviations to have a bigger impact on the statistics. The U-net maps almost double the lake area compared to the dataset by Corr et al. (2022), while Dirscherl et al. (2020) and Dell et al. (2024) only detect 37% and 28%, respectively, of the U-net's area. The IoU between the datasets is generally lower than on the Amery ice shelf (Figure 9a). The highest agreement is between Dirscherl et al. (2020) and Corr et al. (2022) (0.62), where both datasets are based on

Table 4. *Comparison of lake areas (km^2) detected by the newly developed approach (U-net) and several comparison datasets on the southern part of the Amery ice shelf*

Dataset	Absolute lake area	Lake area difference (U-net - dataset)	Dataset overlap with U-net	Area mapped by U-net, but not by dataset	Area mapped by dataset, but not by U-net
U-net	1269.16	N/A	N/A	N/A	N/A
Stokes et al. (2019)	843.33	−425.82	788.68	480.47	54.65
Dell et al. (2024)	1273.80	4.65	916.46	352.70	357.35
Dirscherl et al. (2020)	981.54	−287.62	917.30	351.85	64.24
Moussavi et al. (2020a)	1159.75	−109.40	966.43	302.73	193.33

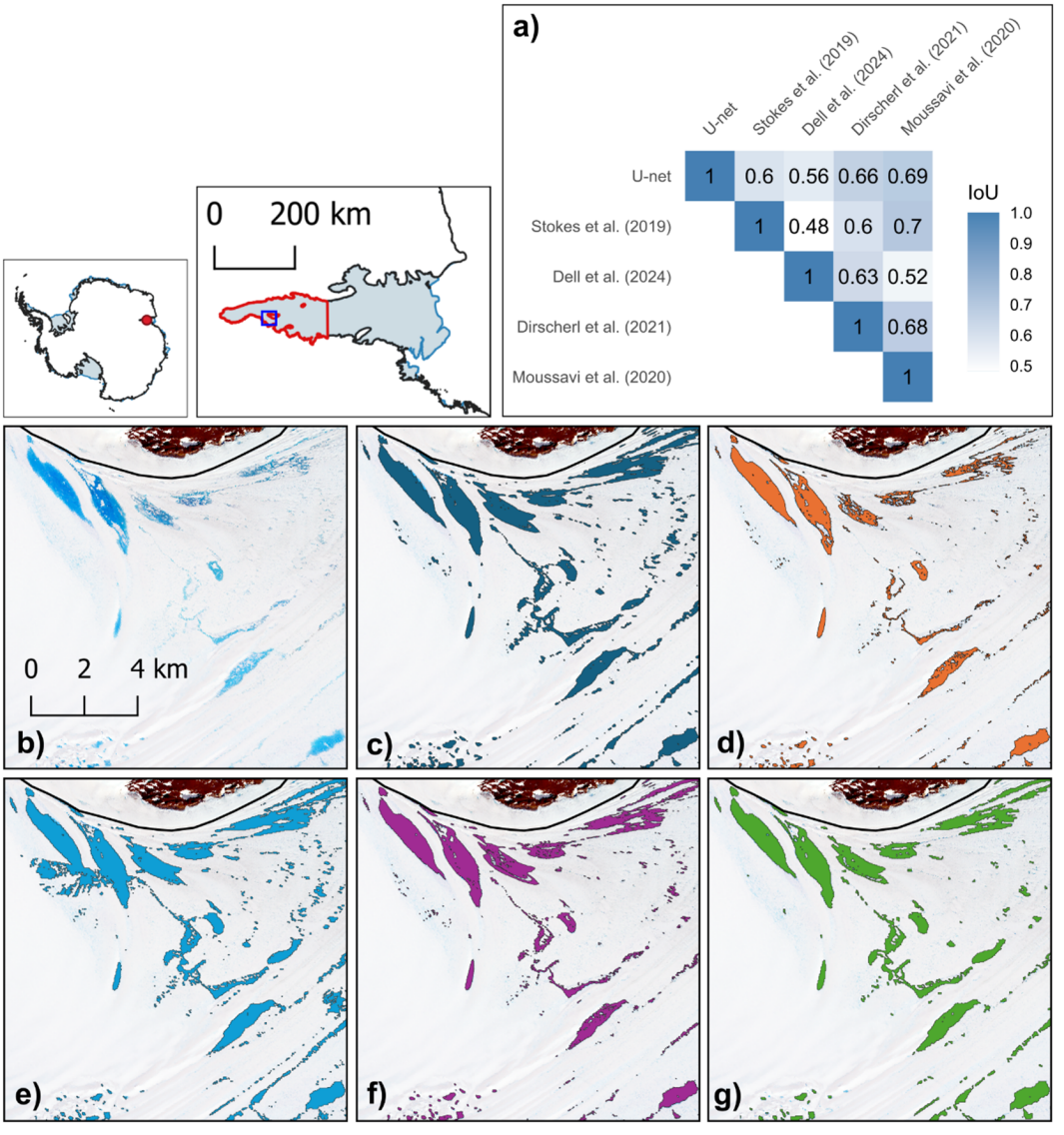


Figure 8. *Comparison of lake areas as detected by the U-net and several comparison datasets on the Amery ice shelf. Maps show the location (red dot) of the ice shelf, the area of analysis (red shape), and the scene displayed in (b)–(g) (blue box). (a) Intersection over union (IoU) between the compared datasets. (b)–(g) Lake mapping for a subset of the Amery ice shelf: (b) RGB composite of Sentinel-2 data for January 2017, (c) U-net, (d) Stokes et al. (2019), (e) Dell et al. (2024), (f) Dirscherl et al. (2020), and (g) Moussavi et al. (2020a). The black line indicates the grounding line. Maps are displayed in Polar Stereographic projection.*

