

Fusing Ice Surface Temperature With the AI4Arctic Dataset for Improved Deep Learning-Based Sea Ice Mapping

Lily de Loë[✉], *Student Member, IEEE*, David A. Clausi[✉], *Senior Member, IEEE*, and K. Andrea Scott[✉], *Member, IEEE*

Abstract—Arctic sea ice mapping is vital for supporting marine navigation, climate monitoring, and efforts by northern communities to adapt to variable ice conditions. Automated mapping approaches can leverage freely accessible satellite data to supplement navigational ice charts, improve operational forecasting, and produce high-resolution sea ice parameter estimates. The AI4Arctic dataset enables deep learning-based mapping using synthetic aperture radar (SAR), passive microwave (PM), and reanalysis data. However, SAR and PM can struggle to resolve ice features due to ambiguous textures, atmospheric effects, and sensor limitations. To provide complementary data, an 84-scene Visible Infrared Imager Radiometer Suite (VIIRS) dataset is co-registered with AI4Arctic to evaluate whether ice surface temperature (IST) measurements can improve estimation of sea ice concentration (SIC), stage of development, and floe size. Input- and feature-level fusion methods, based on the U-Net architecture, are explored. Models are evaluated using the SIC R^2 coefficient and SOD/FLOE F1-score, as well as predicted sea ice maps. In addition, an alternative SIC accuracy score is introduced to assist with evaluating marginal ice predictions. Incorporating IST improves performance across all models compared to the AI4Arctic baseline; this includes single-encoder, dual-encoder, and multidecoder U-Nets. Results highlight significant improvements in the prediction of open water under conditions with low-incidence angle, high atmospheric moisture, and wind roughening. Overall, the best performing dual-encoder model, DUE-Net-V, improves predictions by 2.18–5.01% across all metrics, relative to the baseline. These results support integrating IST in deep learning workflows and highlight the potential for next-generation thermal-infrared sensors to improve automated sea ice mapping.

Index Terms—Data fusion, deep learning, microwave data, synthetic aperture radar (SAR), sea ice, thermal data, U-Net.

I. INTRODUCTION

AUTOMATED sea ice mapping is a valuable tool for navigating changing Arctic conditions, with the potential to support several strategic goals. Diminishing ice extent

Received 4 June 2025; revised 28 August 2025; accepted 11 September 2025. Date of publication 16 September 2025; date of current version 3 October 2025. This work was supported in part by Environment and Climate Change Canada, with compute resources provided by the Digital Research Alliance of Canada, and in part by the Vector Scholarship in Artificial Intelligence, provided through the Vector Institute. (*Corresponding author: Lily de Loë*)

Lily de Loë and David A. Clausi are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G5, Canada (e-mail: lcbdeloe@uwaterloo.ca; dclausi@uwaterloo.ca).

K. Andrea Scott is with the Department of Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G5, Canada (e-mail: ka3scott@uwaterloo.ca).

Digital Object Identifier 10.1109/JSTARS.2025.3610260

contributes to increasing variability in Arctic conditions [1], [2]. This variability, combined with longer shipping seasons, necessitates the development of automated mapping approaches to supplement ice charts, and support strategic planning with shorter lead times [3]. In addition, changing ice dynamics make conditions unpredictable, which increases the risks of encountering hazards. Data pipelines can be created so that remote northern communities can leverage automated mapping and combine it with traditional knowledge to support Indigenous-led community initiatives [4]. Further, Arctic amplification contributes to accelerated warming in this region [5]. Decreasing ice extent impacts Arctic ecosystems, where sea ice acts as habitat for endemic species [6]. Timely, high-resolution mapping approaches can assist in monitoring freeze-up and break-up trends, as well as ice extent, supporting climate research [7].

Advancements in sea ice mapping are increasingly driven by machine learning approaches, which leverage a wealth of free, publicly available remote sensing data. These approaches primarily leverage synthetic aperture radar (SAR) and passive microwave (PM) data. Empirical, algorithm-based products exist, which predominantly predict sea ice concentration (SIC). This includes daily SIC maps derived from PM data [8], from fusing PM and thermal-infrared data [9], or from fusing optical and SAR data [10]. SAR-derived products also exist, which offer finer resolution estimates [11]. Recent advancements in deep learning have demonstrated promising results in estimating the three sea ice parameters relevant to ice charts: SIC, stage of development (SOD), and floe size (FLOE) [12]. These studies leverage convolutional neural networks (CNNs) [13], [14], U-Net architectures, [15], [16], [17], [18], and U-Net variants [19].

The AI4Arctic dataset [20], introduced in 2022, supports the continued development of deep learning-based mapping approaches by providing co-registered Sentinel-1 SAR, AMSR2 brightness temperature (BT), ERA-5 reanalysis data, and ice charts. SAR measurements are not reliant on reflected solar radiation, and boast a spatial resolution that balances sensitivity to changes in ice conditions with swath size and data storage considerations. These instruments measure the backscatter coefficient of the emitted radar pulse; surface roughness, incidence angle, and the physical and dielectric properties of the target medium all impact this measurement [21], as well as the wavelength of the pulse. Passive microwave (PM) products are commonly used for estimating SIC because BT, measured

by the sensor, effectively distinguishes ice from water [13], [22], [23], [24]. Further, instruments operate independently of incident solar radiation, and microwave radiation is capable of penetrating clouds at lower microwave frequencies, such as within the 6.9–18.7 GHz range [25]. A subset of the ERA5 climate variables can influence ice formation and break-up, including wind speed and temperature [26].

Ice charts, which are used as labels, enable the development of multitask, supervised models that are trained to predict SIC, SOD, and FLOE. Accurate estimates of SIC can be used to estimate sea ice extent, the total area covered by $\geq 15\%$ concentration sea ice, which is a key climate change indicator for the Arctic region [27]. Further, high resolution maps of SIC are effective at illustrating the marginal ice zone (MIZ): the transitional region between full ice-cover and open water [28]. Long-term measurements of ice age and thickness, characterized by SOD, are valuable for maritime strategic planning [29], as well as climate monitoring and modeling [7]. Finally, sea ice break-up, captured by floe size, impacts the exchange of momentum between ocean and atmosphere [30]. Knowledge of floe size is particularly important in the MIZ, where small floes accelerate the lateral melt rate [31].

A. Limitations in Remote Sensing Data

While AI4Arctic allows researchers to more easily develop models that combine the strengths of SAR and PM sensors, both data types present limitations that impact the accuracy of model predictions. The interpretability of these remotely sensed measurements can be impacted by sensor limitations, surface textures, and atmospheric effects, discussed below. These factors present challenges for machine learning models, which may struggle to correctly classify ice conditions without additional context (e.g., regional or historical context known by ice analysts but omitted from the input data). The term ambiguity, used throughout this work, refers to measurements that may limit a model's ability to recognize patterns. This includes conditions where a measurement could represent one of two conditions (e.g., wind-roughened water and ice exhibiting similar backscatter signatures [15]), or where moisture (e.g., in the atmosphere for PM [32], or in the snow layer for SAR [33]) could obscure surface conditions.

In SAR imagery, backscatter signatures from surface textures such as wind-roughened water and smooth ice can be misinterpreted as ice or still water, respectively [15]. These signatures introduce ambiguity because they can represent multiple conditions, with additional information being required for a model to correctly distinguish ice from water. In addition, the dielectric properties of snow on the ice surface can influence backscatter mechanisms, where for first year ice (FYI), brine-wetted snow pack causes scattering at the snow-ice interface [33]. Further, backscatter signatures for ice and water vary by incidence angle [34]; this angle, which varies by dataset scene, impacts the appearance of open water and different ice types. For example, open water at near range in a ScanSAR image will have a strong backscatter and a bright appearance in a SAR image, while open water at far range will mostly reflect, leading to a dark

appearance. Thus, incidence angle information is required to correctly interpret these measurements.

PM sensors provide data at a coarse spatial resolution, on the scale of kilometers. Consequently, they are ineffective near coastal boundaries and narrow waterways, where the sensor footprint introduces land contamination in measurements of ocean pixels [35]. In addition, studies have found that SIC estimates using PM data are biased for thin ice due to low BT emissivity [36], [37]. Similarly to SAR, open water measurements can be impacted by atmospheric effects, such as wind-roughening [38]. Further, at certain frequencies, these sensors are sensitive to atmospheric moisture [32]. Thus, readings from individual channels may introduce ambiguity, where limitations in sensor resolution and surface obstruction from atmospheric effects impact the accuracy of surface measurements.

B. Study Contributions

AI4Arctic does not include imagery from the optical, near, or mid-infrared ranges of the electromagnetic spectrum. We propose that models trained using an additional thermal-infrared channel may improve the accuracy of sea ice parameter predictions, where SAR and PM data are challenging for machine learning models to interpret. Specifically, there is a potential for ice surface temperature (IST) measurements to resolve ambiguous backscatter signatures in SAR imagery, and to supplement PM data when accurate surface measurements are obstructed by atmospheric interference. Thermal infrared data may provide insight into ice age and thickness [39], [40]. Further, current products (e.g., from MODIS [41] and Visible Infrared Imager Radiometer Suite (VIIRS) [42]) offer finer resolution data than PM imagery, allowing for accurate measurements where features in PM are unresolved (e.g., near land boundaries). Ultimately, IST measurements, which indicate surface temperature, have the potential to provide a strong control on water at high spatial resolution, which complements AI4Arctic.

