https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Self-supervised learning reduces labelling requirements for sea ice segmentation in Sentinel-1 SAR imagery

Jacob Seston[1], William D. Harcourt[1,2], Georgios Leontidis[2,3], Brice R. Rea[1], Matteo Spagnolo[4], Lauren McWhinnie[5], [6]

1School of Geosciences, University of Aberdeen, Aberdeen, AB24 3UF, UK
2Interdisciplinary Institute, University of Aberdeen, Aberdeen, AB24 3UF, UK
3School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, AB24 3UF, UK
4Department of Earth Sciences, University of Torino, Torino, 10125 Torino TO, Italy
5Institute of Life and Earth Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK
6Department of Geography, University of Victoria, British Columbia, V8W 3P5, Canada

Correspondence to: Jacob Seston (j.seston.23@abdn.ac.uk)

Abstract. Monitoring Arctic sea ice variability is crucial for maritime safety. Synthetic Aperture Radar (SAR) imagery provides an effective means of achieving this through all-weather, day-and-night coverage of the Arctic. Navigation in the Canadian Arctic Archipelago currently relies on operational ice information services, including analyst-derived ice charts, satellite imagery, and ice routing products provided by national ice services. However, the development of machine-learning systems capable of automatically processing large volumes of satellite imagery and accurately identifying ice conditions is constrained by the need for extensive manually labelled datasets. To address this limitation, we developed a self-supervised learning (SSL) approach, which uses unlabelled data to learn general image representations. Specifically, we use Bootstrap Your Own Latent (BYOL), a non-contrastive SSL framework, to pretrain a UNet encoder on unlabelled dual-polarised Sentinel-1 Extra-Wide mode (EW) SAR scenes before fine-tuning with a small set of labelled images. We compare the BYOL-pretrained UNet (called UNet SSL in this study) to four baselines: a control UNet, a fully supervised UNet, a Random Forest classifier, and the Segment Anything Model (SAM). With only three labelled scenes, the BYOL-pretrained UNet achieved higher segmentation accuracy than the fully supervised model trained on seven images, more than twice the number of labelled scenes. The most significant gains occurred in Marginal Ice Zone (MIZ) scenes, where the BYOL-pretrained UNet achieved a Matthews Correlation Coefficient (MCC) of 0.2087, compared with 0.1685 for the fully supervised UNet trained on seven labelled scenes and 0.1449 for the control model trained on three scenes—representing an MCC increase of approximately 24% and 44%, respectively. These improvements were accompanied by a substantial reduction in false negatives and a marked increase in recall, indicating improved discrimination under low-contrast, fragmented floe conditions. Our findings demonstrate that SSL reduces annotation requirements for SAR-based sea ice segmentation, improving model generalisation in both consolidated

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

and fragmented ice conditions. This approach offers a scalable solution to the labelling bottleneck in Arctic monitoring and highlights the potential of BYOL as a general pretraining strategy for SAR-based Earth observation image segmentation.

1. Introduction

The Arctic is warming 4-6 times faster than the rest of the planet and this has led to a reduction in Arctic sea ice extent in both summer and winter (Stroeve and Notz, 2018; Rantanen et al., 2022). This amplified warming is closely linked to diminishing sea ice and associated feedbacks between the surface energy balance and the atmosphere (Screen and Simmonds, 2010; Serreze and Barry, 2011). The sea ice extent minima, usually occurring at the start of September, has decreased by nearly 14% per decade since 1979, while the winter extent maxima continues to shrink at more than 3 % per decade (Stroeve and Notz, 2018). This loss has been accompanied by a dramatic shift in ice age structure: the proportion of Arctic sea ice older than five years has declined from approximately 28% of the basin in the mid-1980s to less than 2% by 2018, leaving the Arctic increasingly dominated by thin, first-year ice (Stroeve and Notz, 2018). The diminishing size of the sea ice pack is reshaping human activity in the region as maritime traffic through Arctic waters is expanding, including increasingly hazardous winter operations (Müller et al., 2023). At present, navigation through Arctic waters relies on sea ice charts that are produced by national ice services through expert analysis of Synthetic Aperture Radar (SAR) imagery and other satellite observations (Dierking, 2013). This creates a time lag between satellite data acquisition, manual interpretation of the imagery, and the release of the ice charts to an online platform. Sea ice charts are crucial for both operational forecasting and climate research, yet they often omit key structural features such as leads and deformation zones because current SAR-based products struggle to provide reliable secondary ice information (Hebert et al., 2015; Sandven et al., 2023), despite their importance as hot spots of ocean–atmosphere interaction (Kortum et al., 2024). Automated methods that reduce manual labelling could accelerate high-resolution sea ice mapping useful for navigation, short-term forecasting of ice conditions, and climate research. However, automating the process of sea-ice mapping is hindered by three linked challenges: the scarcity of labelled SAR data, the inability of existing products to resolve fine-scale structural features, and the difficulty of producing timely, consistent maps at the pace of modern satellite acquisitions.

Deep learning techniques offer a potential solution to these issues by utilising the extensive collections of unlabelled SAR imagery, enabling models to learn transferable representations without requiring large volumes of manually labelled data. SAR is particularly suitable for Arctic monitoring because it can acquire data under all

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

weather conditions and during the polar night, unlike optical sensors, which are limited by cloud cover and the absence of sunlight (Zakhvatkina et al., 2019; Yuan et al., 2025). Methods based on Convolutional Neural Networks (CNNs), such as UNet (Ronneberger et al., 2015), have proven effective for SAR-based sea ice segmentation (Park et al., 2020; Huang et al., 2024), but they depend on large, labelled data sets, which are scarce in the Arctic (Khaleghian et al., 2021; Jiang et al., 2024) and are labour intensive to produce. These challenges have prompted the development of architectures specifically tailored to SAR, such as hybrid convolutional-transformer models designed to better capture spatial and contextual features in radar backscatter (Ristea et al., 2023). For example, Boulze et al. (2020) achieved ~90–92% accuracy in classifying four ice types from Sentinel-1 dual-polarisation SAR, outperforming a texture-based Random Forest while cutting per-scene processing from ~1 hour to ~2 minutes. However, both Random Forests and CNNs struggled with young and first-year ice because of coarse or inconsistent labels and mixed SAR pixels in regions close to manually drawn ice-chart polygon boundaries, where chart generalisation causes a single label to span multiple physical ice types, emphasising that performance is fundamentally constrained by label quality and availability. Manual annotation is costly, time-consuming, subjective, and geographically limited, creating a bottleneck between the vast archives of unlabelled SAR imagery and the comparatively small pool of expert-labelled masks.

Several labelled SAR sea-ice datasets have been developed that complement the present study. The AI4Arctic Sea Ice Challenge dataset and the associated AutoICE Challenge (Stokholm et al., 2024) provide large-scale, multi-class annotations derived primarily from operational ice charts. While highly valuable for regional benchmarking, the chart-based approach results in intentionally generalised polygon boundaries, which are less suited to evaluating pixel-level segmentation accuracy and boundary fidelity.

Similarly, the MOSAiC-based binary lead dataset (Murashkin, 2023) offers extensive coverage of lead structures in the central Arctic Ocean, but the labels are generated using a convolutional neural network rather than manual annotation. As such, it is well suited to statistical analyses of lead occurrence, but less appropriate as an independent reference for evaluating learning-based segmentation methods. In contrast, the present study uses a small number of expert-annotated, pixel-wise masks, designed to support controlled, relative comparisons between models under limited-label conditions. Future work will extend the BYOL-pretrained framework to these larger community datasets.

Self-supervised learning (SSL) has emerged as a promising way to reduce the dependence of deep learning models on labelled data. Instead of relying on human annotations, SSL learns by solving proxy objectives, such as predicting masked content or enforcing consistency across augmented views, that lead the model to acquire useful representations. These tasks typically involve predicting missing information or ensuring that two differently augmented versions of the same image produce similar internal representations (Grill et al., 2020). These

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

representations capture structural and semantic patterns that can later be transferred to downstream tasks with only minimal supervision, often rivalling or surpassing fully supervised baselines (Grill et al., 2020).

Some SSL methods, such as Bootstrap Your Own Latent (BYOL; Grill et al., 2020), remove the need for negative pairs—image patches that are explicitly treated as representing different underlying classes or features during training. In contrastive self-supervised learning frameworks, representation learning is driven by simultaneously pulling together positive pairs and pushing apart negative pairs, meaning that the definition of what constitutes a "negative" sample is central to the training objective. These definitions rely on human design choices. Incorrectly defining negative pairs can actively harm representation learning by forcing the model to separate samples that are physically or semantically related. This is particularly problematic in SAR imagery, where visual dissimilarity arising from incidence-angle effects, speckle, or surface roughness does not necessarily correspond to a true semantic difference between ice types (Casey et al., 2016). Instead of comparing a given image patch against many assumed negatives, BYOL operates using two differently augmented views of the same SAR scene. One network (the online branch) is trained to predict the representation produced by a second, slowly updated target network. Because both views originate from the same underlying ice feature, the model is encouraged to learn stable SAR structures—such as floe texture or ridge geometry—that persist across imaging conditions (Lensu et al., 2022), rather than scene-dependent variations driven by incidence angle or surface state (Macdonald et al., 2024). By eliminating the need to define what constitutes a "dissimilar" SAR sample, BYOL avoids false negatives and reduces sensitivity to scene-specific noise and speckle, making it particularly well suited to SAR data, where complex spatial textures and acquisition-dependent effects can undermine contrastive training schemes that rely on explicit similarity–dissimilarity assumptions.

Recent work demonstrates the growing use of SSL for satellite scene analysis, where annotated data are often scarce and expensive to produce. In optical imagery, SSL methods have been used to cluster or classify land-cover patterns without labels, supporting expert interpretation at regional to national scales (Francis et al., 2023). In hyperspectral imaging, SSL has enabled efficient spectral super-resolution using compact networks trained on synthetic degradation pairs, achieving competitive accuracy with minimal compute and without large labelled corpora (Rajaei et al., 2024). These studies illustrate several advantages of SSL for Earth observation: the ability to exploit abundant unlabelled satellite data, reduced dependence on expert annotation, and the capacity to learn robust spatial and spectral structure with lightweight models. SSL has also begun to gain traction in optical remote sensing (Muzeau et al., 2024) and in SAR scene-level classification (Liu et al., 2024); however, its application to pixel-level segmentation of Arctic sea ice in SAR imagery remains largely unexplored.

While these advances demonstrate the versatility of SSL, SAR-based sea-ice mapping presents additional physical complexities such as speckle, noise, incidence-angle effects, and the sensitivity of backscatter to ice type and

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

surface roughness (Johansson et al., 2018; Lohse et al., 2020; Karlsen et al., 2024). Sea ice itself is heterogeneous and highly dynamic, further complicating classification and monitoring, with SAR backscatter signatures varying seasonally across leads, young ice, first-year ice (FYI), and multi-year ice (MYI) (Guo et al., 2023). Physically, MYI is thicker, salt-depleted, and heavily deformed, characterised by hummocks and ridges that produce bright, granular radar returns, whereas FYI tends to be smoother, more saline, and less consolidated, yielding lower backscatter, except where deformation or surface flooding by seawater increases roughness (Roach et al., 2025). Leads—narrow fractures of open water within the ice pack—occupy a small fraction of the ice cover but represent conduits for the exchange of heat and moisture between the ocean and atmosphere, making their detection particularly useful in studying the energy balance of sea ice (Clemens-Sewall et al., 2023). Their spatial distribution and evolution can be quantified using SAR-derived divergence and deformation metrics (von Albedyll et al., 2024). The Marginal Ice Zone (MIZ) marks the transition from open ocean to pack ice and exhibits extreme spatial and temporal variability, driven by wind, waves, and ocean currents. Its fine-scale fragmentation and overlapping backscatter signatures present major operational challenges, making it difficult to develop segmentation methods that generalise across scattering conditions (Huang and Li, 2023; Itkin, 2025). These characteristics make SAR images of sea ice an ideal but demanding test case for evaluating self-supervised approaches.

Here we introduce, to our knowledge, the first application of non-contrastive self-supervised pretraining (BYOL) to SAR- sea-ice segmentation, evaluating whether such pretraining can reduce annotation requirements without sacrificing accuracy. We pretrain a UNet encoder on unlabelled Sentinel-1 SAR imagery and fine-tune it with limited labelled scenes, which we then compare against widely used models for image segmentation: a fully supervised UNet, a Random Forest classifier, and the Segment Anything Model (SAM). Our research questions are:

How does the segmentation performance of a BYOL-pretrained UNet compare to other widely used models for sea ice segmentation?

How do models perform across contrasting Arctic environments — from consolidated multi-year ice to fragmented marginal-ice zones — and what does this reveal about each model's robustness to different radar scattering conditions? The objective of this study is not to establish state-of-the-art accuracy of sea-ice segmentation models, but to assess their relative performance under controlled, low-label conditions. Specifically, we evaluate whether self-supervised pretraining using BYOL enables a UNet model to achieve performance comparable to, or exceeding, fully supervised baselines while using substantially fewer labelled SAR scenes. All models are therefore trained and evaluated on the same fixed test scenes, and performance differences are interpreted in a comparative sense, focusing on relative gains in robustness, generalisation, and label efficiency rather than on absolute accuracy values.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

2. Study area and data

2.1 Study area

Our study area (Fig. 1) encompasses the western side of the Canadian Arctic Archipelago (CAA; ~75°–83° N, 90°–125° W) and adjacent Arctic Ocean. The region was chosen because it presents a challenging and operationally important environment for sea-ice segmentation: a dense network of narrow straits and islands where land contamination, mixed ice types, and frequent thin ice leads complicate SAR interpretation (Howell et al., 2024). As a major pathway for sea-ice export and a key sector of the Northwest Passage (Cook et al., 2024), the CAA also provides a scientifically relevant testbed for evaluating model performance in conditions that are both climatically and logistically significant.

The climate of the CAA is strongly influenced by the presence and variability of sea ice (Ye et al., 2025), which plays a key role in modulating atmospheric circulation and driving regional weather extremes. Sea ice in the CAA comprises a mixture of FYI that forms each winter and MYI that survives one or more melt seasons. MYI within the CAA is replenished from the Arctic Ocean and the survival of FYI through the melt season, with interannual trends suggesting MYI has remained stable between 2016 and 2022 (Howell et al., 2024). The presence of both MYI and FYI leads to a complex mosaic of sea ice across the CAA: deformed MYI remains concentrated along the northern coasts of the CAA, whilst thinner FYI dominates the southern channels and coastal inlets. Leads and polynyas – open-water features within the pack ice form and evolve throughout the year, enhancing ocean–atmosphere heat exchange and often marking the transition between melt and freeze phases (Roach et al., 2025). Recent climate warming has led to an earlier breakup of sea ice in summer, a longer melt period, and an increase in open-water areas as the pack transitions to younger, thinner FYI (Howell & Brady, 2019). In 2024, September Arctic sea ice extent was the 6th lowest in the satellite era, continuing the long-term decline in total sea ice extent (Meier et al., 2024).