Sentinel-2 (and, in the case of the latter, also Landsat 8), while there is considerable disagreement between the U-net and Dell et al. (2024) (0.21), which use different sensors as input data. The highest average IoU is achieved by Corr et al. (2022) (0.49) and Dirscherl et al. (2020) (0.43), followed by the U-net (0.36) and Dell et al. (2024) (0.28). The visual inspection of a sample area (Figure 9b–f) shows generally good agreement between the U-net and the datasets by Corr et al. (2022) and Dirscherl et al. (2020), which all use Sentinel-2 data as input. The Dell et al. (2024) dataset, interestingly, omits a shallow but wide lake and

Table 5. *Comparison of lake areas (km^2) detected by the newly developed approach (U-net) and several comparison datasets on the Bach ice shelf*

Dataset	Absolute lake area	Lake area difference (U-net - dataset)	Dataset overlap with U-net	Area mapped by U-net, but not by dataset	Area mapped by dataset, but not by U-net
U-net	21.98	N/A	N/A	N/A	N/A
Corr et al. (2022)	12.62	−9.36	11.45	10.30	0.94
Dell et al. (2024)	6.20	−15.79	4.82	17.16	1.38
Dirscherl et al. (2020)	8.20	−13.78	8.16	13.82	0.04

stream network in the center of the detail image (Figure 9e), even though surrounding lakes are at least partly detected and the "slush" class provided with this dataset does not cover this area. However, comparing the differently timed acquisitions by Landsat 8 and Sentinel-2 in this area reveals that this lake system is much less apparent in the Landsat 8 data, upon which the dataset by Dell et al. (2024) is based. In general, the dataset used (either Landsat 8 or Sentinel-2) and thus the acquisition timing seem to play an important role in the mapping results on the Bach ice shelf in January 2017. Due to frequent cloud cover, only a few scenes are available over the ice shelf within the month, and due to the transient nature of surface water, there may occur considerable differences in lake maps even in maximum extent composites.

4. Discussion

4.1. Comparison of U-net and random forests mapping

A major goal of this study is to analyze how a U-net-based classifier compares to an RF-based classification scheme to detect SGLs in Antarctica. The biggest advantage is the increased capability of the U-net to map lakes as consolidated areas in comparison to the sometimes patchy pixel-wise RF product. This results in an increased producer's accuracy and F1 score of the DL approach, particularly at a probability threshold of 90%. This capability can be attributed to the spatial awareness of the U-net and its ability to consider spatial context. Especially when ice or thin clouds cover parts of a lake, the U-net is able to restore the original lake shape to a certain degree. Meanwhile, the RF can only consider the spectral information at a respective pixel, which may diverge significantly from that of uncovered lakes. As a result, there was almost no confusion of lake area with potentially spectrally similar features such as blue ice and dirty ice (Supplementary Figure S3). Additionally, the U-net requires a smaller amount of input features for inference, which makes it suitable for large-scale mapping efforts as it reduces the computing resources required in comparison to the original RF approach.

On the other hand, for some surface features, the RF still performs better than the U-net. We observed some misclassifications of very dark areas covered in topographic shadows as lake areas, most of which were eliminated during post-processing using a topographic mask layer (Supplementary Figure S1). Next, in some cases such as the LarsenC-1-2 validation site, the U-net misclassified dark clouds as lake areas (Figure 7). In this particular case, a comparatively small lake area coincided with extensive cloud coverage. The area of misclassified clouds amounted to 0.09 km^2, whereas the labeled lake area covered only 0.06 km^2 within a total test area of 111.1 km^2. Here, no consistent subsequent masking was possible without serious additional effort, since the cloud information provided in the Sentinel-2 quality bands is unreliable over ice and snow. The application of a higher probability threshold at 99% reduced the number of cloud-related misclassifications but came at the cost of a reduced lake detection rate. We further experimented with adding different bands and indices, such as the New Water Index (NWI) or the visible blue spectral band, to the feature stack because in previous studies these had demonstrated a high