This work extends the dataset introduced by de Loë et al. [43]—the study which verified that VIIRS IST data could be used in conjunction with the AI4Arctic dataset. We make the following, additional contributions.

- 1) We co-locate and georeference 54 additional VIIRS scenes for use with the AI4Arctic dataset. This work extends our preliminary evaluation of IST data [43], providing 84 total scenes and the basis for the largest evaluation of AI4Arctic and VIIRS data.
- 2) We assess the inclusion of IST with AI4Arctic against the winning Artificial Intelligence for Earth Observation (AI4EO) AutoICE Challenge submission by Chen et al. [18], which acts as the baseline for sea ice parameter estimation. This work is accomplished using established evaluation metrics and predicted sea ice maps. Further, we introduce an alternative metric for comparing the accuracy of SIC predictions.
- 3) We evaluate multiple approaches for fusing VIIRS IST data at the input- and feature-levels. Comparing to the best performing fusion approach from de Loë et al. [42], we assess three variants on the U-Net architecture, which

fuse features learned from combinations of IST, SAR, PM, and reanalysis data in two separate encoders. The presented fusion methodologies allow us to address three main hypotheses:

- a) That learning the combined features of Sentinel-1 and VIIRS input channels can leverage the improvements seen from including PM data for ice–water distinction, at a higher resolution than AMSR2, which suffers from land contamination.
- b) That learning the salient features of VIIRS data, independent of the other input channels, may show unique improvements, where features are not consistent between Sentinel-1 and VIIRS, or where VIIRS data is cloud-obscured.
- c) That use of task-specific decoders may allow the network to leverage a common set of learned features to achieve more accurate estimates of conditions that are specific to one sea ice parameter.

To the best of our knowledge, this is the first published research to incorporate VIIRS IST data as inputs in dual-encoder and multidecoder U-Net architectures, and the first to estimate SIC, SOD, and FLOE using these architectures.

II. BACKGROUND

The proposed architectures are designed to perform multitask, supervised semantic segmentation. Accordingly, this section introduces CNN-based approaches to automated sea ice mapping, methods for multimodal data fusion, as well as the performance baseline used to assess inclusion of IST data, and the best performing architecture from the preliminary de Loë et al. [43] study.

A. U-Net Architectures

CNNs are common approach to deep learning-based sea ice parameter estimation. Contributions to the field include Wang et al. [44] and Boulze et al. [14], who used SAR imagery to derive estimates of SIC and SOD, respectively. In addition, Malmgren–Hansen et al.’s [13] estimation of SIC stands as one of the earliest works that fuses SAR and PM data in a deep learning model. The more recent U-Net architecture, designed for medical imaging, outperforms traditional CNNs on image segmentation [45]. Consequently, this architecture has seen increasing adoption when estimating sea ice parameters as a class-based segmentation task.

U-Net architectures consist of a contracting encoder path and expanding decoder path, which are bridged by skip connections. This symmetric structure, from which the name U-Net is derived, allows for the extraction of hierarchical features in the encoder, while restoring the latent representation to its original spatial dimensions in the decoder. Skip connections, which concatenate local encoder feature maps with their corresponding decoder blocks, preserve high-resolution features and spatial information. The output of the network is n segmentation maps (where $n > 1$ for a multitask architecture) at the spatial resolution of the input image(s), for the specified classes.

In the context of sea ice mapping, U-Net architectures have been used to estimate sea ice cover from SAR, improving the accuracy of estimates in the MIZ [15], [16]. More recently, the AI4SeaIce article series [17], [46], [47], [48] established the basis for the use of U-Nets in the AI4EO AutoICE Challenge [29]. As a result of this competition, where models were trained and evaluated on the AI4Arctic dataset, U-Net architectures such as Chen et al. [18] emerged as a leading method for automating sea ice mapping based on multimodal remote sensing data.

Further, variations of the U-Net’s encoding and decoding paths have succeeded at improving the architecture for specific tasks. For example, in sea ice mapping, multiple encoders have been used to fuse multimodal satellite data. In their ViSual_IceD architecture, Rogers et al. [19] separated SAR and multispectral imagery in parallel encoders for ice–water classification. Comparing to a traditional, single-encoder U-Net, this method of learning features from data at different spatial resolutions improves model performance. In addition, in the field of medical imaging, Mahbod et al. [49] introduced a dual-decoder U-Net for nuclei instance segmentation. The decoders simultaneously predict foreground and distance maps, and these outputs are merged to improve segmentation performance for nuclei classification.

B. Data Fusion for Sea Ice Mapping

In remote sensing, data fusion approaches enable models to make predictions based on imagery from multiple different sensors, overcoming the limitations of a single product. Fusion methods can be classified as one of three approaches: input-level, feature-level, and decision-level.

1) *Input-Level Fusion*: In input-level fusion, models learn common features shared across input channels. This method is beneficial because it enables the network to learn features that may not be evident from a single input [50]. This study assesses two architectures that use input-level fusion. We compare the best performing model from de Loë et al. [43], referred to as AutoICE-V, to dual-encoder and multidecoder U-Net variants. We also maintain the same baseline as the de Loë et al. study: the AutoICE U-Net architecture proposed by Chen et al. [18].

2) *Feature-Level Fusion*: Feature-level fusion combines features learned from separate inputs in different parts of the network. This approach is beneficial for fusing inputs from different data sources, allowing the model to identify salient features. Malmgren–Hansen et al. [13] and Rogers et al. [19] both employed feature-level fusion: each model accepts input channels with different resolutions, which are combined to make pixel-wise predictions. The success of these models in fusing high resolution SAR imagery with coarser PM measurements [13], and multispectral imagery [19], demonstrates the benefits of this approach for satellite products.

3) *Decision-Level Fusion*: Decision-level fusion combines results from multiple classifiers, often with a weighting scheme, to improve the final decision for a single task [50]. For example, Wang et al. [51] used a Bayesian framework to combine ice–water information from SAR data with the AMSR-E daily ice concentration product, improving the detail and accuracy of SIC estimates. This approach is also common in algorithm-based

satellite sensor fusions, which do not use machine learning. For example, both Ludwig et al. [9] and Konig et al. [10] (see Section I) merge SIC estimates obtained from multiple remote sensing products.

III. DATA

This study combines the AI4Arctic dataset with VIIRS IST data that meets established criteria. The following sections introduce the data used as input channels, in addition to the method for creating the co-registered VIIRS-AI4Arctic dataset, and its composition.

A. AI4Arctic Data

AI4Arctic consists of geo-referenced SAR and PM imagery, reanalysis data, and auxiliary information. Associated ice charts from the Canadian and Greenland Ice Services are provided, which are used by models during training and evaluation. The full dataset consists of 533 scenes sampled from January 2018 to December 2021, with acquisition frequency varying by region. While 513 scenes were released during the AI4EO AutoICE Challenge and 20 reserved for evaluation, this study uses a separate test set with minimal cloud cover to better assess IST data under diverse conditions.

This study uses the Ready-to-Train AI4Arctic dataset, which provides normalized data [20]. SAR, auxiliary, and ice chart data are provided at a pixel spacing of 80 m, down-sampled from the original 40 m spacing. PM and reanalysis measurements are gridded to a resolution of 2 km. Each scene also includes geolocation and timestamp information.

1) *Sentinel-1 SAR*: Sentinel-1 HH and HV polarized, Extra Wide Swath images are provided, which have been noise corrected using the NERSC algorithm [52]. This imagery reports the backscatter coefficient, measured in dB. Incidence angle and the distance to the nearest coast measurements are provided as auxiliary information at the same pixel spacing.

2) *AMSR2 PM*: The Advanced Microwave Scanning Radiometer 2 (AMSR2) provides horizontally and vertically polarized BT measurements at 6.9, 7.3, 10.65, 18.7, 23.8, 36.5, and 89 GHz. Spatial resolution varies by channel; 6.9 GHz offers the coarsest ground resolution (35×62 km), whereas 89 GHz offers the finest resolution (3×5 km) [53].

3) *ERA5 Reanalysis Data*: Five numerical weather prediction (NWP) parameters are provided from the ERA5 reanalysis dataset: 10 m wind speed (given as components), 2 m air temperature, total column water vapour, total column cloud liquid water, and skin temperature [29].

4) *Ice Charts*: Ice charts report conditions using Egg Codes [12], where polygonal regions are defined in terms of the total SIC, partial concentrations, stages of development, and floe sizes. In this context, SIC is the sum of partial concentrations, reported in tenths, and associated with the n thickest ice conditions in the polygon (typically $n = 3$). SOD relates to the age and thickness of ice, while floe size characterizes the degree of break-up. Similar to SIC, n measurements are reported, each associated with a partial concentration [12].