The CAA has become an increasingly important focus of Arctic shipping, with voyages through Canadian Arctic waters more than quadrupling since 1990 as declining sea ice cover improves seasonal navigability along parts of the Northwest Passage (Cook et al., 2024). However, choke points formed by persistent MYI in narrow channels continue to restrict access, reducing the effective length of the shipping season (Cook et al., 2024). While summer navigation increasingly favours the shallower southern route of the Northwest Passage, the deep-water northern route remains constrained by MYI choke points, with sea-ice area dropping to ~4×10^3 km² at the end of September 2024 (Howell et al., 2025). Accurate, high-resolution monitoring of leads and ice conditions in the CAA and neighbouring Arctic Ocean is vital for safe navigation and planning suitable routes.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Sentinel-1 SAR imagery

We conduct our evaluation using Sentinel-1 SAR scenes from the CAA. We choose scenes that represent two contrasting environments: (1) a consolidated ice pack containing leads (Scene 1), and (2) scenes covering the fragmented MIZ (Scene 2). We used dual-polarised Sentinel-1 scenes acquired in Extra Wide (EW) swath mode and Ground Range Detected (GRD) format. In total, 21 Sentinel-1 scenes were used in this study (Table 1): 12 unlabelled scenes were used for self-supervised pretraining, whilst a total of 9 scenes were manually labelled for training (7 images) and test (2 images). Each pre-processed EW acquisition has an image swath size of 7000 × 7500 pixels at 80 m resolution (~560x600 km image size), corresponding to roughly 5 × 107 pixels per scene and nearly 1 × 109 SAR pixels across the dataset. The labelled acquisitions span from June 2022 to November 2023, capturing seasonal variability across all four Arctic seasons including freeze-up (autumn), maximum extent (winter), melt onset (spring), and sea ice minima (summer). The labelled Sentinel-1 SAR scenes are primarily located over the Queen Elizabeth Islands (the northernmost Canadian archipelago), including: Prince Patrick, Ellef Ringnes, and Mackenzie Islands, with additional coverage extending southward to the waters between Bathurst and Melville Islands and northward into the Arctic Ocean. One scene is centred roughly 500 km north of mainland Alaska, capturing consolidated MYI beyond the island chain. Table 1. Summary of the list of images used in training and testing

File Name	Scene ID (Fig. 1)	Date	Labelled?	Training / Test
S1A_EW_GRDM_1SDH_20180116T075430_20180116T075530_020177_0226B9_9FE3	n/a	16.01.2018	Unlabelled	SSL Training
S1B_EW_GRDM_1SDH_20180213T175444_20180213T175544_009608_011511_8266	n/a	13.02.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20180313T181225_20180313T181325_021000_0240E1_8163	n/a	13.03.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20180417T074606_20180417T074706_021504_0250C3_D211	n/a	17.04.2018	Unlabelled	SSL Training
S1B_EW_GRDM_1SDH_20180515T174633_20180515T174733_010935_01403A_A84D	n/a	15.05.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20180612T180423_20180612T180523_022327_026AB3_	n/a	12.06.2018	Unlabelled	SSL Training

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

AC33				
S1A_EW_GRDM_1SDH_20180717T073809_20180717T073909_022831_0279B9_EBF1	n/a	17.07.2018	Unlabelled	SSL Training
S1B_EW_GRDM_1SDH_20180814T075344_20180814T075444_012256_016952_B1DC	n/a	14.08.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20180911T175548_20180911T175652_023654_0293F5_7CA2	n/a	11.09.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20181016T072958_20181016T073058_024158_02A460_DA8F	n/a	16.10.2018	Unlabelled	SSL Training
S1B_EW_GRDM_1SDH_20181113T074529_20181113T074629_013583_019254_D382	n/a	13.11.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20181218T075437_20181218T075537_025077_02C472_1DB2	n/a	18.12.2018	Unlabelled	SSL Training
S1A_EW_GRDM_1SDH_20221027T161558_20221027T161702_045630_0574C3_1CD3	6	27.10.2022	Labelled	Test
S1A_EW_GRDM_1SDH_20230203T150150_20230203T150254_047073_05A59D_A02F	3	03.02.2023	Labelled	SL Training
S1A_EW_GRDM_1SDH_20230304T151108_20230304T151208_047496_05B3E5_1FD1	7	04.03.2023	Labelled	SL Training
S1A_EW_GRDM_1SDH_20230305T155314_20230305T155414_047511_05B46C_E347	8	05.03.2023	Labelled	Test
S1A_EW_GRDM_1SDH_20230503T151005_20230503T151109_048371_05D160_BE0E	4	03.05.2023	Labelled	SL Training
S1A_EW_GRDM_1SDH_20230602T142043_20230602T142147_048808_05DE9A_F138	1	02.06.2023	Labelled	SL Training
S1A_EW_GRDM_1SDH_20230805T134856_20230805T134956_049741_05FB2E_E2F3	2	05.08.2023	Labelled	SL Training

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

S1A_EW_GRDM_1SDH_20230925T155119_20230925T155224_050486_061495_6FA5	5	25.09.2023	Labelled	SL Training
S1A_EW_GRDM_1SDH_20231110T174737_20231110T174837_051158_062B90_1942	9	10.11.2023	Labelled	SL Training

All Sentinel-1 SAR images were processed using the ESA Sentinel Application Platform (SNAP) following a standard processing chain (Filipponi, 2019):

1. Remove GRD border noise
2. Radiometric calibration to sigma-nought (σ^0) in decibels (dB);
3. Speckle filtering using the Lee Sigma algorithm with a 3 × 3 window;
4. Multilooking with two range and two azimuth looks;
5. Ellipsoid correction using the WGS84 ellipsoid model;
6. Export to GeoTIFF format for integration with machine learning pipelines.
7. Following preprocessing, the spatial resolution of Sentinel-1 scenes was 80 m.

Sentinel-1 Extra-Wide (EW) mode acquisitions exhibit systematic incidence-angle variation across the swath, which affects σ^0 magnitude and backscatter texture. In this study, Sentinel-1 scenes were processed using the standard SNAP radiometric workflow, which partially mitigates incidence-angle effects through σ^0 calibration. No additional explicit incidence-angle normalisation or flattening was applied, as the objective of the study is to compare relative model performance under identical acquisition conditions rather than to optimise absolute accuracy for a single corrected representation. Residual incidence-angle effects are instead handled implicitly through the use of dual-polarised HH–HV inputs and SAR-specific data augmentations designed to encourage invariance to acquisition-related variability. The influence of these effects is further examined through σ^0-resolved performance analysis (Sect. 4.4), which links segmentation skill to physical scattering regimes that partially co-vary with incidence angle.

2.3 Sea ice labels

We manually annotated 9 Sentinel-1 scenes into two classes: sea ice and open water. All annotations are binary. Features such as marginal ice zones (MIZ), leads, thin ice, and melt ponds are not treated as separate classes, but are handled through explicit labelling rules designed to ensure consistency under ambiguous scattering conditions.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Land areas were masked using the "2020 Land Cover of Canada" dataset from Open Canada (Latifovic, 2022). A consistent labelling protocol was applied to allow reproducibility across scenes and seasons (see Figs. 7 and 8 for examples of Test labels):

- Sea Ice was defined as any continuous area of elevated σ^0 in the HH and HV bands, including both thick consolidated floes and thinner, newly formed ice.
- Open Water was characterised by dark, low-return backscatter. Textured water, caused by wind or waves, was also classified as open water.
- Marginal Ice Zones (MIZ) were handled using a consistent labelling convention: within spatially heterogeneous regions containing mixed ice and water scattering, pixels were assigned to the ice class when the dominant contiguous scattering signature corresponded to sea ice, even if intermixed with open-water returns.
- Leads narrower than ~200 m (~2–3 pixels at 80 m resolution) were excluded to ensure consistent detection thresholds.

Where available, Sentinel-2 optical imagery was used for visual cross-reference, particularly to differentiate melt ponds, wind-driven open water, and MIZ features. The final labelled dataset covers all four seasons and includes a range of ice regimes, from dense winter pack ice to fragmented summer floes.

The two test scenes were intentionally selected to represent contrasting levels of segmentation difficulty. The first scene corresponds to relatively consolidated ice conditions and is treated as an easier reference case, whereas the second scene encompasses a Marginal Ice Zone (MIZ) characterised by fragmented ice and mixed scattering, and is used as a challenging test case to assess model robustness under more complex conditions.

To contextualise the labelling approach used in this study, Fig. 2 provides a visual comparison between the manually annotated, pixel-wise labels developed here and two widely used community datasets: the AI4Arctic Sea Ice Challenge dataset and the MOSAiC-based binary lead dataset. The comparison highlights differences in label provenance and spatial granularity, reflecting the distinct objectives of each dataset. While chart-derived and automatically generated labels are well suited to large-scale statistical analyses and benchmarking, the manually produced labels used in this study are designed to support controlled, pixel-level evaluation of segmentation performance under limited-label conditions.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

3. Model intercomparison

An overview of the models and training regimes is shown in Fig. 3.

3.1 Models

In this study, we evaluate the performance of five approaches to segment sea ice using dual-polarised Sentinel-1 SAR imagery:

- UNet (Control): Supervised UNet trained using three labelled scenes (low-data baseline). UNet is widely used in remote sensing and medical imaging as the standard architecture for pixel-wise segmentation.
- UNet (Supervised Learning): Supervised UNet trained using seven labelled scenes. This provides a stronger supervised benchmark for evaluating how performance scales with more labelled data.
- UNet (Self Supervised Learning): UNet with encoder pretrained using SSL (BYOL) on unlabelled SAR imagery, then fine-tuned with 3 labelled scenes. This model tests our central hypothesis that self-supervised pretraining can reduce dependence on labelled data.
- Random Forest (RF): Pixel-wise classifier. Although older, RFs remain widely used for sea ice segmentation tasks due to their simplicity and interpretability (e.g. Marbouti et al, 2020). Including this model allows us to benchmark deep learning performance gains against a classical, low-complexity, baseline.
- Segment Anything Model (SAM): A prompt-based zero-shot segmentation method pretrained on large-scale natural RGB datasets. While SAM was not designed for SAR, it was trained on RGB 11M images and provides an important test of whether such generic pretrained models, so called foundation models, can transfer directly to SAR-based sea ice segmentation without fine-tuning.

3.1.1 UNet (Control and SL)

UNet++ (Zhou et al., 2018) is an extension of the original UNet architecture developed for biomedical image segmentation (Ronneberger et al., 2015). Like its predecessor, it follows an encoder–decoder structure, where the

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

encoder progressively downsamples the input to extract hierarchical features and the decoder upsamples them to produce a dense, pixel-wise segmentation mask. However, UNet++ introduces nested skip connections – dense links between encoder and decoder blocks at multiple depths – that refine feature fusion and improve gradient flow during training. These skip pathways enable finer localisation and more robust multiscale representation learning, which is particularly valuable for segmenting narrow or fragmented features such as sea-ice leads.

In this study, we use a modified UNet++ tailored to the challenges of SAR-based sea ice segmentation. Modifications include Residual convolutional blocks (He et al., 2016) with group normalisation and dropout and Channel Attention Module (CAM) (Woo et al., 2018) in the bottleneck layer. The final layer is a 1×1 convolution that maps to a single-channel probability output. Binary segmentation is achieved by applying a sigmoid activation and thresholding at inference time. Identical UNets were trained, with differing amounts of labelled data. The first was trained on 3 labelled images, called UNet (Control), whilst the second was trained with 7 labelled images, called UNet (SL). The modified architecture is illustrated in Fig. 4.

3.1.2 UNet (SSL)

To reduce reliance on labelled training data, we implemented BYOL (Grill et al., 2020), a SSL framework designed to learn image representations from unlabelled data. Unlike contrastive SSL methods that compare both positive and negative sample pairs, BYOL operates without negative pairs, instead relying on asymmetric prediction between two augmented views of the same image. No labels are used during BYOL pretraining; the network learns by matching embeddings of two augmented views of the same image. The core idea is to train an online network, which is the part of the model updated after each training step, to predict the representation of a target network, a slowly updated copy of itself, using different augmentations of the same input. In this context, augmentations refer to transformations applied to an image that preserve its semantic content (e.g. whether a pixel represents ice or water) while modifying its appearance (see Figs. 5 and 6). These transformations encourage the model to learn representations that are invariant to such changes. Specifically, each Sentinel-1 SAR image is augmented twice to create two views: one is passed through the online encoder and projection head, while the other is passed through the target encoder and projection head. The online network includes an additional prediction head that outputs a vector which is optimised to match the target's latent representation. This architecture is particularly well-suited to SAR imagery as it enables learning from large unlabelled SAR archives and is robust to noise, speckle, and acquisition variation (Xu et al., 2021).

We adapted BYOL to operate within our UNet encoder and with SAR imagery as input by replacing the original ResNet backbone with our modified UNet encoder and using HH-HV SAR pairs as 2-channel inputs. Each SAR image was augmented twice to form a positive pair, i.e. two images with different representations of the same

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

scene. The online encoder and projection head generate a latent vector – a compact numerical representation of the image in feature space – which was compared to the output of a target encoder (an exponential moving average of the online network). The loss function – a measure of how different two outputs are, guiding how the network updates its weights during training – is based on cosine similarity between the predicted and target representations (see sect. 3.3). This was used exclusively during the self-supervised pretraining phase. BYOL's loss function encourages the online network to produce representations that are invariant to augmentation and similar to the target's embeddings. Importantly, since no negative samples are required, BYOL avoids the instability often seen in contrastive methods when semantic similarity does not correlate with visual appearance—a common challenge in SAR data, where different sea ice types may appear visually similar due to backscatter ambiguity.

To qualitatively track what the encoder was learning during self-supervised pretraining, we generated feature visualisations every two epochs. For a single input, we plotted the augmented view alongside the mean and standard deviation of the deepest encoder feature maps. The mean highlights stable, view-invariant structures (Fig. 5b), and the standard deviation reflects variation across channels and sensitivity to textural differences (Fig. 5c). These visualisations were used solely for training diagnostics. Several limiting cases help in interpreting these diagnostics. A high mean with a low standard deviation indicates strong consensus across filters, suggesting that the encoder has confidently identified stable structures, such as consolidated ice. Conversely, a low mean with a high standard deviation indicates weak overall activations and high disagreement across channels, often associated with noisy or ambiguous textures. High mean and standard deviation values indicates strong but varied activations, typical of fragmented floes, such as that found in the open water channels in Fig. 5. In contrast, low mean and standard deviation values implies low information content, typical of uniform open water regions or sea ice floes (Fig. 4). Additionally, to illustrate cross-view consistency—the extent to which the encoder produces similar feature representations for two differently augmented views of the same underlying SAR scene—we include an example with two augmented views of the same patch (Fig. 6) alongside their corresponding encoder mean-feature maps. Cross-view consistency is indicated by the presence of spatially aligned high-activation (bright) regions in both feature maps, showing that the encoder emphasises the same underlying ice structures despite differences in the input augmentations.