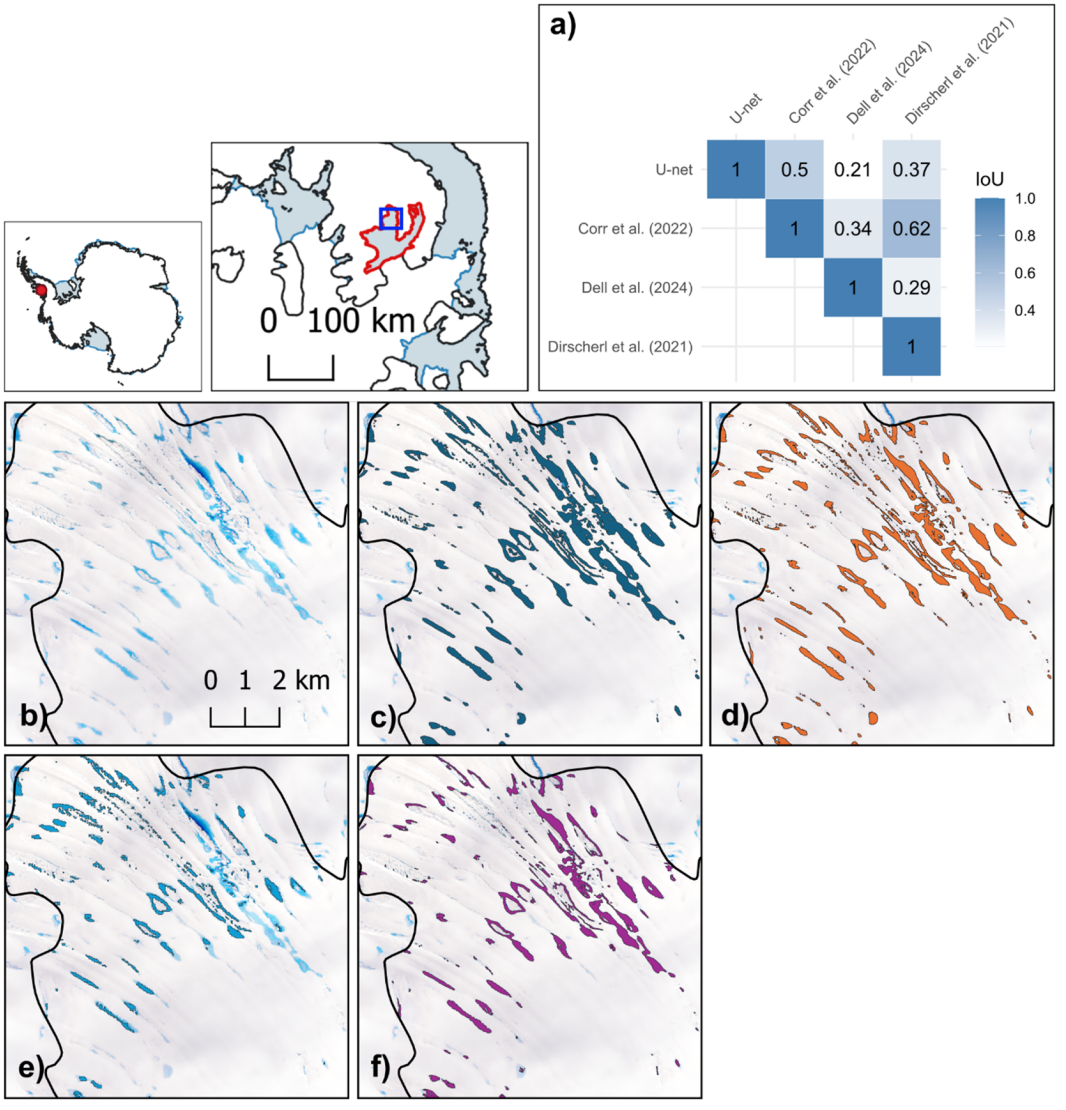


Figure 9. *Comparison of lake areas as detected by the U-net and several comparison datasets on the Bach ice shelf. Maps show the location (red dot) of the ice shelf, the area of analysis (red shape), and the scene displayed in (b)–(f) (blue box). (a) Intersection over union (IoU) between the compared datasets. (b)–(f) Lake mapping for a subset of the Bach ice shelf: (b) RGB composite of Sentinel-2 data for January 2017, (c) U-net, (d) Corr et al. (2022), (e) Dell et al. (2024), (f) Dirscherl et al. (2020). The black line indicates the grounding line. Maps are displayed in Polar Stereographic projection.*

predictive power (Dirscherl et al., 2020). However, none of these experimental setups improved the overestimation of clouds considerably and even increased the number of misclassifications in other areas such as topographic shadows. Eventually, the combination presented in Table 2 yielded the most accurate results. Finally, the RF achieved better results in scenes with very shallow lakes and slush. Especially in the transition areas of consolidated lakes to slush, spectral information becomes more important than shape when surface water is to be mapped. However, in the overall assessment, the U-net at a 90% probability threshold achieves superior classification accuracy compared to the RF approach in 13 of the 16 test sites.

4.2. Comparison with existing SGL and surface water products

Any discussion about observed differences between the compared SGL datasets must acknowledge the differences of the source satellite imagery to each product. Here, the acquisition timing of the image in which lake areas are detected is crucial. SGLs can be highly dynamic surface features, and a lake that may have been detected in Sentinel-2 imagery might be already gone in a later Landsat 8 acquisition. This problem is exacerbated by frequent cloud cover, effectively reducing the mapping frequency, as was the case on the Bach ice shelf in January 2017. Especially comparing datasets relying only on a single sensor may thus yield considerable differences in lake area. In the future, it will therefore be sensible to combine different data sources to achieve the most complete lake inventory and consider the use of SAR data to fill gaps in optical data that occur due to cloud cover. Until then, the interpretation of possible causes of the differences between datasets must be done with caution.

However, the relative comparison can give some indication of the robustness of the large-scale SGL patterns detected by the different datasets. The agreement of the mapping results is greater on the Amery ice shelf than on Bach. The average IoU of the U-net's results in comparison with the other datasets indicates robust mapping results in settings with deep lakes of various morphologies. Meanwhile, on the Bach ice shelf, which is characterized by a larger proportion of shallow lakes, slushy areas, and some thin clouds, there is greater disagreement. Particularly, the transition zones between ponded water and saturated firn are mapped inconsistently between datasets. Our classification approach focuses more on consolidated surface water areas, that is, lakes and streams, which was the original focus of the mapping methods developed by Dirscherl et al. (2020, 2021a). The mapping of slush is a significant challenge for a U-net, as there is almost no trainable characteristic shape or spatial context, and the generation of manually labeled training scenes for the slush class is almost impossible. However, Dell et al. (2024) suggest that slush makes up a significant proportion of the total amount of surface water, including streams and ponded water, on Antarctic ice shelves. Like lake areas, slush reduces albedo, further enhancing surface melting processes that may be poorly represented in climate models used to model ice sheet melting and project sea level rise. Therefore, it may be worthwhile to combine RF and DL-based classification approaches to comprehensively map surface water in the future. In such a framework, the ability of RF to use spectral diversity to map slush could complement the unprecedented accuracy of the U-net to map consolidated lake and stream areas. It should be noted, however, that this would come at the cost of an even more computationally demanding preprocessing workflow.