AI4Arctic ice charts labels report total SIC, dominant SOD, and dominant FLOE. Dominant classes are assigned if there is a majority partial concentration, representing $\geq 65\%$ of the polygon; otherwise, SOD and FLOE pixels are masked. Ice charts are co-registered with Sentinel-1 SAR imagery, and an egg code's information is applied to every pixel contained in that polygon [29]. Each pixel within a polygon inherits its label, although this may not reflect pixel-level SAR variation. Further, analysts may overestimate low SIC, and report more variability for mid-range concentrations, impacting the accuracy of these labels [54].

B. VIIRS IST

The VIIRS provides IST data in the form of 6-min, L2 Swath scenes with a pixel spacing of 750 m. Calculating IST follows a similar approach to MODIS: VIIRS calibrated radiance data from bands M15 and M16 are converted into BT measurements [55], and IST is calculated using a split-window method. IST values are reported in Kelvin, and the product contains flags for features such as land, ocean, and cloud. Targeted uncertainty for this product is $\pm 1\text{K}$ over the range of 213–275 K [56].

C. VIIRS Scene Selection

Scene selection criteria remain consistent with those used in de Loë et al. [43]. However, the threshold for the percentage of cloud contaminated pixels was increased from 50% to 60% in order to increase the size of the dataset. These criteria are summarized in Table I. In addition to those thresholds, VIIRS scenes are graded as “good,” “mid,” or “poor” based on percentages of cloud contaminated pixels and scene overlap with the Sentinel-1 latitude-longitude grids. These grades are used to select scenes for each of the train-validation and test sets, discussed in Section III-D.

1) *Acquisition Window*: Same day acquisition matches the retrieval criteria for the AI4Arctic dataset [20]. The maximum acquisition window is set relative to similar studies; specifically, Malmgren-Hansen et al.’s [13] use of 7 h preceding and 2 h following the Sentinel-1 imagery, and Rogers et al.’s [19] use of 8 h in regions with low drift velocities. Comparing the AMSR2 sensor footprint and the finer spatial resolution of MODIS MSI product (100 m), a threshold of 8 h is sufficient for VIIRS. Further, of the scenes used in the dataset, only 6% had an acquisition window greater than 7 h (see Table I).

2) *Scene Composition*: A minimum area of overlap between the Sentinel-1 and VIIRS scenes ensures that sufficient VIIRS IST data is present in the image, as regions of fill are masked and do not contribute to model predictions. A general temperature threshold is used to identify the percentage of IST pixels registering as sea ice, which freezes at 271.4 K [56]. Because thermal-infrared data may provide insights into ice age and thickness [39], [40], an additional threshold is used to confirm the presence of older ice, ensuring that the dataset evaluates use of IST data under varied ice conditions.

3) *Co-Registration Method*: Viable VIIRS scenes are co-registered, per pixel, with the Sentinel-1 images. The latitude-longitude grids provided by AI4Arctic are up-sampled, and

TABLE I
SUMMARY OF SCENE SELECTION CRITERIA, GRADING

Threshold	Scene Categorization	% of Scenes	Additional Criteria
Acquisition	Same day	100	
	Good	45	$\leq 20\%$ cloud cover; $\geq 90\%$ overlap with Sentinel-1
	Mid	39	$\leq 40\%$ cloud cover; $\geq 90\%$ overlap with Sentinel-1
Hours between Sentinel-1 and VIIRS acquisitions	Poor	16	$\leq 60\%$ cloud cover; $\geq 75\%$ area of overlap
	≤ 5 hours	62	
	≤ 7 hours	32	Scenes are not assigned to a dataset based on acquisition window
VIIRS IST values registering as ice	≤ 8 hours	6	
	$\geq 50\%$ pixels below 271 K; $\geq 15\%$ pixels below 260 K	100	Visual analysis was used to select inference sets that contained varied ice conditions

a polygonal region of intersection between the VIIRS and Sentinel-1 images is found. All VIIRS pixels outside of this region are discarded, and a nearest neighbor approach is used to co-register the scenes. This approach, discussed in de Loë et al. [43], effectively handles differences between IST data (temperatures in Kelvin), flags (integers outside of the valid temperature range), and coordinates that exceed the bounds of the Sentinel-1 scene (set to NaN). Prior to use in a model, flag data is masked and the remaining IST data are normalized using the mean and standard deviation. IST measurements in cloud-contaminated pixels, which report inaccurate measurements, and are masked by the product [56].

D. VIIRS-AI4Arctic Dataset Composition

This study separates the 84-scene VIIRS-AI4Arctic dataset into a 72-scene train-validation set, and 12-scene test set, which is held out for evaluation. Due to the limited size of the dataset, the train-validation set is also used for training, validation, and testing during hyperparameter tuning. The train-validation set is further subdivided into six, 12-scene folds; during hyperparameter tuning, folds 1–5 were used for training and validation, while fold 6 was used for testing. Several considerations, summarized below, inform the selection of these datasets.

1) *Percentage of Cloud Cover*: Due to emphasis on assessing predictions with IST data, scenes assigned “good” and “mid” quality grades were prioritized for fold 6 and the test dataset. Mid quality scenes were included in order to achieve a representative sample of years, months, regions, and ice conditions. Scenes belonging to all three quality grades were assigned to the train-validation set; “poor” scenes with a higher percentage of cloud cover are included, because they are still useful for training.

2) *Spatial-Temporal Distribution*: The formation of sea ice—referred to as freeze-up—begins in early fall; conversely, ice break-up begins in the spring, with melt accelerating throughout the summer [1], [6]. The VIIRS-AI4Arctic dataset encompasses November through May, capturing freeze-up and break-up, but excluding late melt conditions. June through October were omitted from the dataset, based on the limited availability of non-cloudy scenes. The train-validation and test sets were selected to adequately sample across space and time; this decision was prompted by findings from Chen et al. [18] that spatial-temporal encodings increase the accuracy of SOD predictions when the

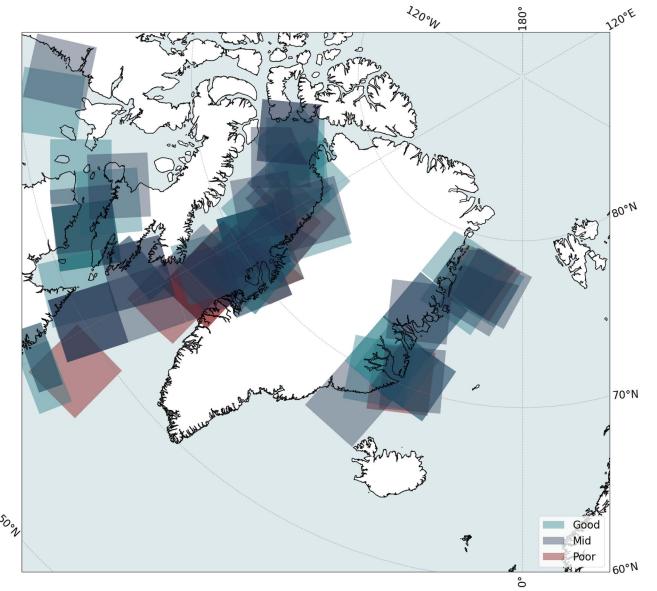


Fig. 1. Study region of the 84 scene dataset. The total scene distribution is 22% (Region 1: Hudson Bay & lower Baffin Bay), 49% (Region 2: upper Baffin Bay & Canadian Archipelago), and 29% (Region 3: Greenland Sea).

sampling is sufficient. The full 84-scene dataset is illustrated in Fig. 1. The 12-scene test set, comprising a subset of scenes, features a comparable distribution of years and regions, and evenly samples acquisition months.

3) *Pixel Class Distribution*: Table II summarizes the distribution of pixels, by ice chart class, for each of the sea ice parameters. The test set was selected such that there would not be any large class imbalances compared to the full train-validation dataset. This class distribution covers scenes with a variety of ice conditions (full ice cover, MIZs, landfast ice, etc.).

4) *Percentage Unmasked Ice Chart Data*: AI4Arctic ice chart labels mask certain classes (e.g., pancake ice), polygons with no dominant ice type, and land. As a result, the amount of masked data in a scene is inconsistent; Table III demonstrates that a fold may contain as much as 58.98% unmasked data, or as little as 15.82%, and that SOD and FLOE charts are disproportionately affected. Resultantly, the test data was selected such that SOD and FLOE would contain a comparable fraction of masked pixels to SIC.