We used 12 unlabelled Sentinel-1 scenes (HH and HV) for pretraining. Scenes were selected based on quality and geographic diversity. Augmentations were designed with SAR-specific considerations:

- Geometric transforms (random crops, flips, rotation) to simulate spatial variability;
- Contrast and brightness variation approximates changes in surface roughness and incidence angle;

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

- Affine distortion introduces geometric deformation, aiding in feature generalisation.

Each augmentation had an 85% probability of being applied. The final patch size – a cropped subset of the Sentinel-1 SAR scene used as a single training sample – was 1024 × 1024 pixels. After pretraining, the encoder weights from the online BYOL model were used to initialise a supervised segmentation model with an identical architecture. Decoder layers and output heads were reinitialised prior to supervised training, with weights set using Xavier uniform initialisation (Glorot & Bengio, 2010) and selected biases reset to zero. Xavier initialisation is a widely used method that scales the starting weights according to the number of input and output connections in a layer, ensuring that activations neither explode nor vanish as they pass through the network. This stabilises early training and improves convergence compared to arbitrary random initialisation.

During the first stage of fine-tuning, the encoder was frozen, meaning its weights were not updated, so that the pretrained representations learned from unlabelled data were preserved while the decoder adapted to the segmentation task. In the second stage, the encoder was gradually unfrozen, allowing its weights to be updated alongside the decoder. This two-step strategy stabilises training and prevents the encoder from overfitting to the limited labelled data too early. It allows the entire network to adapt jointly to the segmentation objective. In doing so, the self-supervised pretraining provides a foundation of generic spatial and textural features of sea ice and open water, which can then be aligned with class labels during fine-tuning. Because the encoder already encodes relevant structure from unlabelled data, effective segmentation can be achieved with substantially fewer labelled examples. The target network's weights are updated using an exponential moving average of the online network, ensuring training stability.

3.1.3 Random Forest (RF) Classifier

To assess a classical machine learning baseline, we implemented an RF classifier, which is a commonly used ensemble machine learning method. The RF model was trained on the same set of seven labelled scenes used in the UNet (SL) experiments, enabling a fair comparison. The feature set consisted of raw HH and HV backscatter values, the HH/HV polarisation ratio, Sobel-derived gradient magnitudes for each channel, global Shannon entropy, and GLCM texture measures (contrast and dissimilarity). These metrics are widely used to pretrain RF classifiers and other machine learning models for SAR-based sea ice classification (Harcourt et al., 2025; Zakhvatkina et al., 2019).

The RF model was trained using 200 trees in four stages (warm start of 50 trees per stage), with a maximum decision tree depth of 15. Warm start refers to incrementally adding new trees while retaining the previously trained ones, improving efficiency and allowing staged evaluation of performance. The model was trained using stratified

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

sampling, so that each subset of data preserves the overall class proportions (ice vs. open water), helping to mitigate class imbalance. However, the approach lacks spatial context beyond the immediate pixel, limiting its ability to detect elongated features such as leads.

3.1.4 Segment Anything Model (SAM)

We also tested version 1 of the SAM foundation model (Kirillov et al., 2023), a prompt-based segmentation framework pre-trained on a data set of 11 million natural RGB images and 1.1 billion masks. For compatibility with the HH-HV dual-channel input, we set HH as red, HV as green, and the ratio HH/HV as blue. Unprompted segmentation was evaluated in zero-shot mode. All SAM inferences were generated using the official Python implementation executed locally, rather than via the online interactive demonstration interface. This approach ensured full control over preprocessing, tiling, and prompt construction for SAR-specific inputs. No fine-tuning was applied.

3.2 Conditional Random Field (CRF) post-processing

To refine the predicted segmentation masks and suppress noise near the image boundary, a dense Conditional Random Field (CRF) was applied as a post-processing step (Fig. 4). In this framework, the UNet outputs act as the unary potentials (the model's initial per-pixel log-probabilities), representing the initial pixel-wise probabilities for each class. The CRF then introduces pairwise potentials that encourage label consistency across the image. Two Gaussian kernels are used: a spatial (smoothness) kernel, which promotes uniform labelling of pixels within local neighbourhoods, and a bilateral (appearance) kernel, which links pixels that are close in position and have similar predicted probabilities. These terms penalise unnecessary label changes between adjacent pixels, reducing noise and sharpening boundaries. These components iteratively adjust the label assignment to minimise the CRF energy function. The result is a refined segmentation that reduces noise in uniform areas, sharpens the delineation of floe edges and leads, and restores structural coherence to fragmented predictions. We employed the fully connected CRF formulation implemented with the pydensecrf library (Krähenbühl and Koltun, 2011). The CRF was applied to the sigmoid probability map produced by the UNet model prior to thresholding.

3.3 Loss function

To address severe class imbalance between open water and sea-ice pixels in the segmentation masks—typically around 90% ice and 10% open water in our labelled scenes— we employed a weighted Focal Loss (Lin et al., 2017) with asymmetric class weights, which has been shown to outperform cross-entropy in SAR-based sea-ice classification tasks with skewed class distributions (Vahedi et al., 2024). This formulation down-weights easy-to-

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

classify background pixels and emphasises learning on the more difficult minority class (water). The Focal Loss was applied only during the supervised training stage of the UNet models, including the Control, SL, and BYOL-fine-tuned variants. It was not used in the self-supervised (BYOL) pretraining phase, which relied solely on the cosine-similarity objective described in Sect. 3.1.2. The focal loss is defined as:

$$L_{Focal} = -\alpha(1 - p_t)^{\gamma} log(p_t) \quad (1)$$

where α is a class-balancing weight used to address class imbalance, p_t denotes the model-predicted probability assigned to the ground-truth class for a given pixel, and γ is a focusing parameter that controls the rate at which easy-to-classify examples are down-weighted. Higher values of γ increase the emphasis on hard-to-classify pixels by reducing the contribution of well-classified samples to the loss.Focal Loss was used across all supervised UNet training configurations, including models initialised from random weights and those fine-tuned after BYOL pretraining. This setup provided stable convergence and significantly improved segmentation accuracy in the MIZ and thin lead regions (see sect. 4), where backscatter signals are particularly ambiguous.

3.4 Experimental Design

We compared three categories of segmentation models – deep learning (UNet and BYOL-pretrained SSL UNet), a classical machine learning baseline (Random Forest), and a foundation model (SAM) – to evaluate their relative performance under limited annotated data conditions. From the nine manually labelled Sentinel-1 scenes, two were reserved as a fixed test set to evaluate all experiments. The remaining seven scenes were used to construct training sets, with two regimes: a 3-label and a 7-label condition. The test data set was strictly held out for final metric reporting. For supervised training, i.e. for UNet (Control), UNet (SL), UNet (SSL), samples consisted of a triplet: HH image, HV image, and binary segmentation mask. HH and HV channels were stacked into a 2-channel tensor, and masks served as ground truth. Samples were cropped to 1024 × 1024 pixels and normalised with z-score statistics computed per image. To allow direct comparison with the UNet (SSL) model, the supervised UNet (SL) was trained under both the 3-label and 7-label conditions described above. For the BYOL-pretrained UNet (SSL), a separate set of 12 unlabelled SAR scenes was utilised for self-supervised pretraining. The encoder was trained using BYOL to generate generalised SAR feature representations. For fine-tuning, the pretrained encoder was trained on the same 3-label conditions as the supervised UNet.

To ensure a fair comparison, all models were evaluated on the same held-out test scenes, and results are interpreted in terms of relative differences between models under identical data and evaluation conditions.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

3.5 Evaluation Metrics

To assess model performance in binary segmentation of sea ice versus open water, we use two primary metrics. The first of these is the Matthews Correlation Coefficient (MCC), which is a balanced metric that accounts for True Positives (TP) and False Positives (FP) (Chicco et al., 2021):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (2)$$

where TP denotes correctly predicted ice pixels, TN denotes correctly predicted water pixels, FP represents water pixels incorrectly classified as ice, and FN represents ice pixels incorrectly classified as water. MCC values range from −1 (total disagreement) to +1 (perfect prediction), with 0 indicating random performance.

We also used the F1 Score, which is the harmonic mean of precision and recall, indicating the model's ability to correctly identify ice pixels without overpredicting:

$$F1 = \frac{2 \times TP}{2\,TP + FP + FN} \quad (3)$$

It is sensitive to both FPs and FNs. Both metrics are computed on a held-out fixed test set of 2 scenes. These scenes were held out from all training to provide an unbiased assessment of generalisation across contrasting ice regimes. Because class balance shifts markedly between scenes and spatial coherence is central to utility, MCC provides the more reliable summary of performance in our setting.

4. Results

4.1 Model performance: Consolidated ice pack (Scene 1)

In the first test case, we evaluated the performance of all models on a scene depicting a consolidated ice pack with well-defined ice floes and linear leads (Scene 1; Fig. 7). The performance of each model applied to Scene 1 is summarised in Table 2. UNet (SSL) outperformed both UNet Control and UNet SL in both the F1 Score and MCC metric. As shown in Fig. 7d, the UNet (SSL) model achieved clear delineation of ice leads and more spatially coherent segmentation across the scene. By contrast, the UNet (Control) model (Fig. 7c) produced increasingly noisy predictions toward the right of the image and failed to capture fine-scale features. The fully supervised UNet (SL), shown in Fig. 7e showed improved definition of lead boundaries but still misclassified some ice regions as

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

water. These qualitative differences align with quantitative metrics: UNet (SSL) reached an F1 Score of 0.9802 and MCC of 0.4389 – higher than UNet (SL) (MCC = 0.2525) and far superior to the Control (MCC = 0.1338).

Table 2. Segmentation Results (F1 Score and MCC) – Scene 1 (Consolidated Ice Pack) (Bold denotes best performing model)

Model	F1 Score	MCC
UNet (Control, 3 scenes)	0.7747	0.1338
UNet (SL, 7 scenes)	0.9283	0.2525
UNet (SSL, 3 scenes)	**0.9802**	**0.4389**
Random Forest	0.9767	0.0352
SAM (Zero-shot)	0.9827	0.4137

We also evaluated two non-UNet baselines: SAM and a classical Random Forest classifier. Despite being trained on natural RGB data, SAM performed surprisingly well in Scene 1 (F1 = 0.9827, MCC = 0.4137; Fig. 7g), capturing the broad ice-water distribution but failing to detect fine-scale leads. In the consolidated ice pack, SAM successfully recovered large-scale ice structure and lead geometry but systematically overfilled leads and open water, producing approximately 71.5% false positives compared to BYOL's approximately 57.1% (Appendix A). The Random Forest model performed poorly in both scenes (e.g. MCC = 0.0352 in Scene 1; Fig. 7f). It significantly overpredicted ice, achieving inflated F1 Scores (up to 0.9834) but producing coarse, spatially incoherent segmentations. The F1 Score is limited in imbalanced settings, especially when high recall is achieved at the cost of accuracy. Across all models evaluated, the Random Forest was the weakest performer by a substantial margin, offering little discriminatory skill in either scattering regime.

4.2 Model performance: Marginal Ice Zone (MIZ) (Scene 2)

Scene 2 (Fig. 8) captures a fragmented MIZ scene with scattered ice floes and diffuse boundaries, which introduces greater textural ambiguity. We find similar trends to those in Scene 1. The UNet (SSL) model (Fig. 8d) was more effective at correctly identifying sea ice, particularly in fragmented and ambiguous regions, than the UNet (Control) and fully supervised UNet (SL) (Fig. 8e), which often misclassified these regions as open water. This resulted in significantly fewer false negatives and a more complete representation of ice extent in the MIZ. This is reflected in

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

the higher MCC achieved by the UNet (SSL) model (0.2087), compared to 0.1685 for the UNet (SL) and 0.1449 for the UNet (Control), supporting its ability to generalise to structurally varied sea ice regimes.

Table 3. Segmentation Results (F1 Score and MCC) – Scene 2 (Marginal Ice Zone) Bold denotes best performing model

Model	F1 Score	MCC
UNet (Control, 3 scenes)	0.5981	0.1449
UNet (SL, 7 scenes)	0.6812	0.1685
UNet (SSL, 3 scenes)	0.8449	**0.2087**
Random Forest	0.9834	0.0201
SAM (Zero-shot)	0.9727	0.0403

While the Random Forest model achieved the highest F1 Score in Scene 2 (Table 3), its segmentation output was visually poor and spatially incoherent. The confusion matrix (Appendix B) shows that 842,116 water pixels (99.82%) were classified as ice, with only 1,538 water pixels (0.18%) correctly classified, while 25,088,922 ice pixels (99.984%) were correctly classified as ice. This explains its inflated F1 Score—high recall but very low precision—and highlights how F1 can be misleading in imbalanced datasets where a model overpredicts the dominant category. The much lower MCC (0.0201) reflects this imbalance more correctly, reaffirming that qualitative assessment and multiple metrics are essential in evaluating segmentation quality. SAM produced more structured and visually accurate segmentations than the Random Forest, particularly in the simpler scene. In this more fragmented setting it misclassified ~94% of water pixels as ice, leaving only ~6% correctly identified as open water. This caused MCC to collapse to 0.0403, despite a superficially high F1 score (0.9727). UNet (SSL) again demonstrated better balance, with approximately 20.8% FPs and approximately 79.2% TNs, achieving a higher MCC of 0.2087 and preserving boundary integrity across mixed floes and textured water. These results indicate that while SAM's large-scale RGB pretraining transfers structural priors such as edge continuity and spatial coherence, these features degrade under radar-specific noise and ambiguous backscatter.