In general, the U-net mapped a larger lake area than the comparison datasets. This is also the case for products which use Sentinel-2 as input data (Stokes et al., 2019; Dirscherl et al., 2020; Corr et al., 2022), too. Visual inspection of the mapping results shows a tendency toward more consolidated lake areas in the DL product. Since the comparison datasets are pixel-based approaches, lake areas, especially when covered partly with ice, appear patchy, likely leading to underestimation of lake area. Another reason for the more extensive coverage of the U-net might be due to the training data, in which we deliberately included the shallow fringes of the lakes as lake area, while the pixel-based classification approaches might produce ambivalent results in these transition areas. On the Bach ice shelf, the presence of clouds may have caused an overestimation of lake area in the U-net results, where dark clouds have been misclassified as lake area close to the ice shelf terminus (see also Section 3.1 on this issue). Without this, the agreement with the other Sentinel-2-based datasets would be considerably higher, as indicated by the detail maps.

The U-net and the dataset by Dell et al. (2024) achieve a low IoU despite a similar amount of mapped lake area. This is likely caused by the different input data and may further be exacerbated by other causes. First, Dell et al. (2024) consider a separate slush class, which is likely a cause for the low IoU with the U-net on Bach and with the Landsat 8-based dataset by Moussavi et al. (2020a) on Amery. This classification likely causes many areas in the transition zone not to be considered in the dataset comparison, as they are not classified as "lake." Furthermore, on the Amery ice shelf, there is an apparent artifact close to the two large lakes in the upper left part of the detail scene (Figure 8e). A similar lake area is not visible either in the other Sentinel-2 or the Landsat 8-based datasets. Checking all available Landsat 8 scenes over the region in January 2017 reveals that on the 25th of January the location was partly

covered in clouds (scene on Path 129, Row 111), which may have caused artefacts in the classification approach. Moreover, we identified several areas with a similar shape further north on the Amery ice shelf, which may have contributed to the overall lake area in this region.

Even though all compared datasets achieved good mapping accuracies as reported in the respective publications, the observed differences between the datasets in the relative comparison demonstrate that a direct comparison of the reported metrics is difficult. A major reason for this is the fact that validation approaches are not standardized. For example, Stokes et al. (2019) do not provide any accuracy metrics compatible with our evaluation approach based on a confusion matrix but only compare the total size of labelled and classified lake areas. The approach presented by Corr et al. (2022) achieved an overall mean accuracy of 0.985, with a mean sensitivity and specificity of 0.815 and 0.994, respectively, which is close to the metrics of the U-net at 90% probability with 0.997 (overall accuracy), 0.915 (sensitivity), and 0.998 (specificity; details in Supplementary Table S3). However, both datasets showed considerable differences when compared directly on the Bach ice shelf. Meanwhile, Dell et al. (2024) use a point-based sampling approach, where testing labels were generated by four experts independently; overall mean accuracies of 84% and 75% were achieved for the ponded water and slush classes, respectively. Again, the comparability to our approach is limited, mainly due to the different input data, the division of surface water into two classes while we only consider ponded water without a sharp delineation of slush, and the use of a point-sampling approach compared to our validation on fully segmented lake labels.

The relative comparison of the different SGL products in this study demonstrated that there is partly good agreement between the products but in some cases also considerable differences. A final judgment on which dataset is "best" in terms of the most complete coverage of lake area, however, is not possible with this approach nor in the framework of this study. Every approach has limitations and strengths, and in a relative comparison as performed in this study, it is not in every case possible to determine whether differences between the datasets are based on data availability, the detection method, cloud coverage, sensor or region specifics, or other factors. Therefore, we recommend a dedicated intercomparison exercise of available SGL products in follow-up work that should include independently collected in-situ data as a baseline, which must be available for all compared datasets.

4.3. Potential for improvements for automated SGL detection

Despite the overall high SGL detection ability of the proposed U-net workflow, some room for improvement remains. An issue that is currently relevant for all DL-based classification efforts is the time-consuming generation of consistent training data. For this study, the application of a consistent labeling rule set on all training and validation scenes significantly improved the model accuracy. However, it posed a considerable challenge to generate the precise training labels required for an image segmentation approach such as the U-net, particularly in the transition areas between consolidated lakes and slush. We experimented with semi-automatic labeling approaches by enhancing and thresholding image histograms but found that a lower amount of consistent and high precision manual training label data yielded significantly better results during inference. It is likely that an increased variety of training data, particularly of problematic surface features, will further increase the U-net's accuracy. This would also present an opportunity to test the results on more ice shelves and glaciers that represent a wider variety of surface features, geographic regions, and temporal variation. In the future, foundation models pretrained on Earth observation imagery might reduce the required effort to generate training data for SGL detection tasks, but for now, it is much less effort to generate training labels for any pixel-based classifier. This is due to the fact that such training data only requires enough samples of spectral information but no spatial context. While a DL-based image segmentation algorithm such as the U-net requires fully segmented training images, conventional training data can have any shape and is thus much easier to generate. However, in an operational processing pipeline with a trained model that would be used to classify large volumes of input data, the DL-based approach would be more efficient than pixel-based approaches such as RF. In such a scenario, the data preprocessing, that is, the preparation of the acquired satellite imagery in such a way that it can be presented to the model for inference, is the computational

bottleneck. In our case, the preprocessing of a satellite scene took around 80–90% of the total processing time, while the actual inference was completed in 10–20%. With DL requiring far fewer input bands for inference than an RF approach to yield comparable results, we expect our new approach to be less computationally expensive in large-scale inference scenarios such as processing the entire Antarctic margin.