TABLE II
SUMMARY OF CLASS DISTRIBUTION BY SEA ICE PARAMETER FOR EACH OF THE TRAIN-VALIDATION AND TEST SETS

SIC	0	10	20	30	40	50	60	70	80	90	100
Train-val [%]	21.91	0.11	1.11	2.11	0.72	1.58	0.70	1.85	5.21	10.03	54.66
Test [%]	24.65	0.19	2.13	1.26	0.51	0.29	1.34	2.82	4.20	12.84	49.78
SOD	Open water	New Ice	Young Ice	Thin FYI	Thick FYI	Old Ice					
Train-val [%]	34.14	2.16	12.29	10.31	32.89	8.22					
Test [%]	30.38	0.41	16.51	9.17	33.44	10.10					
FLOE	Open water	Cake Ice	Small Floe	Medium Floe	Big Floe	Vast Floe	Bergs				
Train-val [%]	31.86	0.44	2.36	7.37	7.94	49.97	0.06				
Test [%]	34.63	0.00	3.05	10.06	7.95	44.19	0.13				

TABLE III
PERCENTAGE OF UNMASKED PIXELS, REPORTED FOR SIC, SOD, AND FLOE

Unmasked Pixels [%]	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Test Set
SIC	40.74	46.48	58.98	53.74	58.45	47.00	57.75
SOD	15.82	25.67	40.17	36.92	40.12	28.42	46.87
FLOE	22.77	30.36	42.08	41.46	38.93	30.57	41.12

Statistics are reported by 12-scene fold. Folds 1-6 are used for training and validation, and the test set is fixed.

IV. METHODOLOGY

A. Data Pre- and Postprocessing

This study assesses the impacts of IST data by comparing predictions using the same 16 AI4Arctic channels selected by Chen et al. [18]. Sentinel-1 HH and HV images are used in conjunction with incidence angle and distance to land. Incidence angle information is provided due to its impact on radar backscatter [34], while distance-to-land maps are useful because the scene's associated PM measurements become inaccurate near land [15], [29]. Of the seven frequencies and two polarizations, four AMSR2 BT channels (18.7 and 36.5 GHz, H and V) are used. These channels balance sensitivity to sea ice and insensitivity to atmospheric effects. Lower-frequency channels have coarser spatial resolutions [16], while higher ones (e.g., 89 GHz) are more sensitive to atmospheric phenomena (e.g., water vapor) [57].

These channels are supported by five of the ERA5 reanalysis parameters, excluding skin temperature. In addition, gridded latitude and longitude data, as well as the scene's acquisition month, are provided as auxiliary inputs and used for spatial-temporal encodings. Finally, three sets of ice charts, corresponding to total SIC, dominant SOD, and dominant FLOE are provided, per scene; these channels are used as reference labels instead of network inputs.

Prior to entering the network, the 80-m spaced Sentinel-1 channels, co-registered IST measurements, and ice charts are downsampled by a factor of 10, reducing these scenes to approximately 500×500 pixels. This downsampling brings the Sentinel-1 resolution close to the original resolution of the VIIRS thermal channel. This downscaling results in a larger receptive field, improving the model's ability to predict the polygon-based ice chart labels, which are inconsistent with pixel-level SAR measurements. Concurrently, the gridded AMSR2, ERA5, and auxiliary channels are upsampled to match this new pixel spacing. Uniform height and width dimensions are enforced across channels by extracting randomly cropped, 256×256

TABLE IV
SUMMARY OF MODELS, BY NAME AND DESCRIPTION

Fusion Type	U-Net Type	Model Name	Description
Input-level fusion	Base U-Net	AutoICE	The Chen et al. [18] architecture, used as a performance baseline.
		AutoICE-V	AutoICE, incorporating VIIRS IST data.
Feature-level fusion	Dual-encoder U-Net	DEU-Net-SV	DEU-Net where encoder 1 contains SAR and VIIRS data.
		DEU-Net-V	DEU-Net where encoder 1 contains only VIIRS data.
	Tri-decoder U-Net	TDU-Net-V	DEU-Net-V encoders, separate decoders for SIC, SOD, and FLOE.

TABLE V
MODEL SPECIFICATIONS, SUMMARIZED BY ARCHITECTURE

Model Specification	AutoICE, AutoICE-V	DEU-Net-SV, DEU-Net-V, TDU-Net-V
Optimizer	SGDM	AdamW
Learning Rate (LR)	0.001	0.0005
LR Coefficients	Weight Decay: 0.01	Betas: 0.9, 0.99
Batch Size	16	4
LR Scheduler		Cosine Annealing (CA)
CA Starting Epochs till Restart	20	10
CA Restart Multiplier	1	2
Down-scaling Factor	10	10
Data Augmentations	Rotation, flip, random scale, cutmix	
Patch Size	256	
Loss Functions	SIC: MSE, SOD and FLOE: CE	
Loss Function Weights	SIC: 1, SOD and FLOE: 3	

Note: stated specifications are those that differ from the torch default parameters.

patches from scenes in the training set. Data augmentation techniques, summarized in Table V, are employed to increase the training dataset's size. During evaluation, validation and test scenes are still downsampled. However, entire scenes are accepted as inputs to the trained model, instead of patches. As a result of this preprocessing, sea ice maps are predicted at the downsampled spatial resolution. At the output of the network, predictions are interpolated such that maps are evaluated at the original 80 m pixel spacing. Ultimately, this approach, seen in Fig. 2, follows previous work using the full AI4Arctic dataset and labels, which achieved enhanced accuracy and improved the capture of contextual information [18].

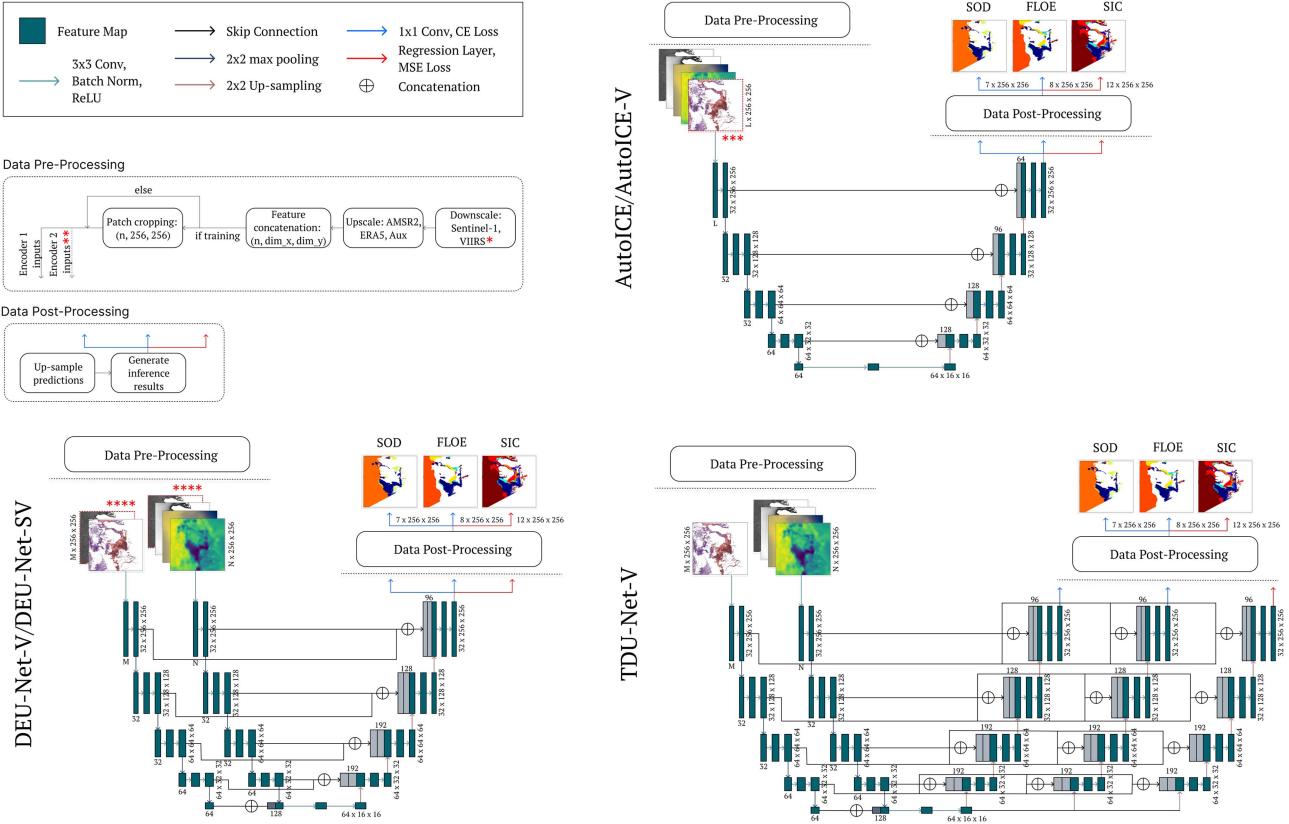


Fig. 2. Summary of the data processing pipeline, and three base architectures used by the study. Input and output feature map dimensions for each architecture are labelled. *Downscaling of VIIRS data applies to AutoICE-V, both DEU-Net architectures, and TDU-Net-V. ** Encoder 2 applies to DEU-Net-V, -SV, and TDU-Net-V. ***IST data is omitted from AutoICE. ****SAR input channels are grouped with VIIRS in DEU-Net-SV, and the remaining AI4Arctic channels in DEU-Net-V.