4.3 Cross-comparison

Across all models, performance metrics were lower in Scene 2 than in Scene 1. Mean MCC values fell by approximately 60–70% between the two scenes (e.g., UNet (SSL) declined from 0.4389 in Scene 1 to 0.2087 in Scene 2; UNet (SL) from 0.2525 to 0.1685; SAM from 0.4137 to 0.0403), and the supervised UNets experienced F1 Score reductions of 20–40%. This reflects the increased difficulty of segmenting fragmented floes and

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

ambiguous SAR textures in the MIZ. Performance differences across ice regimes align with known physical scattering characteristics. In consolidated pack ice (Fig. 7), the smooth backscatter of leads and the high contrast between ice and water favour models with strong structural priors (e.g., SAM), whereas in the MIZ (Fig. 8) the diffuse, mixed-pixel textures require models that capture fine-grained statistical variation, consistent with observed seasonal changes in floe size distributions and fragmentation in Arctic marginal ice zones (Buckley et al., 2024). In practice, speckle and diffuse SAR boundaries trigger these priors to "close" narrow water features, widening ice regions and eroding lead continuity. These results show that SAM relies on smooth object boundaries and continuity priors learned from RGB data, which encourage gap filling at the expense of local texture fidelity. BYOL, in contrast, maintained sharper water–ice delineation and reduced over-segmentation while preserving overall coverage. The improved robustness of UNet (SSL) across these conditions suggests that its pretraining captured structural invariants in radar backscatter – such as floe continuity and speckle-stable texture – allowing it to generalise across both MYI- and FYI-dominated environments. In particular, BYOL's pretraining on unlabelled SAR scenes allows it to internalise textural cues. These results demonstrate that scene-dependent radar backscatter, rather than label scarcity alone, underpins the observed performance hierarchy across models.

Beyond overall accuracy gains, BYOL pretraining for UNet (SSL) reduces FNs – missed detections of sea ice pixels – relative to the supervised UNet (SL) models. In the MIZ (Scene 2), the BYOL-pretrained UNet (SSL) model produced 6.6 million FNs, compared with 12.1 million for the fully supervised UNet (SL) model and 14.4 million for the UNet (Control), representing a reduction of roughly 45% relative to the latter. These values are summarised in the confusion matrices provided in Appendix A. This improvement is visually evident along the boundaries of fragmented floes and thin newly formed ice (Fig. 8), where the supervised UNet (SL) often misclassified low-backscatter regions as open water. UNet (SSL), by contrast, retained these marginal ice features as continuous structures, capturing narrow leads, diffuse floe edges, and partially consolidated ice that the supervised models failed to detect. Similar trends were observed in the consolidated ice pack, where BYOL recovered faint linear leads overlooked by both supervised networks. These reductions in FNs suggest that self-supervised pretraining can mitigate some class-specific weaknesses in SAR imagery of sea ice. This improvement corresponds to an increase in recall from 0.517 for the UNet (SL) to 0.737 for the Unet (SSL) in the MIZ, confirming its enhanced ability to capture ambiguous or marginal ice.

The improved performance of UNet (SSL) compared to the other models demonstrates that self-supervised pretraining enables the model to learn generalisable spatial and textural representations from unlabelled SAR imagery, making it significantly more label-efficient. UNet (SSL) using BYOL, trained on just three labelled images, matched or exceeded the performance of a fully supervised model trained on more than twice as many labels. This provides evidence that UNet (SSL) using BYOL can be effective in low-label regimes. BYOL's cross-

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

view invariance curbs FPs in textured water and preserves boundary structure, whereas SAM's RGB-trained smoothness priors overfill ice in speckled scenes and RF's pixel-wise decision rule misses spatial context. Most FNs in both UNet (SSL) and UNet (SL) outputs occur in low σ^0 regions, typically associated with thin or melting ice. Conversely, false positives cluster in bright, rough-water patches where wind or incidence angle effects elevate HH returns.

4.4 Model performance across ice types and HH backscatter ranges

Fig. 9 shows model performance (MCC) as a function of HH backscatter (σ^0), computed by pooling 64 × 64 tiles within 0.5 dB HH bins. This links segmentation accuracy directly to physical scattering regimes and, by extension, to dominant ice types within each σ^0 interval. In the consolidated ice pack (Scene 1), all models exhibited clear σ^0-dependent behaviour. SAM performed best in moderately bright regimes (≈ −15 to −14 dB) where floe surfaces were smooth and boundaries well defined. UNet (SSL) maintained consistently high MCC across a broader interval (≈ −16 to −12 dB), reflecting robustness to speckle and subtle textural variation. At higher backscatter values, additional trends emerge. Both UNet (Control) and the Random Forest show a marked increase in MCC between approximately −12 dB and −9 dB, corresponding to very bright scattering regimes typically associated with rough, consolidated ice. Beyond −9 dB, however, performance declines across all models. This suggests that once backscatter becomes uniformly high, additional discriminatory information is limited, reducing the effectiveness of both learned and hand-crafted features. Conversely, all models exhibit a local increase in performance near the ice–water transition, where strong contrast in σ^0 provides a clear separability signal. This behaviour is expected given the dominance of the ice class and the relative ease of identifying ice in high-contrast regimes, reinforcing that gains at the transition are driven primarily by radiometric contrast rather than nuanced structural understanding.

In the MIZ (Scene 2), where scattering from water, thin ice, and deformed floes overlap, the contrasts between models widened. UNet (SSL) outperformed other models between -22 dB and -16 dB with a positive MCC. In this region all other models approached random performance (i.e. MCC was around 0) whilst UNet (SSL) performed consistently better (MCC ~0.2). This reflects the UNet (SSL) encoder's ability to recognise structural coherence and textural relationships that are not tied to radar backscatter alone. SAM's transferability collapsed above -18 dB. Notably, both UNet (Control) and UNet (SL) showed a marked increase in accuracy from −17 dB to around −11 dB, indicating that bright, high-contrast ice was easier for supervised models to classify, particularly when strong training examples were available. UNet (SSL), by contrast, peaked earlier and declined in the high radar backscatter regions, suggesting that supervised learning benefits more directly from high-contrast labels, while SSL is more effective in low-contrast regimes.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Overall, the σ^0-binned results reveal distinct performance trends across scattering regimes:

- UNet (Control) and UNet (SL) excel in bright, high-contrast scattering regimes, typical of rough multi-year ice or strong ice–water boundaries.
- UNet (SSL) excels in low to moderate σ^0 ranges, where scattering is more ambiguous and structural cues rather than brightness determine class separability.

This complementarity suggests that future hybrid approaches—e.g., supervised fine-tuning that explicitly leverages SSL-derived representations in high-σ^0 regimes—may further improve performance, particularly in operational settings that require robustness across the full range of Arctic scattering conditions.

5. Discussion

5.1 Sea ice segmentation with fewer labels

This study demonstrates that self-supervised learning can significantly reduce the reliance on annotated data for accurate SAR-based sea ice segmentation. Among the three UNet variants tested—the baseline UNet (Control), the fully supervised UNet (SL), and the BYOL-pretrained UNet (SSL)— the BYOL-pretrained encoders i.e. UNet (SSL) consistently produced the strongest overall segmentation performance, surpassing the fully supervised UNet (SL), the RF baseline, and SAM. Absolute metric values should therefore be interpreted in the context of this controlled comparison, as the primary contribution of the study lies in demonstrating relative performance gains under reduced annotation budgets, rather than maximising accuracy on a specific benchmark. However, individual models still displayed local advantages under certain radiometric conditions. For example, the fully supervised UNet (SL) performed best in the high σ^0 regions where the contrast between ice and water is strongest. Furthermore, the improved performance of the BYOL-pretrained UNet (SSL) under limited-label conditions can be explained by its ability to extract generalisable structural and textural features from unlabelled SAR imagery. Unlike the UNet (Control) and UNet (SL) models, which must learn discriminative features entirely from a small labelled dataset, the BYOL encoder benefits from pretraining across a wider range of SAR patterns. This enables it to capture subtle variations in backscatter associated with sea ice characteristics, such as the roughness differences between thin ice and open water or the texture of fragmented floes, even when only a few labelled examples are available for fine-tuning. BYOL's non-contrastive design is also advantageous for SAR imagery, where negative-pair assumptions break down because visual dissimilarity does not reliably indicate class differences, and overlapping σ^0 ranges can make distinct ice types appear similar (Singha et al., 2018). These representational advantages of SSL are reflected in the higher MCC of the BYOL model, particularly in the marginal ice zone (MIZ), where scattering is more variable. By encoding a broader feature space during pretraining, the model is

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

better able to recognise marginal or low-contrast ice types that the supervised UNet (SL) misses, reducing FNs without substantially compromising precision.

Our findings are consistent with a growing body of work demonstrating the value of self-supervised learning for satellite image analysis. Ayush et al. (2021) showed that SSL can close almost the entire performance gap with fully supervised models for land-use classification on the Functional Map of the World dataset, even surpassing the supervised baseline after fine-tuning. Jain et al. (2022) introduced RS-BYOL, a remote-sensing adaptation of BYOL using teacher–student distillation and reported that it outperformed models pretrained on benchmark data sets such as ImageNet (Deng et al., 2009) across multispectral and SAR tasks. Complementing these results, Muzeau et al. (2024) demonstrated that SSL can extract stable structural features from SAR using masked Siamese Vision Transformers, while Liu et al. (2024) showed that contrastive learning improves SAR scene classification under limited labels. Together, these studies highlight the broader potential of SSL for Earth Observation across optical, multisensor, and SAR modalities. Our results extend this evidence to the Arctic cryosphere by demonstrating that non-contrastive pretraining also benefits dense, pixel-level segmentation of sea ice—an inherently more challenging task than patch-level classification due to speckle, mixed pixels, diffuse boundaries, and overlapping σ^0 ranges. BYOL's consistent performance across both consolidated and fragmented ice regimes indicates that it's learned representations capture structural patterns that generalise across scattering conditions, providing robustness that supervised models struggle to achieve with limited annotated data.

5.2 Improvement on baseline models

The Segment Anything Model (SAM), despite being trained exclusively on natural RGB imagery, achieved surprisingly high performance in the consolidated ice pack, but this apparent strength in structured scenes did not generalise to the MIZ. This scene-dependent performance is consistent with the findings of Shankar et al. (2024), who observed similar behaviour when testing SAM across glaciological features and remote sensing modalities. They reported that while SAM performed well on clear, object-like features such as glacier termini and icebergs in open water, its accuracy declined markedly in mélange, crevassed terrain, and noisy SAR imagery—particularly without prompt optimisation or domain-specific tuning. Wallace et al. (2025) reached a similar conclusion for centimetre-resolution UAV imagery of glacier crevasses, showing that off-the-shelf SAM and SAM 2 models provide only moderate segmentation skill and argued that domain shift and complex ice-fracture patterns require targeted fine-tuning or few-shot adaptation before such models can be relied on operationally. Their results and ours highlight that SAM's transferability to cryospheric imagery depends critically on scene structure, whereas

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

BYOL's self-supervised pretraining yields representations inherently adapted to the statistical properties of radar backscatter.

In contrast, the Random Forest (RF) model performed poorly in both test scenes, with MCC values reaching as low as 0.0352 in the consolidated ice pack case (Scene 1; Table 2 and Fig. 7). Although the model was trained on a range of SAR-relevant features, including polarisation ratios, edge filters, and texture measures, it was unable to learn spatial or contextual dependencies beyond the immediate pixel neighbourhood. This limitation is well documented in the remote sensing literature. Sinha et al. (2019) demonstrated that RF-based models are highly sensitive to spatial autocorrelation in training data and tend to regress towards the mean in the absence of strong global spatial patterns. Their findings showed that RF models often underestimate extremes and fail to preserve boundary integrity when applied to geospatial prediction tasks, particularly when the spatial structure of the input data varies across scales. This behaviour is consistent with our results, where the Random Forest produced over-smoothed boundaries and in-filled open water, reflecting the same tendency to lose fine-scale spatial structure under variable scattering conditions. These results show the limitations of classical machine learning methods for dense prediction tasks on spatially complex, noisy data such as SAR imagery (Sinha et al., 2019). In addition, our RF predictions over-smooth floe edges and infill water between closely spaced floes, inflating recall and depressing precision; the result is high F1, driven by class dominance, but very low MCC once FPs and TNs are accounted for. The Random Forest's inflated F1 Score, despite its poor visual performance, underscores the limitations of single-metric evaluation in imbalanced segmentation tasks. Its high recall and class dominance artificially boost F1, yet its MCC remains low, and its predictions lack spatial coherence. This result reinforces the need to combine visual inspection, spatial metrics, and balanced indicators like MCC in future remote sensing evaluations.

5.3 Comparison to other self-supervised approaches

Alternative contrastive self-supervised methods such as SimCLR (Chen et al., 2020) and MoCo (He et al., 2020) learn by maximising agreement between augmented views of the same image (positive pairs) while pushing apart views from different images (negative pairs). SimCLR relies on very large batch sizes to ensure many negatives are available at once, whereas MoCo uses a dynamic memory queue to provide a larger, more consistent set of negatives across batches. However, these approaches are challenging to scale to remote-sensing imagery, where large spatial dimensions (typically 512–1024 pixels rather than the 224 pixels used in ImageNet) and multi-channel inputs (e.g. HH, HV, VV, or optical stacks) inflate memory requirements and constrain batch size. This limitation is well recognised in the literature. For example, Alosaimi et al. (2023) showed that high-resolution remote-sensing scenes make contrastive training inefficient under few-shot regimes, and Piao et al. (2023) identified GPU memory

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

as a fundamental bottleneck for large-batch contrastive learning. Consequently, frameworks like SimCLR, which depend on large negative pools, are less suitable for SAR, whereas non-contrastive approaches such as BYOL are naturally more efficient and stable under limited-memory conditions.

In the context of remote sensing, Jain et al. (2022) demonstrated that a BYOL-based distillation model (RS-BYOL), trained on VV–VH polarised SAR and multispectral imagery, outperformed ImageNet-pretrained baselines on land-cover classification and segmentation tasks. Their results confirm that non-contrastive self-supervised methods can learn invariant features from remote-sensing data without labelled supervision. Our findings extend this principle to Arctic sea ice imagery, showing that similar invariance mechanisms improve segmentation where brightness overlap and texture ambiguity are common. Whereas Jain et al. (2022) used cross-modal information to enrich their representations, the present work demonstrates that comparable robustness can be achieved through SAR-only pretraining when paired with physically consistent augmentations and dual-polarised inputs. Recent theoretical work further supports this interpretation. Garrido et al. (2023) showed that contrastive and non-contrastive frameworks share similar underlying mathematical foundations, suggesting that the specific loss formulation—that is, the mathematical objective used to measure and minimise the difference between representations—may be less critical than the overall network architecture and training dynamics. Furthermore, Nguyen et al. (2023) proposed Dimensional Contrastive Learning (DimCL), which enhances feature diversity within learned embeddings; when applied to BYOL, it improved representation robustness under limited supervision. These findings align with our results, where feature diversity learned through augmentation and spatial invariance translated directly into stronger segmentation consistency across distinct ice regimes.