Finally, further improvement regarding the limitations of optical satellite data can be made by including optical Landsat and/or SAR data in the product. A higher acquisition frequency allows checking the temporal continuity of detected lake areas. Thus, it can be validated whether a time gap occurred due to cloud obstruction or if a cloud was misclassified as a lake by the U-net. A lake that vanished from the time series between two observations is likely covered by a cloud in the central observation. Meanwhile, a supposed lake area that occurs only at a single observation within the time series will likely be a cloud misclassification.

5. Conclusion and outlook

In this study, we presented an approach to map supraglacial lakes around the Antarctic margin using multispectral Sentinel-2 satellite imagery in an automated DL classification (U-net) approach. The mapping results were compared to its RF-based predecessor by validation on manually generated high-precision, fully segmented SGL maps. The U-net at a 90% probability threshold outperformed the RF classifier in 13 of the 16 validation scenes, achieving an F1 score of 0.913 (RF: 0.837), a producer's accuracy of 0.915 (0.750), and a user's accuracy of 0.912 (0.945). The U-net was able to map lake areas more completely and consistently in comparison to the pixel-based RF approach, and was thus able to detect lake areas under thin clouds or ice cover. Remaining challenges are the transition areas between lakes and slush, where spectral information is more important than lake shape or spatial context, as well as the misclassification of some clouds as lakes. However, the fact that the method yielded satisfactory classification results in regions on the West and East Antarctic ice shield as well as the Antarctic Peninsula demonstrates its transferability. In a relative comparison with mapping results from five other SGL products on two of Antarctica's ice shelves, the U-net mapped the largest lake area. However, the agreement with the comparison datasets, as measured by the IoU, was mixed, as was the IoU among the comparison datasets. This is likely due to different data sources of the datasets (either Landsat 8, Sentinel-2, or both) and cloud cover causing different timings on which the highly dynamic SGLs are mapped. Here, additional data sources such as SAR can potentially yield more complete lake maps at a higher temporal frequency. Additionally, the mapping method contributes to disparities between the datasets, with the U-net, again, being able to produce more consolidated lake areas than the pixel-based classification approaches. However, due to the relative nature of this comparison, for which independent ground-truth data covering all datasets were not available, it is not possible to quantify to what degree over- or underestimation of lake areas may have taken place.

Even though the generation of training data is more challenging for the U-net, the method presented enables the creation of pan-Antarctic datasets on SGL dynamics, as inference requires fewer input features, thus reducing computational effort when applied on large scales. Future assessments should explore the method's integration with other satellite sensors to enhance dataset completeness. Dense temporal information on SGL dynamics is vital for hydrological, firn, and surface mass balance models, improving simulations of surface melt. SGL datasets from observational data are essential for validating lateral water transport parameterizations and understanding processes that lead to ice thinning, which could contribute to sea level rise from global warming.

Abbreviations

API	Antarctic Peninsula
ASPP	Atrous spatial pyramid pooling
$AWEI_{nsh}$	automated water extraction index
BOA	bottom-of-atmosphere

DEM	digital elevation model
DN	digital numbers
EAIS	East Antarctic ice sheet
IoU	intersection over union
LeakyReLU	leaky nonlinear rectified linear unit
ML	machine learning
MSI	multispectral instrument
NDSI	normalized difference snow index
NDWI	normalized difference water index
NIR	near-infrared
NWI	new water index
RF	random forests
RGB	red, green, blue
SAR	synthetic aperture radar
SGL	supraglacial lake
SWI	soil/water index
SWIR	shortwave infrared
WAIS	West Antarctic ice sheet

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2026.10065.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10065.

Acknowledgments. The authors would like to thank Dr. Mariel Dirscherl for the careful preparatory work on the methodology.

AI statement. For the preparation of the manuscript, the authors used the tool DeepL Write (https://www.deepl.com/en/write) to improve the writing style and correct grammatical errors in some parts of pre-written text. The tool was not used to generate new text, content, citations, or to summarize other existing works. After applying the tool, the authors carefully checked the results and took full responsibility for the final version of the manuscript.

Author contribution. Conceptualization: J.K., C.B., and A.D. Data curation: J.K. and C.B. Formal analysis: J.K. Funding acquisition: A.D. Investigation: J.K. Methodology: J.K. and C.B. Project administration: A.D. Software: J.K. and C.B. Supervision: B.W., S.L., and A.D. Validation: J.K. Visualization: J.K. Writing—original draft: J.K. Writing—review and editing: C.B., B.W., S.L., and A.D. All authors approved the final submitted draft.

Competing interests. The authors declare none.

Data availability statement. The labels for training and testing the U-net approach are published at: https://doi.org/10.5281/zenodo.15723721. The Sentinel-2 imagery from which SGLs are derived is freely available from the Copernicus Browser at https://browser.dataspace.copernicus.eu/. The SGL data products used for the comparison are available as follows: Stokes et al. (2019): https://doi.org/10.5285/a9f2e4b5-9c2e-4ea5-8c0c-db5f6585128a; Moussavi et al. (2020a): https://doi.org/10.15784/601401; Corr et al. (2022): https://doi.org/10.5281/zenodo.5642755; Dell et al. (2024): https://doi.org/10.17863/CAM.108421; and Dirscherl et al. (2020) and the U-net approach: https://doi.org/10.5281/zenodo.15723721. The data used for the base maps is available from the British Antarctic Survey (https://data.bas.ac.uk/) and the Antarctic Digital Database (https://add.scar.org/).

Funding statement. This work was supported by the European Space Agency (ESA) under Contract No. 4000141701/23/I-DT-bgh within the framework of the "Future EO-1 EO Science for Society Permanently Open Call".