B. Summary of U-Net Architectures

Three variants on the U-Net architecture were developed, which assess methods of fusing IST data at the feature-level. These architectures are compared to the baseline AutoICE architecture [18], and the AutoICE-V [43] method for fusing IST data at the input-level. Two dual-encoder U-Nets are trialled, which 1) group inputs by finer (SAR and IST) and coarser (PM, ERA5, aux) spatial resolution, and 2) learn IST features independently from the AI4Arctic channels. Following tuning, these models were compared to AutoICE-V. Finally, a tri-decoder U-Net, which employs separate decoders for SIC, SOD, and FLOE, was trained based on the encoding path from the best performing architecture. Models are summarized, by name, in Table IV.

These architectures build on the base AutoICE structure [18], which consists of four sets of encoder and decoder blocks, connected by a bridge layer. A single block is comprised of two sets of convolutional layers, followed by batch normalization and the ReLU activation function. During tuning, changes to this structure yielded no performance improvements (see Section IV-C). Thus, all architectures consist of these four-block sets, varying the number of encoder and decoder paths. Feature map dimensions associated with each block are specified in Fig. 2. For each of SOD and FLOE, outputs from the final decoder block are fed into a 1×1 convolution layer followed

by the softmax activation function; the model predicts the pixel as belonging to one of seven (SOD) or eight (FLOE) discrete classes. Concurrently, the decoder output for SIC is fed into a regression layer. Prior to evaluation, SIC predictions are rounded to the nearest tenth, matching the ice chart labels. Based on the task's treatment as a regression or classification problem, mean squared error (MSE) and cross entropy (CE) loss are used, respectively. A cumulative, weighted loss function, comprised of the task-specific losses, informs model training. A summary of model specification is provided in Table V, while Fig. 2 summarizes the model architectures.

Both the AutoICE and AutoICE-V architectures are retrained on this study's 84-scene dataset to provide 1) a fair comparison to the baseline, and 2) an assessment of input- and feature-level fusion approaches, evaluated on a larger dataset, and comprising a wider variety of ice conditions. Two feature-level fusion approaches are compared to identify the key differences in learning combined features from IST and SAR, versus features learned solely from IST data. Previous works (see Section II-B) fuse data from two instruments; thus, these models learn features from each data type in separate encoder branches. Following this approach, two-, three-, and four-encoder U-Nets could be evaluated. Ultimately, this study seeks to test existing findings

with DEU-Net-SV, grouping the 80-m SAR and co-registered IST data in one encoder, and the gridded AMSR2, ERA5, and auxiliary data in another. In addition, due to the masking of cloud-contaminated pixels in IST data, we anticipate an imbalance in features learned from the Sentinel-1 data. Thus, DEU-Net-V intends to assess if salient features learned from IST can reduce misclassification from SAR and PM data. Last, TDU-Net-V investigates the use of multiple decoders, which has achieved success in other domains [49], but is yet untested for multitask sea ice mapping. This study assesses 1) how use of multiple decoders could improve segmentation of SIC, SOD, and FLOE, and 2) if improvements in individual parameters reduce overall accuracy because ice chart labels are assigned over shared polygonal regions.

C. Experimental Setup

All experiments are conducted using an NVIDIA A100-SXM4-40 GB GPU, 60 GB of memory, and Pytorch version 2.0.1 with CUDA 11.8. These resources are accessed using the Narval cluster, supported by the Digital Research Alliance of Canada. Models are trained for 300 epochs, with each epoch consisting of 500 iterations. This training approach is consistent with Chen et al. [18], in order to compare their baseline model to predictions made using IST data.

Models are evaluated using the average result of holdout k-fold cross validation, and the datasets described in Section III-D. Each run, one train-validation fold is assigned to validation, while the remaining scenes are used for model training. A grid-search approach was used to evaluate key hyperparameters (learning rate, optimizer, batch size, number of convolutional filters, etc.). Different network depths were assessed, but no meaningful improvements from the AutoICE configuration were found. Optimized model specifications are summarized in Table V.

During evaluation, all of folds 1–6 are used as the train-validation set. The test set is fixed in order to evaluate model performance when trained on different class distributions. This process is repeated six times for each architecture, which allows each fold to be assigned to the validation set. Test and validation scores are then averaged across the six folds, which are consistent across the models. While models are evaluated on consistent validation and test scenes, training patch selection is randomized.

D. Evaluation Metrics

For comparison to other studies leveraging the AI4Arctic dataset, all metrics used during the AI4EO AutoICE challenge are reported. These include: the R^2 coefficient (see (1)) for SIC, the F1 score (2) for SOD and FLOE, and a combined score (see (3)), which weights the importance of each parameter for ice analysts.

The coefficient of determination, referred to as R^2 , is given as

$$R^2 = 1 - \left[\sum_{i=1}^N \left((y_i^t - y_i^p)^2 \right) / \left(\sum_{i=1}^N \left((y_i^t - \hat{y}^t)^2 \right) \right) \right] \quad (1)$$

where N denotes the number of pixels, y_i^t denotes the true value of the i th pixel according to the ice chart, \hat{y}^t denotes the mean value over all true pixels containing sea ice, and y_i^p denotes the pixel's predicted SIC class [29]. An R^2 score of one (100%) corresponds to a perfect fit of the data, whereas a score of zero reflects no linear relationship between the target and predicted values.

The F1 score, which is similarly expressed as a percentage, is defined as follows:

$$F1 = (TP) / (TP + 1/2(FP + FN)) \quad (2)$$

where TP denotes the number of true positive pixels, FP denotes the number of false positive pixels, and FN denotes the number of false negative pixels [29]. Here, a high F1 score (around one or 100%) reflects an accurate prediction of the ice chart labels (TP), and minimal misclassification (FP and FN). For a multiclass classification problem, F1 scores are calculated on a per-class basis and averaged across the classes.

Finally, a weighted sum of these metrics, C , is used to generate the combined score [29]

$$C = (2/5) * R_{SIC}^2 + (2/5) * F1_{SOD} + (1/5) * F1_{FLOE}. \quad (3)$$

In addition, we present a SIC accuracy metric, which is used to provide a visual assessment of model performance within the three parent categories of ice formation: open water ($0 \leq \text{SIC} < 20\%$), marginal ice ($20 \leq \text{SIC} \leq 80\%$), and consolidated ice ($\text{SIC} > 80\%$). Previous works, (e.g., Cheng et al. [54]) evaluate the accuracy of SIC labels within $\pm 10\%$ of the true concentration, in order to alleviate errors by ice analysts and account for variability in ice concentration within a polygon. We apply a broader threshold to the marginal ice classes, assessing a model's ability to accurately predict within this category, which is underrepresented in the dataset (see Table II). SICs between 15 and 80% are commonly accepted as thresholds for the MIZ [28], [30], [58]. Because the ice chart labels report concentrations in tenths, 20% is used as the minimum threshold for marginal ice. Concentrations below this threshold are assigned to the open water category. Similarly, we assign concentrations exceeding the marginal ice threshold to the category of consolidated ice. While this terminology is traditionally applied to 100% concentration ice [12], 90% SIC is assigned to this category because it exceeds the marginal ice threshold, and is within $\pm 10\%$ of the true concentration. This metric, described in (4), provides a scene-level comparison of SIC predictions across these three categories

$$\text{Accuracy} = 1/N * \sum_i^N 1(y_i = \hat{y}_i) \quad (4)$$

where a scene's accuracy is calculated by dividing the scene's correct pixels [the sum of ($y_i = \hat{y}_i$) for each pixel i] by its total pixels, N . Prior to this calculation, pixels are assigned to one of the three ice formation classes based on the predicted SIC. These results are compared to the SIC ice chart, after the thresholds are applied. An example of the applied thresholds is given in Fig. 3.