5.4 Implications for sea ice monitoring

Our results show that comparable sea ice segmentation performance can be achieved using only three labelled Sentinel-1 scenes, offering a scalable route toward automated sea ice monitoring. Self-supervised pretraining enables models to learn directly from the vast archives of unlabelled SAR imagery collected by satellites such as Sentinel-1 ERS-1/2, RADARSAT and more, reducing dependence on manual annotation and improving temporal coverage. Compared with traditional supervised approaches (Park et al., 2020; Khaleghian et al., 2021; Huang et al., 2024), this allows encoders to be trained once on generic SAR patterns and then adapted to tasks such as lead detection, ice-edge tracking, or floe-size mapping with minimal supervision. The demonstrated advances in label-efficient segmentation and self-supervised SAR pretraining could aid operational ice services and forecasting centres by providing more consistent, high-resolution inputs for data assimilation and climate analysis. Maritime users would also benefit by receiving finer delineation of leads and detailed mapping of marginal ice zones, which can support route planning, optimise navigation windows, and reduce risk in hazardous ice conditions.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Although the findings are robust across both test scenes, they reflect the scope of the present data set and experimental design. First, the pretraining dataset is modest and geographically concentrated in the CAA, potentially biasing the learned representations toward regional scattering regimes and incidence-angle distributions. Second, we treat segmentation as binary (ice vs. water), applying a conservative MIZ rule and excluding sub-pixel and very narrow leads (<~200 m at 80 m resolution); this simplifies evaluation but suppresses ice-type variability that is operationally relevant. Third, the 80 m ground-range resolution of the Sentinel-1 EW product limits the retrieval of fine-scale ice structures such as narrow leads, brash ice, and melt ponds. These features are often smaller than a pixel, meaning that backscatter values represent mixed scattering from both ice and water. Such mixed pixels introduce ambiguity and reduce the apparent sharpness of boundaries, which likely constrains maximum segmentation accuracy even for well-trained models.

Future work should look to extend the framework developed in this paper beyond binary segmentation to multi-class ice mapping, distinguishing between thin ice, first-year ice, multi-year ice, melt ponds, and open water. Furthermore, expanding SSL pretraining to multi-year, multi-sensor archives and incorporating temporal coherence would move this approach closer to near-real-time, large-scale Arctic sea ice monitoring – linking research-grade segmentation to practical, operational decision-making. Importantly, modern UNet-based frameworks, such as the one used here, can process complete Sentinel-1 scenes in only a few seconds once trained, meaning that improving label efficiency upstream directly translates into faster, more consistent downstream mapping over vast Arctic regions. Integrating such SSL-based segmentation within data-driven forecasting systems (e.g. ICENet; Andersson et al., 2021) or as enhanced observational inputs for physical sea-ice and coupled ocean–ice models would bridge representation learning and process understanding, including data-driven surrogate models of sea-ice state variables such as thickness (Durand et al., 2024), supporting both scientific analysis and operational Arctic sea ice monitoring. UNet-based segmentation produces spatially complete, gridded ice–water fields from SAR imagery that can be used directly as sea ice concentration or ice-state inputs in forecasting systems, meaning that improved segmentation quality translates into more accurate initial ice fields for prediction, consistent with evidence that assimilating satellite-derived sea ice concentration improves seasonal forecast skill (Zhang et al., 2021)

6. Conclusion

This study demonstrates that self-supervised learning, specifically Bootstrap Your Own Latent (BYOL), can substantially reduce reliance on labelled data for sea-ice segmentation from Sentinel-1 SAR imagery. A UNet encoder pretrained with BYOL achieved superior performance to fully supervised and classical approaches while using less than half the annotated data. In a consolidated ice pack scene, the BYOL-pretrained model reached an F1

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

score of 0.9802 and MCC of 0.4389, exceeding both the fully supervised UNet trained on seven labelled scenes (F1 = 0.9283, MCC = 0.2525) and the control model trained on three (F1 = 0.7747, MCC = 0.1338). In the more challenging marginal ice zone, BYOL again performed best, achieving F1 = 0.8449 and MCC = 0.2087, compared with 0.6812/0.1685 for the supervised UNet and 0.5981/0.1449 for the control. These results confirm that BYOL pretraining substantially enhances segmentation accuracy and generalisation under limited-label conditions, outperforming both conventional deep learning and classical machine-learning models such as Random Forest (MCC ≤ 0.035) and the SAM foundation model in the marginal-ice scene (MCC ≈ 0.040). In contrast, SAM performed strongly in the consolidated-ice scene (MCC ≈ 0.414) but failed to generalise to more complex radar environments. These results show that self-supervised representation learning can perform well even with limited annotated data.

By pretraining directly on SAR imagery without negative pairs, the BYOL framework mitigates many of the limitations of contrastive methods for radar data, such as the difficulty of defining semantically meaningful dissimilarities. The performance differences observed across scenes can be interpreted in terms of the physical properties of radar backscatter in sea ice. In the consolidated ice pack, strong σ^0 contrast between bright, deformed MYI and the darker open-water background provides clear discriminatory cues, enabling all deep-learning models—including SAM—to perform relatively well. By contrast, the marginal ice zone contains thin ice, slush, brash, and small floes whose σ^0 values frequently overlap with rough water, producing ambiguous or low-contrast signatures. These mixed scattering mechanisms reduce brightness separability and create diffuse boundaries, leading the supervised UNet and SAM to fragment continuous ice or misclassify rough water as thin ice. The BYOL-pretrained encoder, however, leverages structural invariants learned from unlabelled SAR scenes—such as floe continuity, lead geometry, and speckle-stable texture—to maintain segmentation coherence even when intensity-based cues are unreliable. In practical terms, this allowed BYOL to correctly identify thin, low-backscatter ice and preserve floe boundaries in regions where σ^0 differences alone were insufficient for supervised or classical methods. The use of SAR-specific augmentations and CRF post-processing further improved spatial coherence, helping to recover fine-scale ice features while suppressing speckle and noise.

Overall, this work contributes to the growing body of label-efficient deep learning applied in remote sensing and demonstrates that self-supervised pretraining offers a practical, scalable solution to the annotation bottleneck in Arctic sea ice monitoring. While promising, our approach remains limited by the size and diversity of the pretraining data set, and future work should explore scaling BYOL to larger, geographically and temporally varied SAR archives. Extension to multi-class ice type segmentation and the incorporation of domain adaptation strategies may further enhance generalisation to new sensors and conditions. As SAR data volumes continue to grow,

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

approaches like BYOL hold considerable promise for improving the accessibility and robustness of sea ice mapping in support of climate research and safe navigation in polar regions.

Figures

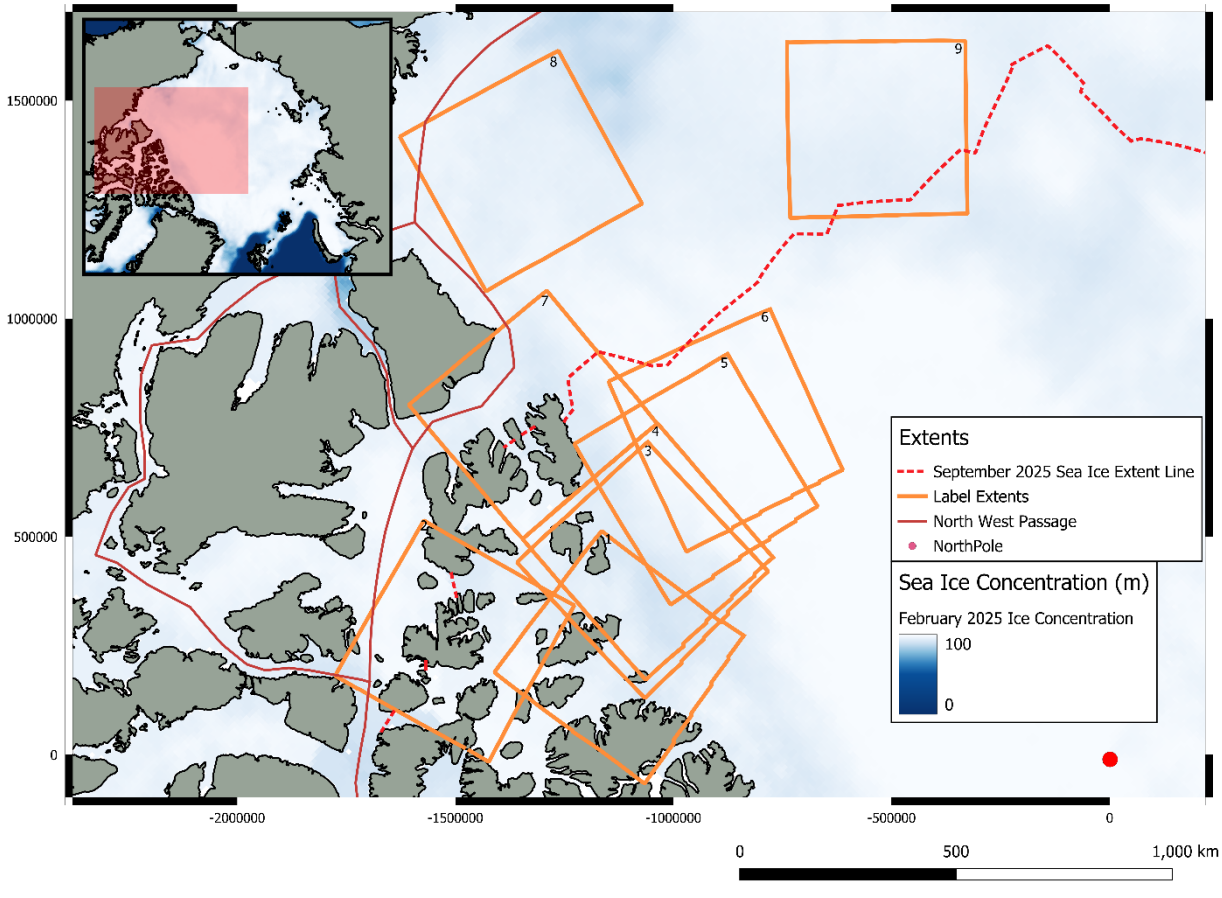

Figure 1. Spatial distribution of labelled Sentinel-1 SAR scenes used in this study. The main panel shows the extent of manually labelled segmentation overlays across the Canadian Arctic Archipelago (CAA) and adjacent Arctic Ocean. The background represents daily sea-ice concentration on 1 February 2025 from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI-SAF) Northern Hemisphere product (OSI SAF, 2025). The inset map provides Arctic-wide context, showing the location of the main panel as a red footprint outline.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

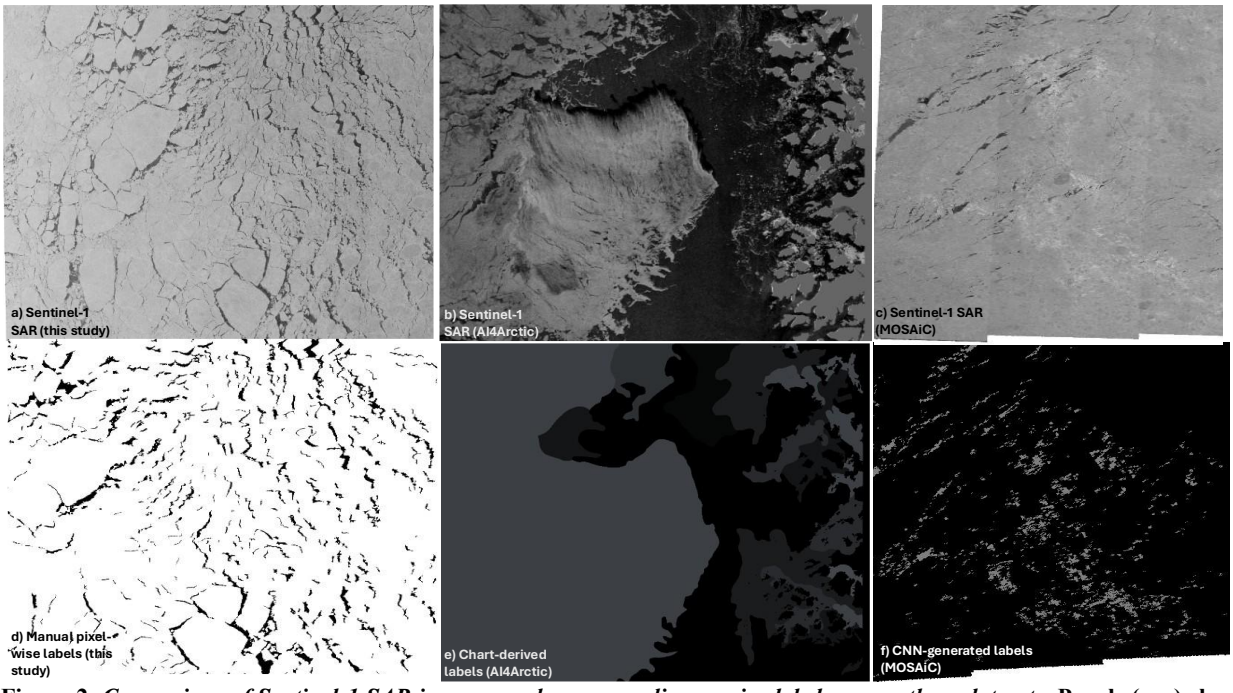

Figure 2. ***Comparison of Sentinel-1 SAR imagery and corresponding sea-ice labels across three datasets.*** **Panels (a–c) show Sentinel-1 SAR backscatter for (a) this study, (b) the AI4Arctic Sea Ice Challenge dataset, and (c) the MOSAiC-based dataset. Panels (d–f) show the corresponding labels for (d) this study, (e) AI4Arctic, and (f) MOSAiC. Labels in this study (d) are manually annotated at pixel level for a binary ice–water classification. AI4Arctic labels (e) are derived from operational ice charts and exhibit spatial generalisation consistent with regional mapping objectives, while the MOSAiC labels (f) are generated using a convolutional neural network and focus on lead structures within the central Arctic Ocean. The figure highlights differences in label provenance and spatial granularity that reflect the distinct objectives of each dataset and motivate the use of manually annotated pixel-wise labels for controlled model comparison under limited-label conditions.**