References

Arthur JF, **Stokes C**, **Jamieson SS**, **Carr JR and Leeson AA** (2020) Recent understanding of Antarctic supraglacial lakes using satellite remote sensing. *Progress in Physical Geography: Earth and Environment 44*(6), 837–869. https://doi.org/10.1177/0309133320916114.**Arthur JF**, **Stokes CR**, **Jamieson SSR**, **Rachel Carr J**, **Leeson AA and Verjans V** (2022) Large interannual variability in supraglacial lakes around East Antarctica. *Nature Communications 13*(1), 1711. https://doi.org/10.1038/s41467-022-29385-3.

Banwell A (2017) Ice-shelf stability questioned. *Nature 544*(7650), 306–307. https://doi.org/10.1038/544306a.

Banwell AF, **MacAyeal DR and Sergienko OV** (2013) Breakup of the Larsen B Ice Shelf triggered by chain reaction drainage of supraglacial lakes: Larsen B chain reaction lake drainage. *Geophysical Research Letters 40*(22), 5872–5876. https://doi.org/10.1002/2013GL057694.

Banwell AF, **Willis IC**, **Macdonald GJ**, **Goodsell B and MacAyeal DR** (2019) Direct measurements of ice-shelf flexure caused by surface meltwater ponding and drainage. *Nature Communications 10*(1), 730. https://doi.org/10.1038/s41467-019-08522-5.

Baumhoer CA, **Dietz AJ**, **Heidler K and Kuenzer C** (2023) IceLines – A new data set of Antarctic ice shelf front positions. *Scientific Data 10*(1), 138. https://doi.org/10.1038/s41597-023-02045-x.

Bell RE, **Banwell AF**, **Trusel LD and Kingslake J** (2018) Antarctic surface hydrology and impacts on ice-sheet mass balance. *Nature Climate Change 8*(12), 1044–1052. https://doi.org/10.1038/s41558-018-0326-3.

Box JE and Ski K (2007) Remote sounding of Greenland supraglacial melt lakes: Implications for subglacial hydraulics. *Journal of Glaciology 53*(181), 257–265. https://doi.org/10.3189/172756507782202883.

Burton-Johnson A, **Black M**, **Fretwell PT and Kaluza-Gilbert J** (2016) An automated methodology for differentiating rock from snow, clouds and sea in Antarctica from Landsat 8 imagery: A new rock outcrop map and area estimation for the entire Antarctic continent. *The Cryosphere 10*(4), 1665–1677. https://doi.org/10.5194/tc-10-1665-2016.

Buzzard SC, **Feltham DL and Flocco D** (2018) A mathematical model of melt lake development on an Ice Shelf. *Journal of Advances in Modeling Earth Systems 10*(2), 262–283. https://doi.org/10.1002/2017MS001155.

Chen F (2021) Comparing methods for segmenting supra-Glacial lakes and surface features in the Mount Everest region of the Himalayas using Chinese GaoFen-3 SAR images. *Remote Sensing 13*(13), 2429. https://doi.org/10.3390/rs13132429.

Chen L-C, **Papandreou G**, **Kokkinos I**, **Murphy K and Yuille AL** (2018) DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence 40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.

Corr D, **Leeson A**, **McMillan M**, **Zhang C and Barnes T** (2022) An inventory of supraglacial lakes and channels across the West Antarctic Ice Sheet. *Earth System Science Data 14*(1), 209–228. https://doi.org/10.5194/essd-14-209-2022.

De Roda Husman S, **Lhermitte S**, **Bolibar J**, **Izeboud M**, **Hu Z**, **Shukla S**, **Van Der Meer M**, **Long D and Wouters B** (2024) A high-resolution record of surface melt on Antarctic ice shelves using multi-source remote sensing data and deep learning. *Remote Sensing of Environment 301*, 113950. https://doi.org/10.1016/j.rse.2023.113950.

Dell RL, **Banwell AF**, **Willis IC**, **Arnold NS**, **Halberstadt ARW**, **Chudley TR and Pritchard HD** (2022) Supervised classification of slush and ponded water on Antarctic ice shelves using Landsat 8 imagery. *Journal of Glaciology 68*(268), 401–414. https://doi.org/10.1017/jog.2021.114.

Dell RL, **Willis IC**, **Arnold NS**, **Banwell AF and De Roda Husman S** (2024) Substantial contribution of slush to meltwater area across Antarctic ice shelves. *Nature Geoscience 17*(7), 624–630. https://doi.org/10.1038/s41561-024-01466-6.

Dirscherl M, **Dietz AJ**, **Kneisel C and Kuenzer C** (2020) Automated mapping of Antarctic supraglacial lakes using a machine learning approach. *Remote Sensing 12*(7), 1203. https://doi.org/10.3390/rs12071203.

Dirscherl M, **Dietz AJ**, **Kneisel C and Kuenzer C** (2021a) A novel method for automated supraglacial lake mapping in Antarctica using Sentinel-1 SAR imagery and deep learning. *Remote Sensing 13*(2), 197. https://doi.org/10.3390/rs13020197.

Dirscherl M, **Dietz AJ and Kuenzer C** (2021b) Seasonal evolution of Antarctic supraglacial lakes in 2015–2021 and links to environmental controls. *The Cryosphere 15*(11), 5205–5226. https://doi.org/10.5194/tc-15-5205-2021.

European Space Agency (2022) Sentinel-2 MSI Level-1C TOA Reflectance. Available at https://doi.org/10.5270/S2_-742ikth (accessed 21 June 2025).

Feyisa GL, **Meilby H**, **Fensholt R and Proud SR** (2014) Automated water extraction index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment 140*, 23–35. https://doi.org/10.1016/j.rse.2013.08.029.

Gerrish L, **Ireland L**, **Fretwell PT and Cooper P** (2024, November 12) High Resolution Vector Polygons of the Antarctic Coastline - VERSION 7.10. (Version 7.10). NERC EDS UK Polar Data Centre, 2 files, 212 MB. Available at https://doi.org/10.5285/4ECD795D-E038-412F-B430-251B33FC880E (accessed 21 June 2025).