Several reasons motivate this metric's use: 1) this metric is meant to supplement R^2 , providing easier visual analysis; 2)

TABLE VI

VALIDATION AND TEST METRICS AVERAGED OVER THE SIX CROSS-VALIDATION FOLDS, REPORTED AS THE AVERAGE AND STANDARD DEVIATION

VALIDATION SET					
Model	Combined Score	SIC R ² [%]	SIC Acc [%]	SOD F1 [%]	FLOE F1 [%]
AutoICE	80.76±6.47	87.77±4.10	89.79±3.16	74.33±11.61	79.35±7.07
AutoICE-V	81.22±6.37	88.29±4.28	89.78±3.62	75.63±11.07	78.28±7.56
DEU-Net-SV	81.94±7.02	88.97±3.68	90.46±3.09	76.01±13.47	79.95±7.71
DEU-Net-V	81.84±7.03	88.72±4.03	90.55±3.07	76.49±9.60	78.66±7.01
TDU-Net-V	81.83±5.69	87.84±3.62	90.29±3.10	79.79±6.05	75.60±10.80
TEST SET					
Model	Combined Score	SIC R ² [%]	SIC Acc [%]	SOD F1 [%]	FLOE F1 [%]
AutoICE	82.58±1.44	90.33±1.40	84.44±3.15	75.28±3.00	81.69±1.64
AutoICE-V	83.79±1.03	91.37±0.95	85.99±1.00	76.79±2.64	82.39±1.53
DEU-Net-SV	83.40±1.78	91.95±0.93	86.59±2.20	74.82±2.70	83.48±2.73
DEU-Net-V	84.42±0.72	92.53±0.76	88.67±0.56	76.92±1.15	85.13±1.40
TDU-Net-V	83.38±1.86	91.99±0.80	87.19±2.22	74.59±3.56	83.73±2.34

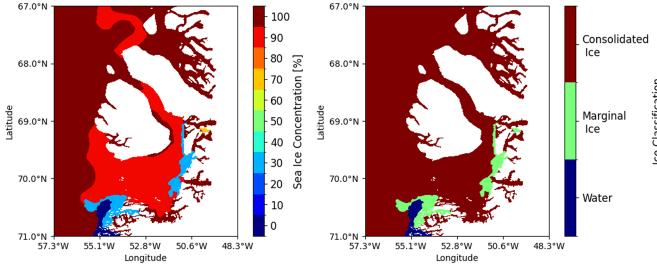


Fig. 3. Example of the SIC thresholds applied to 2019-01-29 on the west coast of Greenland. Comparison of SIC (left) and the three ice formation categories (right).

existing models do not effectively segment marginal ice polygons. This approach allows us to visually assess the general shape of the MIZ, acknowledging that use of total concentration and pixel-level differences between the satellite imagery and charts may cause these discrepancies; 3) this metric penalizes predictions outside of these three classes—something that R^2 does not provide, because it measures relative distance between predicted and target classes.

V. RESULTS

A. Relative Model Performance

Table VI reports the average and standard deviation of each metric for the validation and test sets. Here, the validation set comprises all scenes in folds 1–6, while the test set remains constant. Notably, higher standard deviation is reported across the validation set for all models, with the largest standard deviation across scores being reported for SOD and floe size. Referring to Table III, these parameters have the folds containing the least data, and the largest variation in percentage of unmasked data (15.82–46.87% for SOD and 22.77–42.08% for FLOE, relative to 40.74–58.98% for SIC). Further, the percentage of masked data across parameters differs at the scene level (e.g., Figs. 4–7). This discrepancy results from AI4Arctic’s definition

of a dominant ice type, where there must be a dominant partial concentration for SOD and FLOE labels to be assigned [20]. Thus, classification errors in folds containing minimal unmasked data result in low, outlying scores, which skew the reported average and standard deviation.

In comparison, test set scores report more consistent standard deviations across parameters, in line with comparable data representation of 57.75% unmasked SIC pixels, 46.87% unmasked SOD pixels, and 41.12% unmasked FLOE pixels. Some variability in model predictions for both validation and test can be attributed to different class distributions across train sets, in addition to training on random scene patches.

Comparing results, by metric, for the test set, the greatest overall performance improvement is found using the DEU-Net-V architecture, which improves predictive accuracy by 2.44% (SIC), 5.01% (SIC accuracy), 2.18% (SOD), 4.21% (FLOE), and 2.23% (combined score). The greatest improvement is seen in SIC accuracy, which does not impact the combined score. The following sections compare a subset of test set examples, illustrating these improvements.

B. Comparison of Predicted Sea Ice Maps

Figs. 4–7 provide a visual comparison of predicted sea ice maps for the five architectures, across varied regions and seasons. Fig. 4 compares all five architectures across the three sea ice parameters. Fig. 6 directly assesses the impacts of IST data on model predictions by comparing AutoICE and AutoICE-V, then comparing to DEU-Net-V as the best performing model. Finally, Fig. 7 highlights a case where use of task-specific decoders improves model predictions. These predictions leverage coincident remote sensing data retrieved on the scene date. The predicted sea ice maps, generated at inference-time, are compared to AI4Arctic ice charts. Example input channels for each scene are provided to aid visual analysis.

1) *Performance Across Models*: Fig. 4 illustrates mid-season freeze-up conditions off the coast of Labrador, Canada. The chart depicts mixed polygons of marginal and consolidated ice,

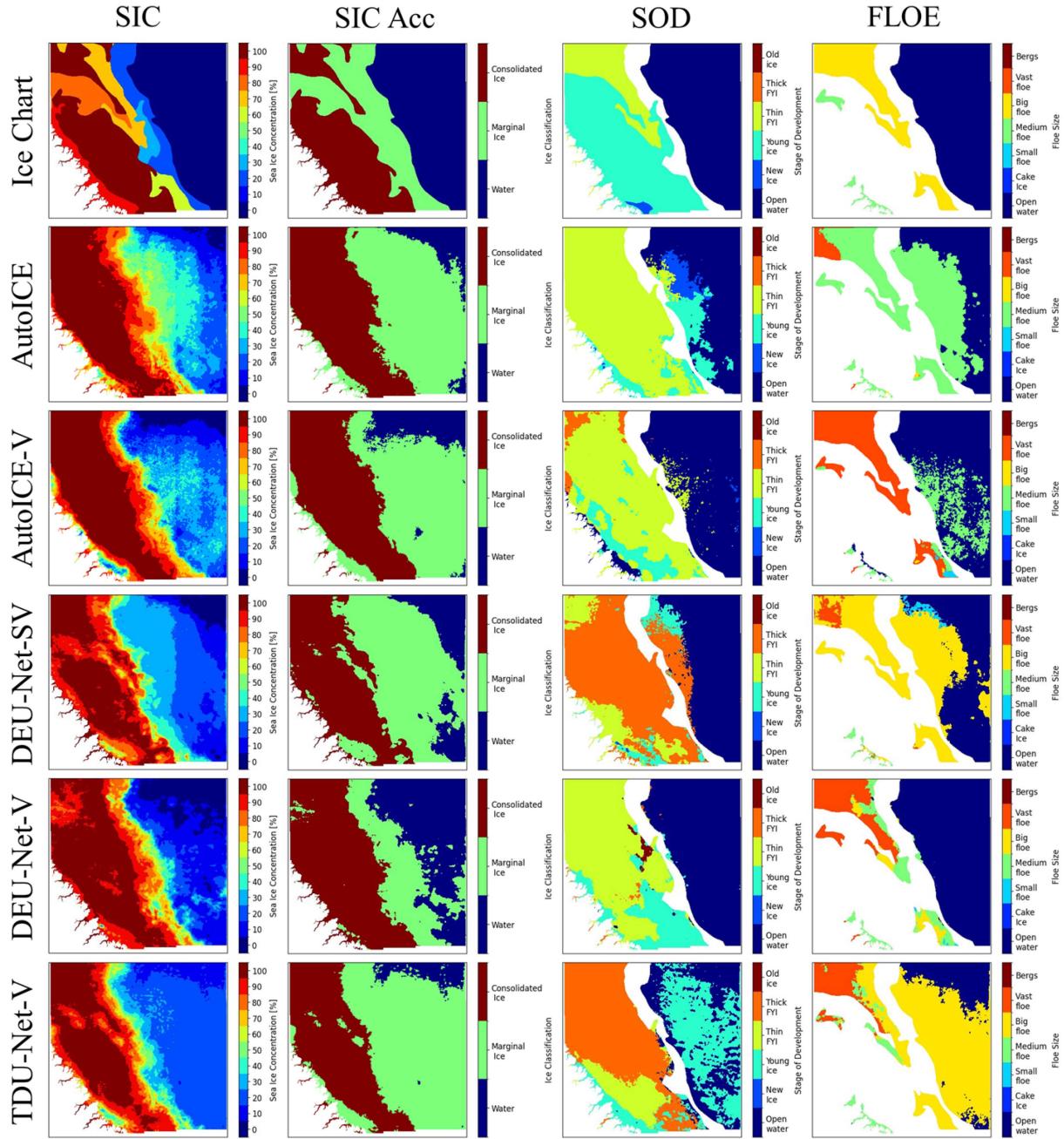


Fig. 4. Summary of model predictions for all architectures, compared for 2020-01-24 off the coast of Labrador.

in addition to an open water region. Referring to the input channels (see Fig. 5), the SAR imagery records open water with comparable-to-higher backscatter measurements than the ice, which can result from low incidence angle or wind roughening. The associated BT measurements exhibit noise over water, likely due to atmospheric moisture or similar wind roughening. These ambiguous features result in classification errors across all models—predominantly in the prediction of open water.

While IST data is largely masked in the open water region, the unmasked measurements exceeding 271 K do positively impact predictions, as seen in the more accurate DEU-Net-V SIC map. Compared to AutoICE-V and DEU-Net-SV, in regions where

IST data is largely masked, features learned in combination with ambiguous backscatter signatures yield lower accuracy. Additionally, relative to the AutoICE baseline, IST data is seen to improve predictions for floe size, where big floe is underestimated.