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

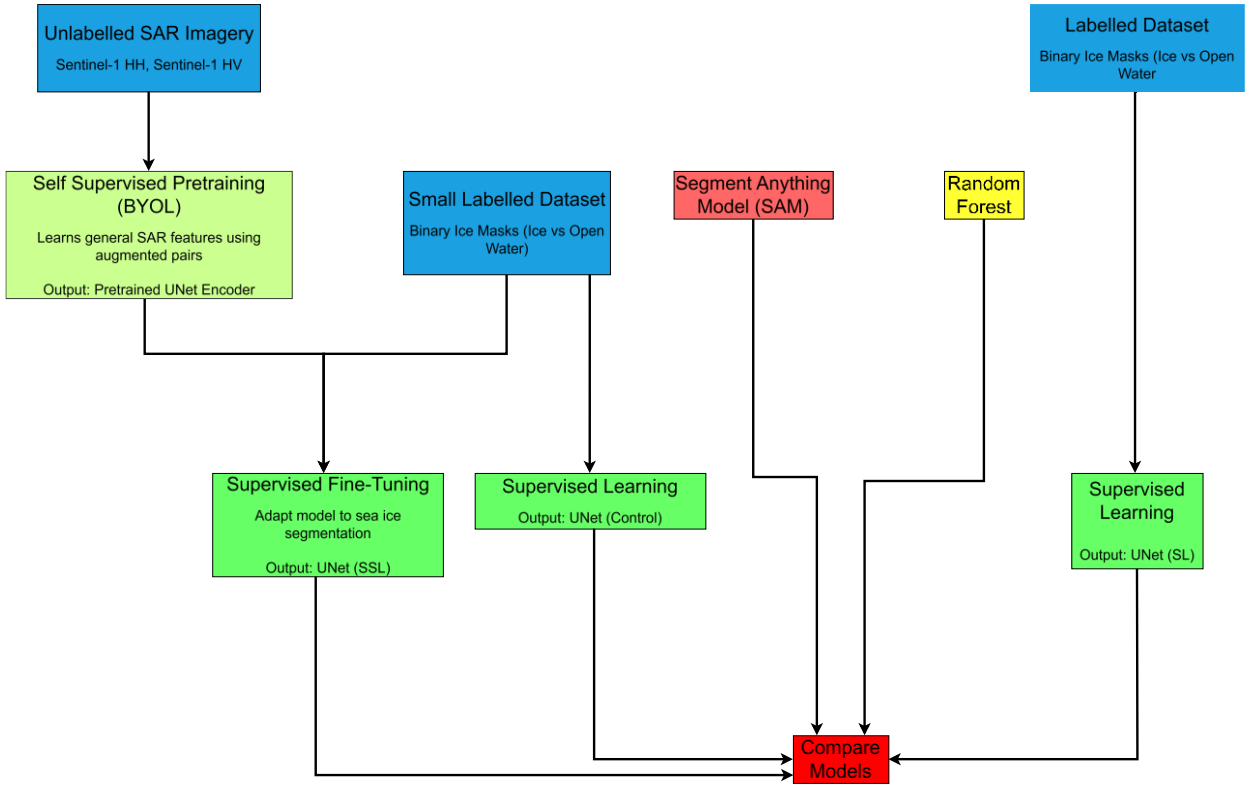

Figure 3. Overview of the Model Comparisons.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

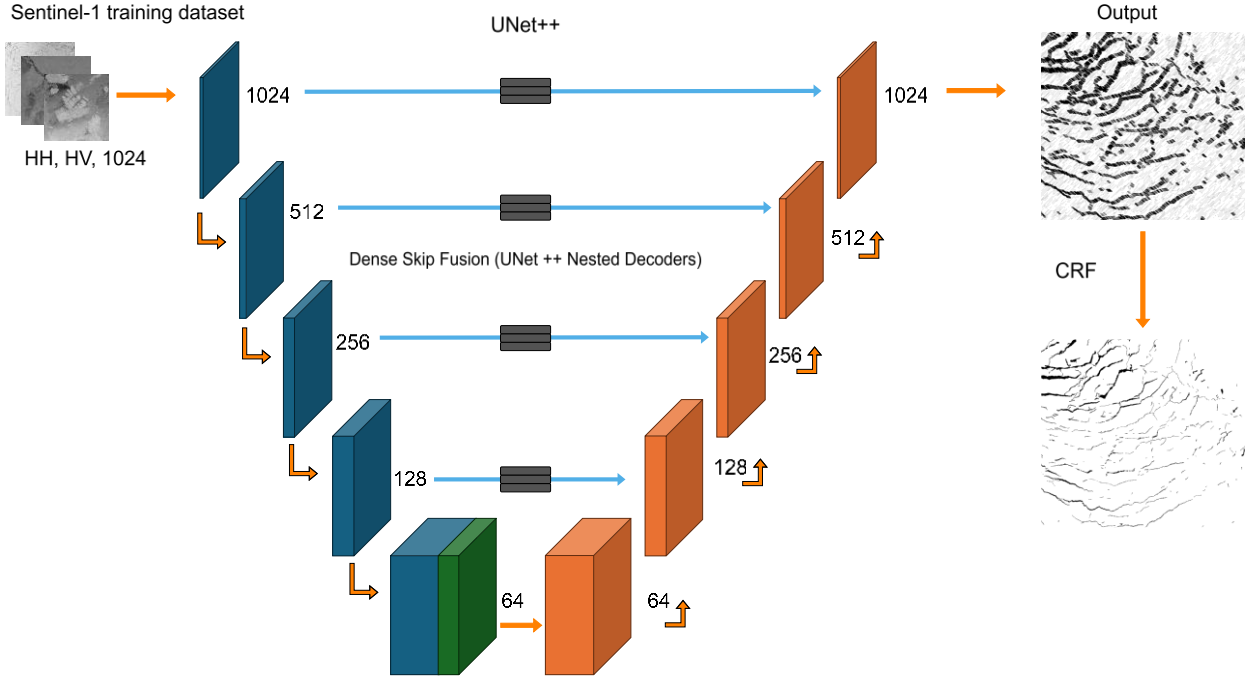

Figure 4. Simplified schematic of the modified UNet++ architecture used for SAR sea-ice segmentation. Blue boxes represent encoder blocks that progressively reduce spatial resolution, orange boxes represent decoder blocks. Blue arrows indicate skip connections, orange arrows denote the main feed-forward path through convolution and upsampling operations. The bottleneck (64 × 64) includes a Channel Attention Module (green box). Numbers indicate spatial resolution at each stage. Intermediate nested decoder layers of the UNet++ structure are omitted for clarity. Grey boxes denote the locations of intermediate nested decoder pathways in the full UNet++ architecture, which are not shown explicitly for clarity.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

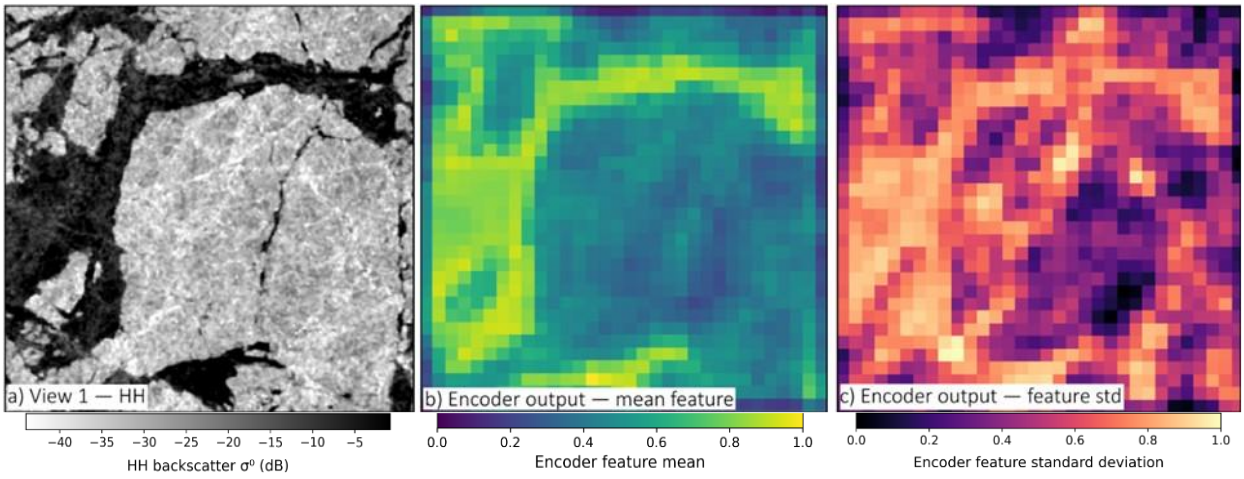

Figure 5. Example BYOL encoder feature visualisation. (a) An augmented HH input with (b) the corresponding encoder feature mean, and (c) standard deviation from the deepest UNet encoder layer (64 × 64). This 1024 × 1024 patch represents approximately 82 × 82 km of the sea-ice surface.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

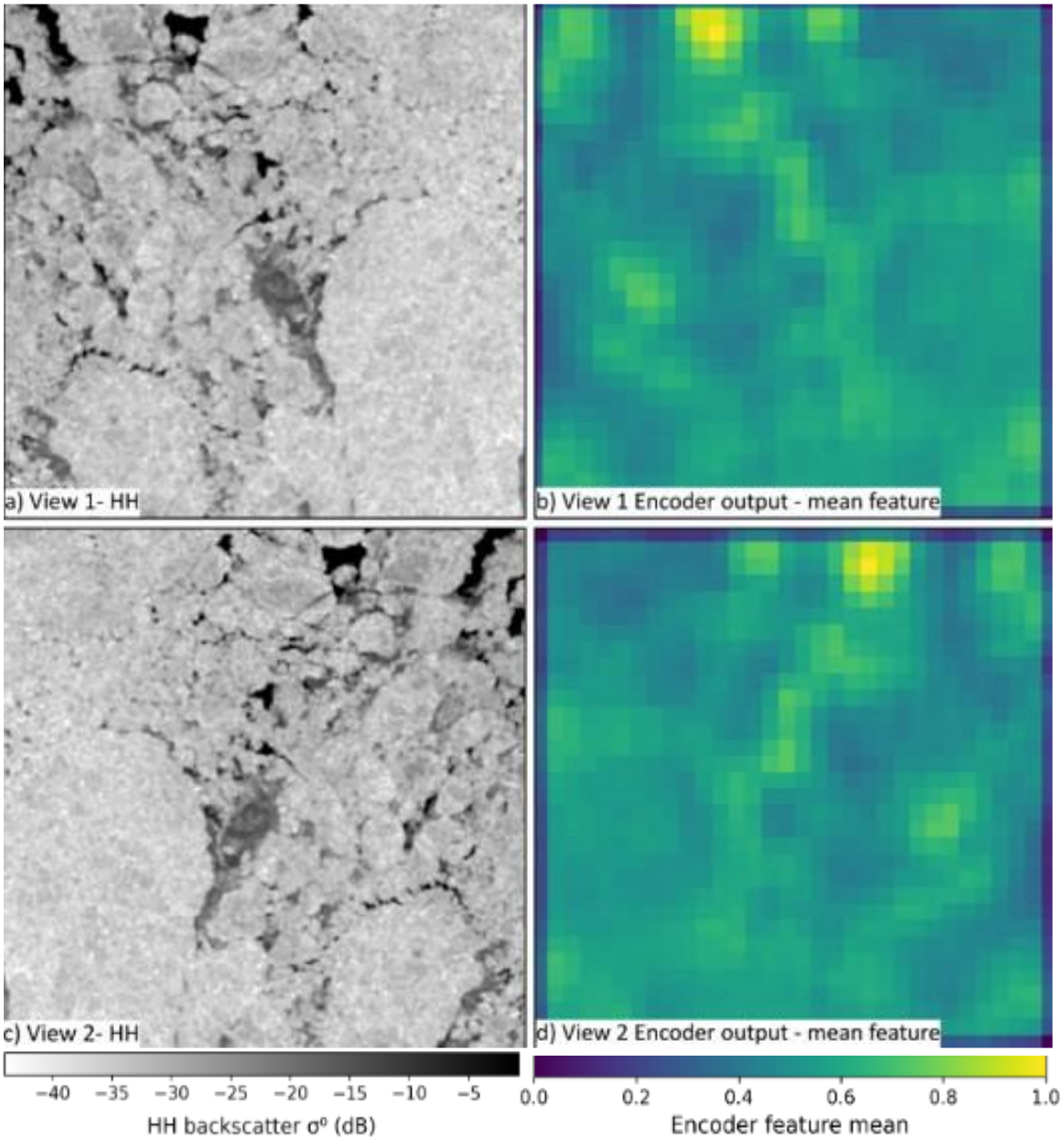

Figure 6. Example of BYOL encoder feature consistency across augmented views. (a) Augmented View 1 (HH channel) and (b) its corresponding encoder mean-feature map. (c) Augmented View 2 (HH channel) and (d) its corresponding encoder mean-feature map. Higher mean values in (b) and (d) indicate stable, high-activation features shared across both views, showing that the encoder has learnt invariant spatial structures such as floe continuity and lead geometry despite differing augmentations.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

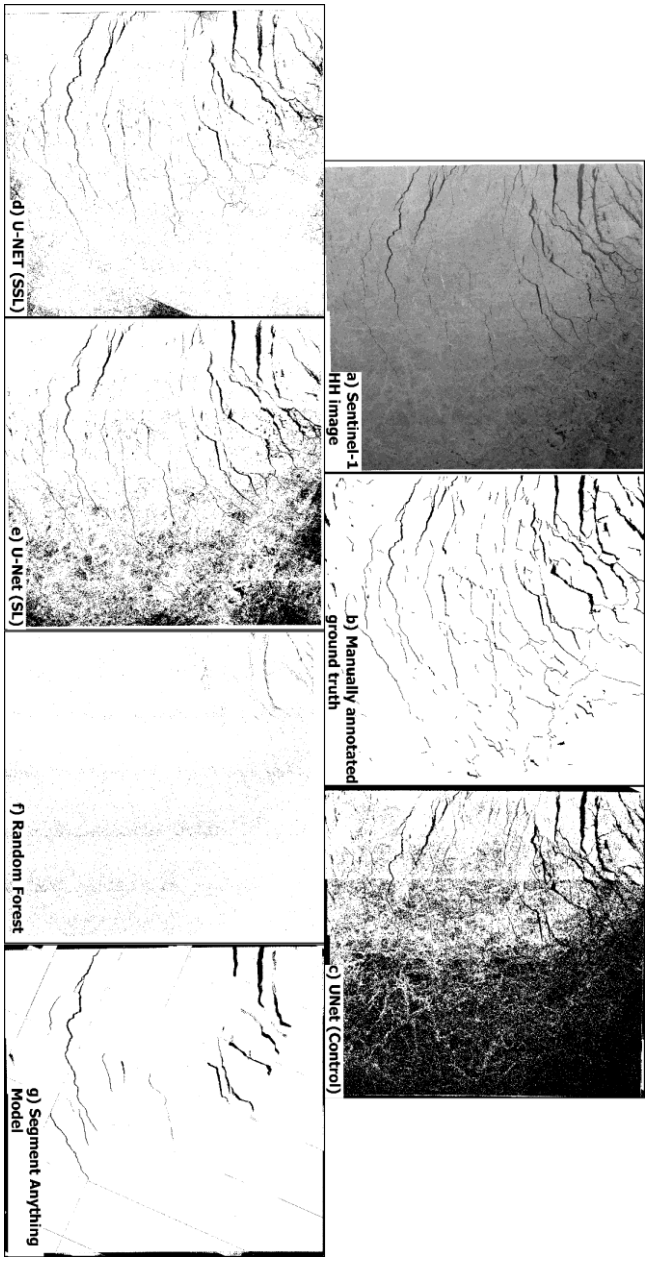

Figure 7. Scene 1 (Consolidated Ice Pack). Comparison of ice-water segmentation across models for the test scene. Panels show: (a) Sentinel-1 HH σ^0 image, (b) ground-truth binary mask, (c) UNet (Control; 3 labelled scenes), (d) UNet (SSL)with BYOL pretraining (3 labelled scenes), (e) UNet (SL; 7 labelled scenes), (f) Random Forest, and (g) SAM (zero-shot). White denotes sea ice and black denotes open water.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