Gilbert E and Kittel C (2021) Surface melt and runoff on Antarctic Ice Shelves at 1.5°C, 2°C, and 4°C of future warming. *Geophysical Research Letters 48*(8), e2020GL091733. https://doi.org/10.1029/2020GL091733.

Hu J, **Huang H**, **Chi Z**, **Cheng X**, **Wei Z**, **Chen P**, **Xu X**, **Qi S**, **Xu Y and Zheng Y** (2022) Distribution and evolution of supraglacial lakes in Greenland during the 2016–2018 melt seasons. *Remote Sensing 14*(1), 55. https://doi.org/10.3390/rs14010055.

Hubbard B, **Luckman A**, **Ashmore DW**, **Bevan S**, **Kulessa B**, **Kuipers Munneke P**, **Philippe M**, **Jansen D**, **Booth A**, **Sevestre H**, **Tison J-L**, **O'Leary M and Rutt I** (2016) Massive subsurface ice formed by refreezing of ice-shelf melt ponds. *Nature Communications 7*(1), 11897. https://doi.org/10.1038/ncomms11897.

Jiang D, **Li X**, **Zhang K**, **Marinsek S**, **Hong W and Wu Y** (2022) Automatic supraglacial lake extraction in Greenland using Sentinel-1 SAR images and attention-based U-net. *Remote Sensing 14*(19), 4998. https://doi.org/10.3390/rs14194998.

Lai C-Y, **Kingslake J**, **Wearing MG**, **Chen P-HC**, **Gentine P**, **Li H**, **Spergel JJ and Van Wessem JM** (2020) Vulnerability of Antarctica's ice shelves to meltwater-driven fracture. *Nature 584*(7822), 574–578. https://doi.org/10.1038/s41586-020-2627-8.

Liang Y-L, **Colgan W**, **Lv Q**, **Steffen K**, **Abdalati W**, **Stroeve J**, **Gallaher D and Bayou N** (2012) A decadal investigation of supraglacial lakes in West Greenland using a fully automatic detection and tracking algorithm. *Remote Sensing of Environment 123*, 127–138. https://doi.org/10.1016/j.rse.2012.03.020.

Lutz K, **Bahrami Z and Braun M** (2023) Supraglacial lake evolution over Northeast Greenland using deep learning methods. *Remote Sensing 15*(17), 4360. https://doi.org/10.3390/rs15174360.

Machguth H, **MacFerrin M**, **Van As D**, **Box JE**, **Charalampidis C**, **Colgan W**, **Fausto RS**, **Meijer HAJ**, **Mosley-Thompson E and Van De Wal RSW** (2016) Greenland meltwater storage in firn limited by near-surface ice formation. *Nature Climate Change 6*(4), 390–393. https://doi.org/10.1038/nclimate2899.

Morriss BF, **Hawley RL**, **Chipman JW**, **Andrews LC**, **Catania GA**, **Hoffman MJ**, **Lüthi MP and Neumann TA** (2013) A ten-year record of supraglacial lake evolution and rapid drainage in West Greenland using an automated processing algorithm for multispectral imagery. *The Cryosphere 7*(6), 1869–1877. https://doi.org/10.5194/tc-7-1869-2013.

Moussavi M, **Abdalati W**, **Halberstadt ARW**, **Pope A and Trusel LD** (2020a) Supraglacial Lakes in Antarctica. (Version 1). U.S. Antarctic Program (USAP) Data Center. Available at https://doi.org/10.15784/601401 (accessed 21 June 2025).

Moussavi MS, **Abdalati W**, **Pope A**, **Scambos T**, **Tedesco M**, **MacFerrin M and Grigsby S** (2016) Derivation and validation of supraglacial lake volumes on the Greenland Ice Sheet from high-resolution satellite imagery. *Remote Sensing of Environment 183*, 294–303. https://doi.org/10.1016/j.rse.2016.05.024.

Moussavi M, **Pope A**, **Halberstadt A**, **Trusel L**, **Cioffi L and Abdalati W** (2020b) Antarctic supraglacial lake detection using Landsat 8 and Sentinel-2 imagery: Towards continental generation of lake volumes. *Remote Sensing 12*(1), 134. https://doi.org/10.3390/rs12010134.

Niu L, **Tang X**, **Yang S**, **Zhang Y**, **Zheng L and Wang L** (2023) Detection of Antarctic surface meltwater using Sentinel-2 remote sensing images via U-net with attention blocks: A case study over the Amery Ice Shelf. *IEEE Transactions on Geoscience and Remote Sensing 61*, 1–13. https://doi.org/10.1109/TGRS.2023.3275076.

Pignatale FC (2022, November 21) Sen2Cor 2.11.00 Software Release Note. Available at https://step.esa.int/thirdparties/sen2cor/2.11.0/docs/OMPC.TPZG.SRN.003%20-%20i1r0%20-%20Sen2Cor%202.11.00%20Software%20Release%20Note.pdf (accessed 21 June 2025).

Pope A, **Scambos TA**, **Moussavi M**, **Tedesco M**, **Willis M**, **Shean D and Grigsby S** (2016) Estimating supraglacial lake depth in West Greenland using Landsat 8 and comparison with other multispectral methods. *The Cryosphere 10*(1), 15–27. https://doi.org/10.5194/tc-10-15-2016.

Robel AA and Banwell AF (2019) A speed limit on Ice Shelf collapse through hydro fracture. *Geophysical Research Letters 46*(21), 12092–12100. https://doi.org/10.1029/2019GL084397.