DEU-Net-SV and TDU-Net-V both provide better approximations of the region of 70 and 80% concentration, although these features are largely predicted as 90% consolidated ice. In addition, these improvements are in conjunction with greater overestimation of marginal ice, similarly to the predictions by both AutoICE architectures. For DEU-Net-SV, this region is differentiable from consolidated ice in both SAR and

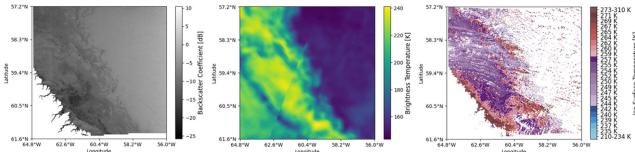


Fig. 5. HH SAR (left), 36.5 GHZ H BT (middle), and IST (right) input channels for 2020-01-24. Note: certain AI4Arctic BT scenes are a mosaic of multiple AMSR2 swaths, with pixels selected based on minimizing acquisition time with Sentinel-1 [20].

PM measurements. Thus, this fusion approach is capable of capturing features that are less evident in the IST data. However, overestimation of ice concentration in the open water region demonstrates the benefits of learning IST features independently from SAR and PM measurements. These findings are supported by the input-level fusion approaches, which report the highest concentrations in the open water region.

Comparing the single and task-specific decoders, DEU-Net-SV's misclassification of open water for SOD and FLOE relates to the misclassification of open water as 30% SIC. Similarly, DEU-Net-V's more accurate classification of 0% concentration improves SOD and FLOE predictions. Across the AutoICE and DEU-Net architectures, misclassification of open water is largely consistent across the three sea ice parameters. Conversely, TDU-Net-V inconsistently predicts open water, and reports potentially incompatible predictions of low concentration ice, young ice, and big floe in the same region.

Of the five models, DEU-Net-V achieves the best SOD prediction, although no model accurately segments the thin FYI feature. In addition, while DEU-Net-V misclassifies the central region of big floe, predictions from other folds, which are trained on different class distributions, do not predict this error. This variability is an example of sensitivity to the pixel class distribution, which contributes to the standard deviations reported in Table VI.

2) *Impacts of Input-Level Fusion:* Fig. 6 depicts early season melt off the west coast of Greenland. This chart is characterized by a region of full ice cover, in addition to an inland region surrounding Disko Island that features mixed marginal and consolidated ice conditions, and open water.

Comparing predicted SIC maps for input-level fusion, AutoICE-V more accurately models the consolidated ice region south of Disko Island, in addition to reducing misclassification of marginal ice north of Disko Island. For SOD, AutoICE-V reduces misclassification of thin FYI in the same region. Neither model correctly predicts the inland regions as thin FYI; however, use of IST data produces a more accurate prediction of the thick FYI region. Finally, comparing floe size, the models achieve comparable accuracy on open water and vast flow. AutoICE accurately predicts the region south of Disko Island as big floe whereas AutoICE-V more accurately predicts the same ice condition in the region between Disko Island and the peninsula.

Referring to the input channels (see Fig. 6), the region of 90% concentration south of Disko Island registers a backscatter signature that is challenging for models to correctly interpret.

Here, regions of smooth ice appear similarly to the scene's smaller regions of open water, and the entire feature registers as darker than the large regions of open water and consolidated ice. BT measurements distinguish this region from open water and consolidated ice; however, IST data in combination with BT further reduces ambiguity. Notably, both models appear to correlate SIC and FLOE predictions. For AutoICE-V, misclassification as vast floe is consistent with the classification of 90% SIC. Similarly for AutoICE, correct classification of big floe is more consistent with the model's misclassification of 90% SIC as 70%–80%. Finally, comparing to the SAR imagery, the resolutions of the VIIRS and AMSR2 products are too coarse to improve predictions of ice conditions in the tributaries. Overall, fusion of IST at the input-level appears to improve predictions most significantly over regions with homogeneous conditions (e.g., in the case of SOD).

Further, comparing DEU-Net-V to AutoICE-V, feature-level fusion of IST data is found to more accurately predict the region of consolidated ice south of Disko Island, in addition to the two, small open water regions north and south of the island. In addition, this architecture reduces misclassification of thin FYI above Disko Island, and vast floe in the southern region of consolidated ice. However, AutoICE-V more correctly predicts big floe between Disko Island and the peninsula, whereas DEU-Net-V classifies the region as vast floe, in line with its overestimation of SIC. Both models report similar predictions to the AutoICE baseline in the tributaries; this is expected, as the spatial resolution of the IST data does not accurately capture this region.

3) *Impacts of Task-Specific Decoders:* Fig. 7 illustrates early melt conditions on the west coast of Greenland. The scene is characterized by a large region of consolidated ice, followed by a MIZ and a region of open water, which is expanding with melt along the ice edge. Example input channels, included in Fig. 7, illustrate noise over the open water region in the AMSR2 data, as well as distinct SAR backscatter signatures, which differentiate thin and thick FYI from old ice. While much of the open water IST data is masked, there is a strong delineation between the 100% and 80% concentrations at the boundary of the MIZ.

Comparing TDU-Net-V to the DEU-Net-V architecture, this model is more susceptible to noise in the BT measurements, resulting in incorrect predictions of 10% SIC. However, TDU-Net-V improves predictions of thin and thick FYI (SOD), as well as small floes (FLOE). Here, task specific decoders are found to improve predictions where input channel features (e.g., the SAR backscatter signature) coincide with an individual sea ice parameter. Specifically, a homogeneous region for one parameter is composed of multiple polygons, which may report mixed conditions for another (e.g., comparing SIC and SOD). In these cases, the separate decoders are able to reconstruct different representations based on common features extracted by the shared encoders. However, this architecture is unable to improve the segmentation of SIC polygons in the MIZ—a task all models struggle with in this cross validation fold. Further, because much of this zone is masked for SOD and FLOE, we cannot comment on if this trend persists when labels are available for all three parameters.

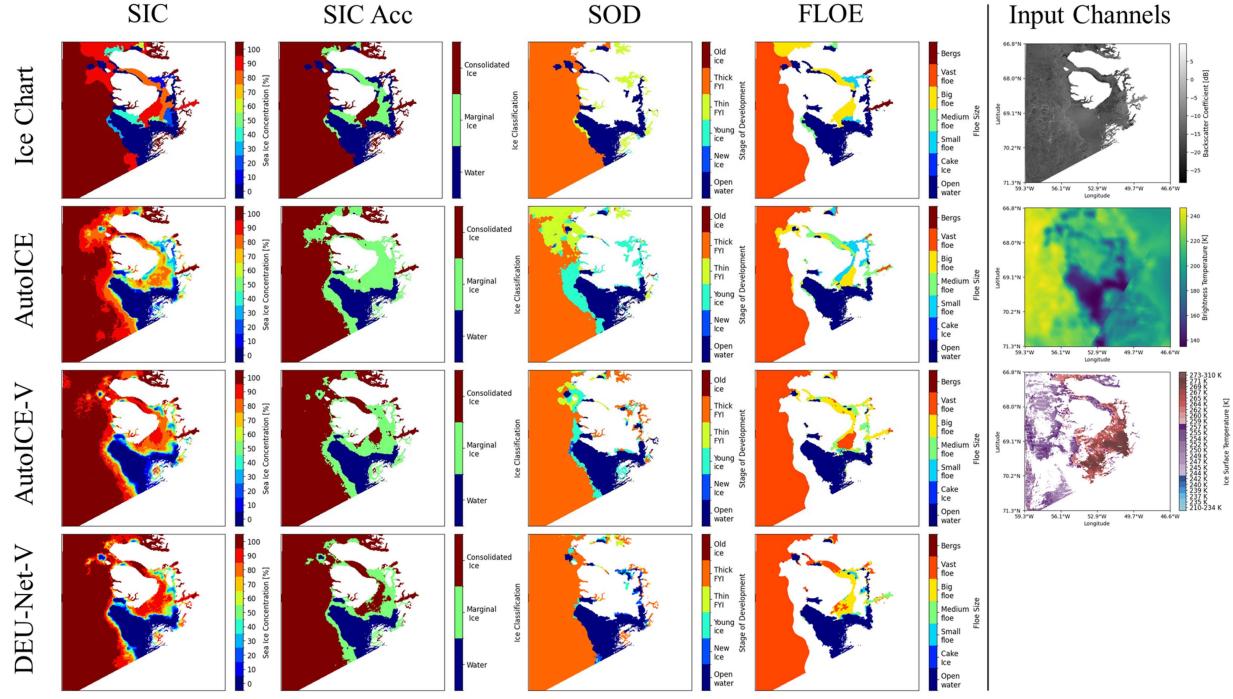


Fig. 6. Summary of AutoICE, AutoICE-V, and DEU-Net-V model predictions for 2021-04-19 on the west coast of Greenland. The SIC, SIC Acc, SOD, and FLOE columns display model predictions. Associated HH SAR (top), 36.5 GHZ H BT (middle), and IST (bottom) input channels are provided on the right for comparison with the ice chart labels.