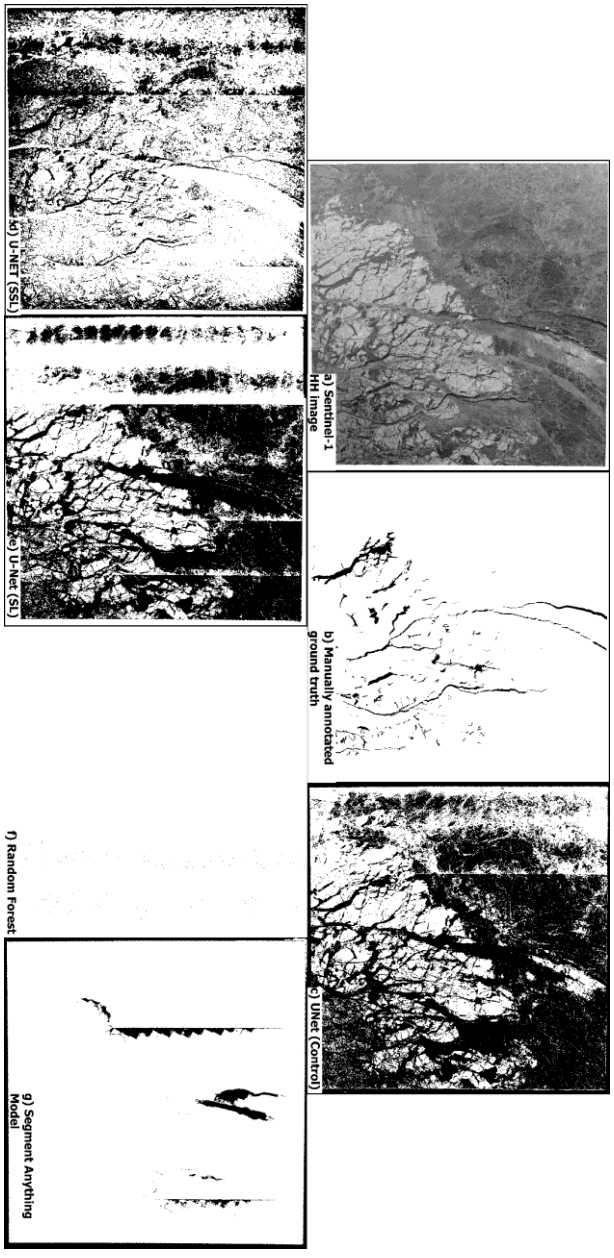

Figure 8. Scene 2 (Marginal Ice Zone). Comparison of ice–water segmentation across models for the test scene. Panels show: (a) Sentinel-1 HH backscatter σ^0 (b) ground-truth binary mask, (c) UNet (Control; 3 labelled scenes), (d) UNet (SSL) with BYOL pretraining (3 labelled scenes), (e) UNet (Supervised; 7 labelled scenes). White denotes sea ice and black denotes open water.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

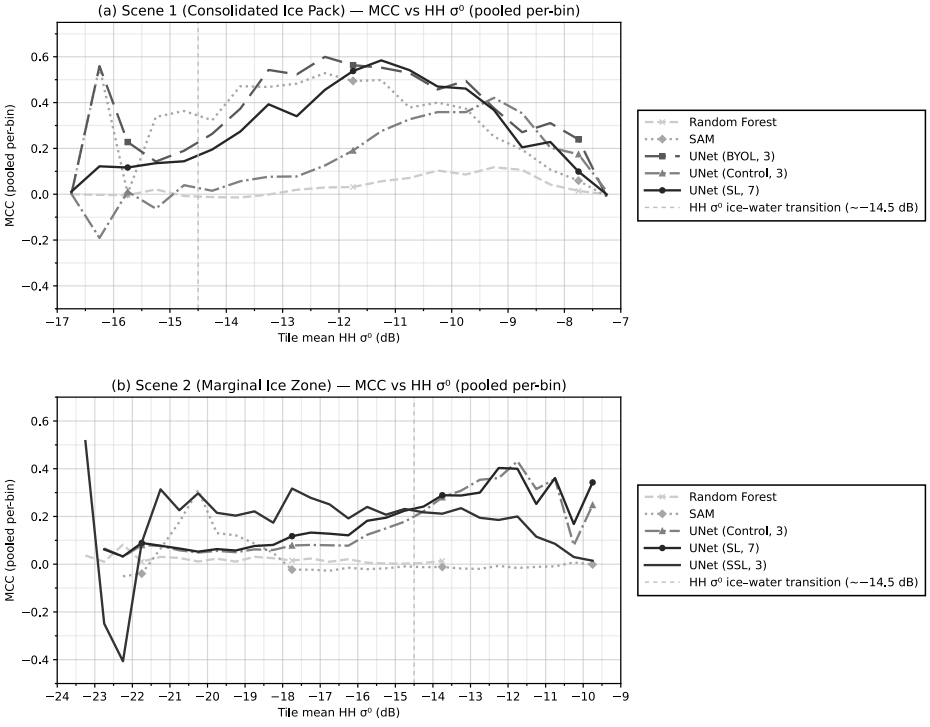

Figure 9. Model performance as a function of HH backscatter (σ^0, dB) for (a) Scene 1 (consolidated-ice scene) and b) Scene 2 (MIZ scene).

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Appendices

Appendix A

Confusion matrices for Scene 1.

Table A1. Confusion Matrix for Scene 1 Supervised learning model

	Predicted Water	Predicted Ice
True Water	584780 (57.3%)	436037 (42.7%)
True Ice	3291105 (12.9%)	22202629 (87.1%)

Table A2. Confusion Matrix for Scene 1 BYOL model

	Predicted Water	Predicted Ice
True Water	438342 (42.9%)	582475 (57.1%)
True Ice	589753 (2.3%)	24903981 (97.7%)

Table A3. Confusion Matrix for Scene 1 Random Forest model

	Predicted Water	Predicted Ice
True Water	25989 (2.5%)	994828 (97.5%)
True Ice	210615 (0.83%)	25283119 (99.17%)

Table A4. Confusion Matrix for Scene 1 Control model

	Predicted Water	Predicted Ice
True Water	852732 (83.5%)	168085 (16.5%)
True Ice	12623265 (49.5%)	12870469 (50.5%)

Table A5. Confusion Matrix for Scene 1 SAM model

	Predicted Water	Predicted Ice
True Water	291147 (28.5%)	729670 (71.5%)
True Ice	162581 (0.64%)	25331153 (99.36%)

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix B

Table B1. Confusion Matrix for Scene 2 Supervised learning model

	Predicted Water	Predicted Ice
True Water	808007 (95.8%)	35647 (4.2%)
	12114113 (48.3%)	12978850 (51.7%)

Table B2. Confusion Matrix for Scene 2 BYOL model

	Predicted Water	Predicted Ice
True Water	668162 (79.2%)	175492 (20.8%)
True Ice	6609154 (26.3%)	18483809 (73.7%)

Table B3. Confusion Matrix for Scene 2 Random Forest model

	Predicted Water	Predicted Ice
True Water	1538 (0.18%)	842116 (99.82%)
True Ice	4041 (0.016%)	25088922 (99.984%)

Table B4. Confusion Matrix for Scene 2 Control model

	Predicted Water	Predicted Ice
True Water	822784 (97.5%)	20870 (2.5%)
True Ice	14379484 (57.3%)	10713479 (42.7%)

Table B5. Confusion Matrix for Scene 2 SAM model

	Predicted Water	Predicted Ice
True Water	49091 (5.8%)	794563 (94.2%)
True Ice	581740 (2.3%)	24511223 (97.7%)

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Acknowledgements

This research was financially supported by the Natural Environment Research Council (NERC) and the University of Aberdeen through QUADRAT DTP. We also acknowledge the generous support of alumni and friends in establishing the University of Aberdeen's Interdisciplinary Institute, which enabled this research.

Code Availability Statement

The model training and evaluation code used in this study has been archived in a FAIR-aligned Zenodo repository:
Seston, J. (2026). BYOL UNet SAR. Zenodo. https://doi.org/10.5281/zenodo.18339243

The archived version corresponds to the experiments and results presented in this manuscript. The repository is currently private for the review process and will be made publicly accessible upon publication.Data Availability Statement

The manually labelled Sentinel-1 SAR sea-ice masks used in this study are openly available at "Manually Labelled Sea Ice Masks for Sentinel-1 SAR Imagery in the Canadian Arctic (2022–2023)". Zenodo, December 1, 2025. https://doi.org/10.5281/zenodo.17780145.

Author Contributions

JS led the study and was responsible for methodology development, software implementation, data curation, formal analysis, visualisation, and preparation of the original manuscript draft. WDH contributed to study conceptualisation, methodology development, supervision, funding acquisition, and manuscript review and editing. GL, BRR, MS, and LM contributed to study conceptualisation and supervision, and reviewed and edited the manuscript. All authors approved the final version of the manuscript.

Competing Interests

The authors declare that they have no competing interests.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

References

von Albedyll, L., Hendricks, S., Hutter, N., Murashkin, D., Kaleschke, L., Willmes, S., Thielke, L., Tian-Kunze, X., Spreen, G., and Haas, C.: Lead fractions from SAR-derived sea ice divergence during MOSAiC, The Cryosphere, 18, 1259–1285, https://doi.org/10.5194/tc-18-1259-2024, 2024.

Alosaimi, N., Alhichri, H., Bazi, Y., Ammour, N., and Melgani, F.: Self-supervised learning for remote sensing scene classification under the few shot scenario, Sci. Rep., 13, 433, https://doi.org/10.1038/s41598-022-27313-5, 2023.

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Bowyer, J., Holland, P. R., Pritchard, H. D., and Suckling, E. B.: Seasonal Arctic sea ice forecasting with probabilistic deep learning, Nat. Commun., 12, 5124, https://doi.org/10.1038/s41467-021-25257-4, 2021.

Ayush, K., Uzkent, B., Meng, C., Tanmay, M., Burke, M., Lobell, D., and Ermon, S.: Geography-aware self-supervised learning, in: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 10115–10125, https://doi.org/10.48550/arXiv.2011.09980, 2021.

Boulze, H., Korosov, A., and Brajard, J.: Classification of sea ice types in Sentinel-1 SAR data using convolutional neural networks, Remote Sens., 12, 2165, https://doi.org/10.3390/rs12132165, 2020.

Buckley, E. M., Cañuelas, L., Timmermans, M.-L., and Wilhelmus, M. M.: Seasonal evolution of the sea ice floe size distribution in the Beaufort Sea from two decades of MODIS data, The Cryosphere, 18, 5031–5043, https://doi.org/10.5194/tc-18-5031-2024, 2024.

Casey, J. A., Howell, S. E. L., Tivy, A., and Haas, C.: Separability of sea ice types from wide-swath C- and L-band synthetic aperture radar imagery acquired during the melt season, Remote Sens. Environ., 174, 314–328, https://doi.org/10.1016/j.rse.2015.12.021, 2016.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G.: A simple framework for contrastive learning of visual representations, in: Proc. 37th Int. Conf. Mach. Learn. (ICML), 1597–1607, https://doi.org/10.48550/arXiv.2002.05709, 2020.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Chicco, D., Tötsch, N., and Jurman, G.: The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation, BioData Min., 14, 13, https://doi.org/10.1186/s13040-021-00244-z, 2021.

Clemens-Sewall, D., Polashenski, C., Frey, M. M., Cox, C. J., Granskog, M. A., Macfarlane, A. R., Fons, S. W., Schmale, J., Hutchings, J. K., von Albedyll, L., Arndt, S., Schneebeli, M., and Perovich, D.: Snow loss into leads in Arctic sea ice: minimal in typical wintertime conditions, but high during a warm and windy snowfall event, Geophys. Res. Lett., 50, https://doi.org/10.1029/2023GL102816, 2023.

Cook, A. J., Dawson, J., Howell, S. E. L., Holloway, J. E., and Brady, M.: Sea ice choke points reduce the length of the shipping season in the Northwest Passage, Commun. Earth Environ., 5, 362, https://doi.org/10.1038/s43247-024-01477-6, 2024.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 248–255, https://doi.org/10.1109/CVPR.2009.5206848, 2009.

Dierking, W.: Sea ice monitoring by synthetic aperture radar, Oceanography, 26(2), 100–111, https://doi.org/10.5670/oceanog.2013.33, 2013.

Durand, C., Finn, T. S., Farchi, A., Bocquet, M., Boutin, G., and Ólason, E.: Data-driven surrogate modeling of high-resolution sea-ice thickness in the Arctic, The Cryosphere, 18, 1791–1815, https://doi.org/10.5194/tc-18-1791-2024, 2024.

Filipponi, F.: Sentinel-1 GRD preprocessing workflow, Proc., 18, 11, https://doi.org/10.3390/ECRS-3-06201, 2019.

Francis, J., Bright, J., Esnaashari, S., Hashem, Y., Morgan, D., and Straub, V. J.: Unsupervised feature extraction of aerial images for clustering and understanding hazardous road segments, Sci. Rep., 13, 10922, https://doi.org/10.1038/s41598-023-38100-1, 2023.

Garrido, Q., Chen, Y., Bardes, A., Najman, L., and LeCun, Y.: On the duality between contrastive and non-contrastive self-supervised learning, in: Proc. Int. Conf. Learn. Represent. (ICLR), https://doi.org/10.48550/arXiv.2206.02574, 2023.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: Proc. 13th Int. Conf. Artif. Intell. Stat. (AISTATS), Proc. Mach. Learn. Res., 9, 249–256, https://proceedings.mlr.press/v9/glorot10a.html, 2010.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z. D., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., and Valko, M.: Bootstrap your own latent: a new approach to self-supervised learning, arXiv [preprint], arXiv:2006.07733, https://doi.org/10.48550/arXiv.2006.07733, 2020.

Guo, W., Itkin, P., Singha, S., Doulgeris, A. P., Johansson, M., and Spreen, G.: Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture, The Cryosphere, 17, 1279–1297, https://doi.org/10.5194/tc-17-1279-2023, 2023.