Ronneberger O, **Fischer P and Brox T** (2015) U-net: Convolutional networks for biomedical image segmentation. In Navab N, Hornegger J, Wells WM and Frangi AF (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Vol. *9351*. Cham: Springer International Publishing, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Rott H, **Abdel Jaber W**, **Wuite J**, **Scheiblauer S**, **Floricioiu D**, **Van Wessem JM**, **Nagler T**, **Miranda N and Van Den Broeke MR** (2018) Changing pattern of ice flow and mass balance for glaciers discharging into the Larsen A and B embayments, Antarctic Peninsula, 2011 to 2016. *The Cryosphere 12*(4), 1273–1291. https://doi.org/10.5194/tc-12-1273-2018.

Royston S and Gudmundsson GH (2016) Changes in ice-shelf buttressing following the collapse of Larsen A Ice Shelf, Antarctica, and the resulting impact on tributaries. *Journal of Glaciology 62*(235), 905–911. https://doi.org/10.1017/jog.2016.77.

Shen X-Y, **Ke C-Q**, **Li H-L**, **Cai Y**, **Xiao Y and Li M-M** (2025) Evolution of supraglacial lakes over the pan-Antarctic ice sheet between 2014 and 2022: Assessment and the control factors. *Advances in Climate Change Research 16*(2), 337–349. https://doi.org/10.1016/j.accre.2025.02.005.

Sneed WA and Hamilton GS (2007) Evolution of melt pond volume on the surface of the Greenland Ice Sheet. *Geophysical Research Letters 34*(3), 2006GL028697. https://doi.org/10.1029/2006GL028697.

Spergel JJ, **Kingslake J**, **Creyts T**, **Van Wessem M and Fricker HA** (2021) Surface meltwater drainage and ponding on Amery Ice Shelf, East Antarctica, 1973–2019. *Journal of Glaciology 67*(266), 985–998. https://doi.org/10.1017/jog.2021.46.

Stokes CR, **Sanderson JE**, **Miles BWJ**, **Jamieson SSR and Leeson AA** (2019) Widespread distribution of supraglacial lakes around the margin of the East Antarctic Ice Sheet. *Scientific Reports 9*(1), 13823. https://doi.org/10.1038/s41598-019-50343-5.

Tedesco M, **Willis IC**, **Hoffman MJ**, **Banwell AF**, **Alexander P and Arnold NS** (2013) Ice dynamic response to two modes of surface lake drainage on the Greenland ice sheet. *Environmental Research Letters 8*(3), 034007. https://doi.org/10.1088/1748-9326/8/3/034007.

Trusel LD, **Frey KE**, **Das SB**, **Karnauskas KB**, **Kuipers Munneke P**, **Van Meijgaard E and Van Den Broeke MR** (2015) Divergent trajectories of Antarctic surface melt under two twenty-first-century climate scenarios. *Nature Geoscience 8*(12), 927–932. https://doi.org/10.1038/ngeo2563.

Trusel LD, **Pan Z and Moussavi M** (2022) Repeated tidally induced hydro fracture of a supraglacial lake at the Amery Ice Shelf grounding zone. *Geophysical Research Letters 49*(7), e2021GL095661. https://doi.org/10.1029/2021GL095661.

Tuckett PA, **Ely JC**, **Sole AJ**, **Lea JM**, **Livingstone SJ**, **Jones JM and Van Wessem JM** (2021) Automated mapping of the seasonal evolution of surface meltwater and its links to climate on the Amery Ice Shelf, Antarctica. *The Cryosphere 15*(12), 5785–5804. https://doi.org/10.5194/tc-15-5785-2021.

Tuckett PA, **Ely JC**, **Sole AJ**, **Livingstone SJ**, **Davison BJ**, **Van Wessem JM and Howard J** (2019) Rapid accelerations of Antarctic Peninsula outlet glaciers driven by surface melt. *Nature Communications 10*(1), 4311. https://doi.org/10.1038/s41467-019-12039-2.

Van Wessem JM, **Den Broeke MR**, **Wouters B and Lhermitte S** (2023) Variable temperature thresholds of melt pond formation on Antarctic ice shelves. *Nature Climate Change 13*(2), 161–166. https://doi.org/10.1038/s41558-022-01577-1.

Wehrlé A, **Lüthi MP**, **Kneib-Walter A**, **Nap A**, **Rousseau H**, **Jouvet G and Walter F** (2026) Velocity and calving response of a major Greenland ice stream to a lake drainage event. *Nature Geoscience 19*(1), 84–89. https://doi.org/10.1038/s41561-025-01858-2.

Wessel B, **Huber M**, **Wohlfart C**, **Bertram A**, **Osterkamp N**, **Marschalk U**, **Gruber A**, **Reuß F**, **Abdullahi S**, **Georg I and Roth A** (2021) TanDEM-X PolarDEM 90 m of Antarctica: Generation and error characterization. *The Cryosphere 15*(11), 5241–5260. https://doi.org/10.5194/tc-15-5241-2021.

Williamson AG, **Willis IC**, **Arnold NS and Banwell AF** (2018) Controls on rapid supraglacial lake drainage in West Greenland: An exploratory data analysis approach. *Journal of Glaciology 64*(244), 208–226. https://doi.org/10.1017/jog.2018.8.

Yang K and Smith LC (2013) Supraglacial streams on the Greenland Ice Sheet delineated from combined spectral–shape information in high-resolution satellite imagery. *IEEE Geoscience and Remote Sensing Letters 10*(4), 801–805. https://doi.org/10.1109/LGRS.2012.2224316.

Cite this article: Koehler J, Baumhoer C, Wouters B, Lhermitte S and Dietz A (2026). Advancing supraglacial lake mapping in Antarctica: A deep learning approach using Sentinel-2 imagery. *Environmental Data Science*, 5: e38. doi:10.1017/eds.2026.10065