Harcourt, W. D., Shahin, M., Stearns, L. A., and Shankar, S.: Structural weaknesses in ice mélange revealed by high-resolution ICEYE SAR imagery, J. Glaciol., 71, e108, https://doi.org/10.1017/jog.2025.10085, 2025.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R.: Momentum contrast for unsupervised visual representation learning, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 9726–9735, https://doi.org/10.1109/CVPR42600.2020.00975, 2020.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Hebert, D. A., Allard, R. A., Metzger, E. J., Posey, P. G., Preller, R. H., Wallcraft, A. J., Phelps, M. W., and Smedstad, O. M.: Short-term sea ice forecasting: an assessment of ice concentration and ice drift forecasts using the U.S. Navy's Arctic Cap Nowcast/Forecast System, J. Geophys. Res.-Oceans, 120, 8327–8345, https://doi.org/10.1002/2015JC011283, 2015.

Howell, S. E. L. and Brady, M.: The dynamic response of sea ice to warming in the Canadian Arctic Archipelago, Geophys. Res. Lett., 46, 13119–13125, https://doi.org/10.1029/2019GL085116, 2019.

Howell, S. E. L., Cabaj, A., Babb, D. G., Landy, J. C., Dawson, J., Mahmud, M., and Brady, M.: Near sea ice-free conditions in the northern route of the Northwest Passage at the end of the 2024 melt season, EGUsphere [preprint], EGUsphere-2025-2029, https://doi.org/10.5194/tc-19-6711-2025, 2025.

© Author(s) 2026. CC BY 4.0 License.

Howell, S. E. L., Babb, D. G., Landy, J. C., Glissenaar, I. A., McNeil, K., Montpetit, B., and Brady, M.: Sea ice transport and replenishment across and within the Canadian Arctic Archipelago, 2016–2022, The Cryosphere, 18, 2321–2333, https://doi.org/10.5194/tc-18-2321-2024, 2024.

Huang, B. Q. and Li, X.-M.: Wave attenuation by sea ice in the Arctic marginal ice zone observed by spaceborne SAR, Geophys. Res. Lett., 50, e2023GL105059, https://doi.org/10.1029/2023GL105059, 2023.

Huang, Y., Ren, Y., and Li, X.: Deep learning techniques for enhanced sea-ice types classification in the Beaufort Sea via SAR imagery, Remote Sens. Environ., 308, 114204, https://doi.org/10.1016/j.rse.2024.114204, 2024.

Itkin, P.: Novel methods to study sea ice deformation, linear kinematic features, and coherent dynamic clusters from imaging remote sensing data, The Cryosphere, 19, 1135–1156, https://doi.org/10.5194/tc-19-1135-2025, 2025.

Jain, P., Schoen-Phelan, B., and Ross, R.: Self-supervised learning for invariant representations from multi-spectral and SAR images, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 15, 7797–7808, https://doi.org/10.1109/JSTARS.2022.3204888, 2022.

Jiang, M., Chen, X., Xu, L., and Clausi, D. A.: IceGCN: an interactive sea ice classification pipeline for SAR imagery based on graph convolutional networks, Remote Sens., 16, 2301, https://doi.org/10.3390/rs16132301, 2024.

Johansson, A. M., Brekke, C., Spreen, G., and King, J. A.: X-, C-, and L-band SAR signatures of newly formed sea ice in Arctic leads during winter and spring, Remote Sens. Environ., 204, 162–180, https://doi.org/10.1016/j.rse.2017.10.032, 2018.

Karlsen, T., Johansson, M., Lohse, J., and Doulgeris, A. P.: Incidence angle dependency and seasonal evolution of L- and C-band SAR backscatter over landfast sea ice, Ann. Glaciol., 65, e29, 1–14, https://doi.org/10.1017/aog.2024.30, 2024.

Khaleghian, S., Ullah, H., Kræmer, T., Eltoft, T., and Marinoni, A.: Deep semi-supervised teacher–student model based on label propagation for sea ice classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 14, 10761–10772, https://doi.org/10.1109/JSTARS.2021.3119485, 2021.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R.: Segment Anything, Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Paris, France, 3992–4003, https://doi.org/10.1109/ICCV51070.2023.00371, 2023.

Kortum, K., Singha, S., Spreen, G., Hutter, N., Jutila, A., and Haas, C.: SAR deep learning sea ice retrieval trained with airborne laser scanner measurements from the MOSAiC expedition, The Cryosphere, 18, 2207–2222, https://doi.org/10.5194/tc-18-2207-2024, 2024.

Krähenbühl, P. and Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials, Adv. Neural Inf. Process. Syst., 24, 1–9, https://doi.org/10.48550/arXiv.1210.5644, 2011.

Latifovic, R., Canada Centre for Remote Sensing (CCRS), Canada Centre for Mapping and Earth Observation (CCMEO), and Natural Resources Canada (NRCan): 2020 land cover of Canada, Government of Canada, Ottawa [data set], https://open.canada.ca/data/en/dataset/ee1580ab-a23d-4f86-a09b-79763677eb47, last access: 3 September 2025, 2022.

Lensu, M., Similä, M., Haapala, J., and Karvonen, J.: Ice ridge density signatures in high-resolution SAR imagery of the Baltic Sea, The Cryosphere, 16, 4363–4384, https://doi.org/10.5194/tc-16-4363-2022, 2022.

Lohse, J., Doulgeris, A. P., and Dierking, W.: Mapping sea-ice types from Sentinel-1 considering the surface-type-dependent effect of incidence angle, Ann. Glaciol., 61, 260–270, https://doi.org/10.1017/aog.2020.45, 2020.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal loss for dense object detection, Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2980–2988, https://doi.org/10.1109/ICCV.2017.324, 2017.

Liu, F., Qian, X., Jiao, L., Zhang, X., Li, L., and Cui, Y.: Contrastive learning-based dual dynamic graph convolutional network for SAR image scene classification, IEEE Trans. Neural Netw. Learn. Syst., 35, 390–404, https://doi.org/10.1109/TNNLS.2022.3174873, 2024.

Macdonald, G. J., Scharien, R. K., Duncan, K., Farrell, S. L., Rezania, P., and Tavri, A.: Arctic sea ice topography information from RADARSAT Constellation Mission (RCM) synthetic aperture radar (SAR) backscatter, Geophys. Res. Lett., 51, e2023GL107261, https://doi.org/10.1029/2023GL107261, 2024

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Marbouti, M., Antropov, O., Praks, J., Eriksson, P. B., Arabzadeh, V., Rinne, E., and Leppäranta, M.: TanDEM-X multiparametric data features in sea ice classification over the Baltic Sea, Geo-Spat. Inf. Sci., 24, 313–332, https://doi.org/10.1080/10095020.2020.1845574, 2020.

Meier, W. N., Petty, A., Hendricks, S., Bliss, A., Kaleschke, L., Divine, D., Farrell, S., Gerland, S., Perovich, D., Ricker, R., Tian-Kunze, X., and Webster, M.: NOAA Arctic Report Card 2024: Sea ice, NOAA Tech. Rep. OAR ARC 24-06, U.S. National Oceanic and Atmospheric Administration, Office of Oceanic and Atmospheric Research, https://doi.org/10.25923/aksk-7p66, 2024.

Müller, M., Knol-Kauffman, M., Jeuring, J., and Palerme, C.: Arctic shipping trends during hazardous weather and sea-ice conditions and the Polar Code's effectiveness, npj Ocean Sustain., 2, 12, https://doi.org/10.1038/s44183-023-00021-x, 2023.

Murashkin, D.: Binary sea-ice lead maps derived from Sentinel-1 SAR images during the MOSAiC expedition, PANGAEA, https://doi.org/10.1594/PANGAEA.962904 , 2023.

Muzeau, M., Frontera-Pons, J., Ren, C., and Ovarlez, J.-P.: SAFE: a SAR feature extractor based on self-supervised learning and masked Siamese ViTs, arXiv [preprint], https://doi.org/10.48550/arXiv.2407.00851, 2024.

Nguyen, T., Pham, T. X., Zhang, C., Luu, T. M., Vu, T., and Yoo, C. D.: Dimensional contrastive learning for improving self-supervised learning, IEEE Access, 11, 21534–21544, https://doi.org/10.1109/ACCESS.2023.3236087, 2023.

OSI SAF: Sea ice concentration climate data record, release 3.1 (multimission), EUMETSAT Ocean and Sea Ice Satellite Application Facility [data set], https://doi.org/10.15770/EUM_SAF_OSI_0023, 2025.

Park, J.-W., Korosov, A. A., Babiker, M., Won, J.-S., Hansen, M. W., and Kim, H.-C.: Classification of sea ice types in Sentinel-1 synthetic aperture radar images, The Cryosphere, 14, 2629–2645, https://doi.org/10.5194/tc-14-2629-2020, 2020.

Piao, X., Synn, D., Park, J., and Kim, J.-K.: Enabling large batch size training for DNN models beyond the memory limit while maintaining performance, IEEE Access, 11, 98571–98582, https://doi.org/10.1109/ACCESS.2023.3312572, 2023.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Rajaei, A., Abiri, E., and Helfroush, M. S.: Self-supervised spectral super-resolution for fast hyperspectral and multispectral image fusion, Sci. Rep., 14, 29820, https://doi.org/10.1038/s41598-024-81031-8, 2024.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Commun. Earth Environ., 3, 168, https://doi.org/10.1038/s43247-022-00498-3, 2022.

Ristea, N.-C., Anghel, A., and Datcu, M.: Sea ice segmentation from SAR data by convolutional transformer networks, arXiv [preprint], arXiv:2306.07649, https://doi.org/10.48550/arXiv.2306.07649, 2023.

Roach, L. A., Smith, M. M., Herman, A., and Ringeisen, D.: Physics of the seasonal sea ice zone, Annu. Rev. Mar. Sci., 17, 355–379, https://doi.org/10.1146/annurev-marine-121422-015323, 2025.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, Cham, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Sandven, S., Spreen, G., Heygster, G., Girard-Ardhuin, F., Farrell, S. L., Dierking, W., and Allard, R. A.: Sea ice remote sensing—recent developments in methods and climate data sets, Surv. Geophys., 44, 1653–1689, https://doi.org/10.1007/s10712-023-09781-0, 2023.

Screen, J., and Simmonds, I.: The central role of diminishing sea ice in recent Arctic temperature amplification, Nature, 464, 1334–1337, https://doi.org/10.1038/nature09051, 2010.

Serreze, M. C., and Barry, R. G.: Processes and impacts of Arctic amplification: a research synthesis, Glob. Planet. Change, 77, 85–96, https://doi.org/10.1016/j.gloplacha.2011.03.004, 2011.

Shankar, S., Stearns, L. A., and van der Veen, C. J.: Semantic segmentation of glaciological features across multiple remote sensing platforms with the Segment Anything Model (SAM), J. Glaciol., 70, e4, https://doi.org/10.1017/jog.2023.95, 2024.

Sinha, P., Gaughan, A., Stevens, F., Nieves, J., Sorichetta, A., and Tatem, A.: Assessing the spatial sensitivity of a random forest model: application in gridded population modeling, Comput. Environ. Urban Syst., 75, 132–143, https://doi.org/10.1016/j.compenvurbsys.2019.01.006, 2019.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Singha, S., Johansson, M., Hughes, N., Hvidegaard, S. M., and Skourup, H.: Arctic sea ice characterization using spaceborne fully polarimetric L-, C-, and X-band SAR with validation by airborne measurements, IEEE Trans. Geosci. Remote Sens., 56, 3715–3734, https://doi.org/10.1109/TGRS.2018.2809504, 2018.

Stokholm, A., Høyer, J. L., Pedersen, L. T., Karvonen, J., Kreiner, M. B., and Longépé, N.: The AutoICE Challenge: automated sea-ice classification from Sentinel-1 SAR using operational ice chart labels, The Cryosphere, 18, 3471–3494, https://doi.org/10.5194/tc-18-3471-2024

Stroeve, J., and Notz, D.: Changing state of Arctic sea ice across all seasons, Environ. Res. Lett., 13, 103001, https://doi.org/10.1088/1748-9326/aade56, 2018.

Vahedi, B., Lucas, B., Banaei-Kashani, F., Barrett, A. P., Meier, W. N., Khalsa, S. J. S., and Karimzadeh, M.: Partial label learning with focal loss for sea ice classification based on ice charts, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 17, 13616–13633, https://doi.org/10.1109/JSTARS.2024.3413003, 2024.

Wallace, S., Durrant, A., Harcourt, W. D., Hann, R., and Leontidis, G.: Exploring Segment Anything foundation models for out-of-domain crevasse drone image segmentation, Proc. 6th Northern Lights Deep Learning Conf. (NLDL), Proc. Mach. Learn. Res., 265, 255–268, https://proceedings.mlr.press/v265/wallace25a.html, 2025.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: CBAM: Convolutional block attention module, in: Proc. Eur. Conf. Comput. Vis. (ECCV), 3–19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.

Xu, Y., Sun, H., Chen, J., Lei, L., Ji, K., and Kuang, G.: Adversarial self-supervised learning for robust SAR target recognition, Remote Sens., 13, 4158, https://doi.org/10.3390/rs13204158, 2021.

Ye, K., Cohen, J., Chen, H. W., Zhang, S., Luo, D., and Hamouda, M. E.: Attributing climate and weather extremes to Northern Hemisphere sea ice and terrestrial snow: progress, challenges and ways forward, npj Clim. Atmos. Sci., 8, 166, https://doi.org/10.1038/s41612-025-01012-0, 2025.

Yuan, H., Guo, Q., Ren, Y., Fu, H., and Li, X.-M.: Long-term Pan-Arctic evaluation of a Sentinel-1 SAR sea ice extent product and insights into model integration, Remote Sens., 17, 3166, https://doi.org/10.3390/rs17183166, 2025.

https://doi.org/10.5194/egusphere-2026-376
Preprint. Discussion started: 4 February 2026
© Author(s) 2026. CC BY 4.0 License.

Zhang, Y., Bushuk, M., Winton, M., Hurlin, B., Yang, X., Delworth, T., and Jia, L.: Assimilation of satellite-retrieved sea ice concentration and prospects for September predictions of Arctic sea ice, J. Clim., 34, 2107–2126, https://doi.org/10.1175/JCLI-D-20-0469.1, 2021.

Zakhvatkina, N., Smirnov, V., and Bychkova, I.: Satellite SAR data-based sea ice classification: an overview, Geosciences, 9, 152, https://doi.org/10.3390/geosciences9040152, 2019.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA/ML-CDS 2018), Lect. Notes Comput. Sci., 11045, 3–11, Springer, Cham, https://doi.org/10.1007/978-3-030-00889-5_1, 2018.