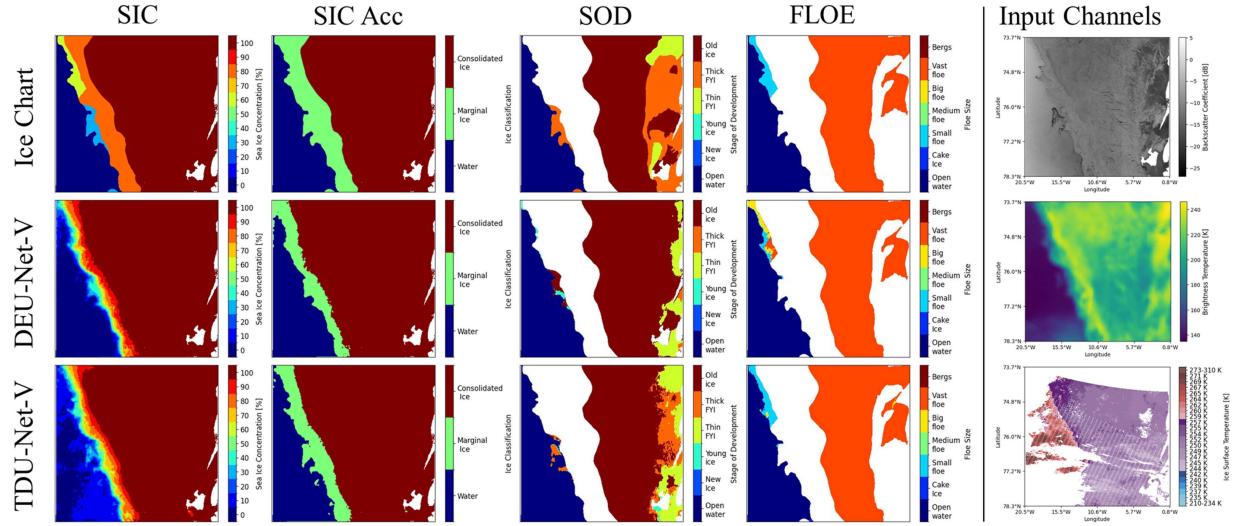


Fig. 7. Summary of DEU-Net-V and TDU-Net-V model predictions for 2021-03-01 off the west coast of Greenland. The SIC, SIC Acc, SOD, and FLOE columns display model predictions. Associated HH SAR (top), 36.5 GHZ H BT (middle), and IST (bottom) input channels are provided on the right for comparison with the ice chart labels.

VI. DISCUSSION

A. Impacts of Fusion Approaches

Overall, learning the salient features of IST data using the DEU-Net-V architecture best reduces classification errors resulting from ambiguous SAR and PM measurements. Across all models, IST data used in combination with BT measurements reduces the impacts of ambiguous SAR textures. Further, IST

measurements consistently improve the classification of open water where incidence angle impacts backscatter measurements, atmospheric moisture introduces noise in BT readings, and where wind roughening causes ambiguity in ice–water classification for both sensors.

Comparing the dual-encoder architectures, complimentary features learned between SAR and IST data (DEU-Net-SV) result in higher rates of misclassification when IST data is

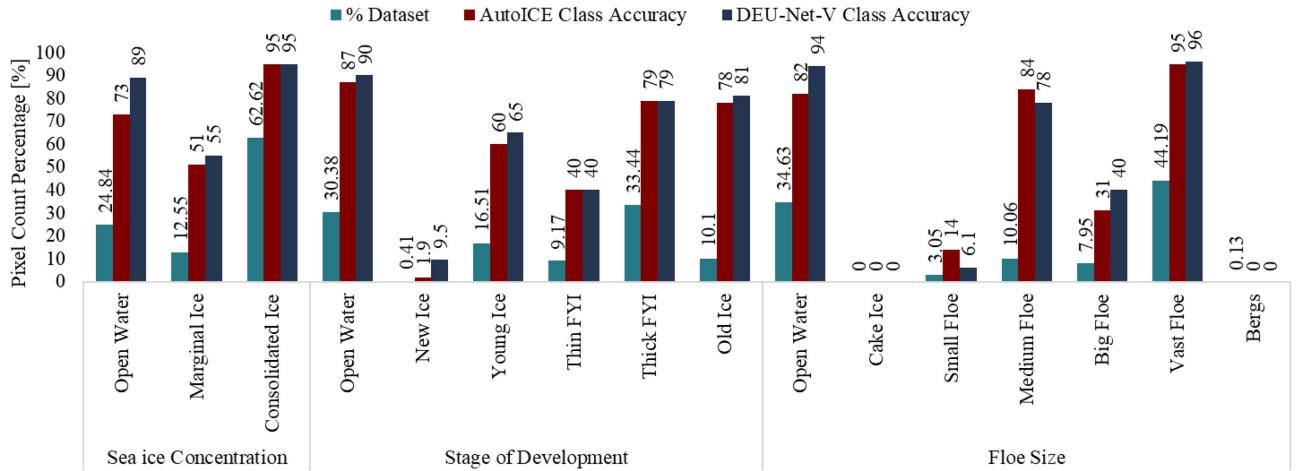


Fig. 8. Accuracy of AutoICE and DEU-Net-V predicting the correct class of SIC, SOD, and floe size, compared to the representation of each class in the test dataset.

masked and SAR measurements are challenging to interpret. Thus, learning complimentary SAR and PM features, where BT effectively distinguishes between ice and water, improves predictions. Conversely, salient features learned from IST data are able to reduce SAR and PM ambiguities where data is present. Overall, masked data and spatial resolution—impacting small floes and tributaries—constrain the potential for IST to improve accuracy. A single, shared decoder improves contextual awareness, improving the classification of shared classes; specifically, open water. This trend is consistent across single, and dual-encoder architectures, including the AutoICE baseline. Conversely, TDU-Net-V is found to predict incompatible conditions across the three parameters (e.g., young ice and big floes in the same region, as in Fig. 4). However, task-specific decoders, used in TDU-Net-V, are effective at reconstructing features that are specific to a single sea ice parameter. These findings result in an architecture that outperforms the AutoICE baseline, but is ultimately less competitive than the other models incorporating IST data.

Finally, the AutoICE-V architecture, presented in de Loë et al. [43], continues to achieve competitive performance on the expanded VIIRS-AI4Arctic dataset. Use of IST data in this fusion approach improves predictions relative to the AutoICE baseline, reducing overestimation of open water as low concentration ice (see Fig. 4), and improving underestimation of consolidated ice (see Fig. 6). However, feature-level fusion of IST data is found to better mitigate classification errors overall.

B. DEU-Net-V Performance

In addition, class-wise accuracy scores are compared between the best performing model, DEU-Net-V, and the AutoICE baseline. In Fig. 8, accuracy scores are reported for the three sea ice parameters over the test set's six cross-validation runs. These scores are also compared to the pixel class distribution in the test set, excluding masked data (e.g., land pixels). Across all parameters, there is a correlation between classes with

higher representation—consistent across datasets (see Table II)—achieving higher accuracy scores. Further, model predictions, which exhibit minor variability across cross-validation folds, are most consistent in classifying consolidated ice, which is the most represented class.

1) *SIC*: For SIC, we choose to compare the three ice formation categories in place of the 11 discrete concentrations. This choice results from low representation of the 10%–80% classes, accounting for between 0.19%–4.2% of unmasked test set pixels. Consequently, class-wise accuracy scores for these concentrations are disproportionately affected by classification errors, similar to cross-validation folds with significant masked data.

Overall, use of IST data and the DEU-Net-V architecture improves classification of open water and marginal ice, while both models predict consolidated ice with comparable accuracy. Referring to Fig. 8, improvements in predicting the open water and consolidated ice classes coincide with reduced misclassifications of marginal ice. These results demonstrate significant improvements in the classification of open-water regions where SAR and PM data is challenging to interpret, or obstructed by atmospheric interference.

Further, the models overwhelmingly misclassify marginal ice as consolidated ice; 42% of AutoICE misclassifications are attributed to consolidated ice, compared to 38% for DEU-Net-V. These errors predominantly occur with 70% and 80% concentration labels, supporting the conclusion that the over-representation of consolidated ice skews predictions toward these classes.¹ Ultimately, we do not anticipate significant improvements in segmenting MIZ polygons due to pixel-level differences in total concentration, and uncertainty in ice chart labels for marginal concentrations.

2) *SOD*: The dual-encoder architecture consistently outperforms class-wise accuracy scores for SOD. Referring to Fig. 8,

¹For AutoICE, 79% of 70% SIC pixels are misclassified as consolidated ice, and 64% for 80% SIC. These statistics are 76% and 52%, respectively, for DEU-Net-V.

accuracy is improved for all classes, excluding identical scores for thin and thick FYI, where the IST product retrieves similar measurements. This lack of improvement may also result from lower dataset representation, and learned ice type relationships between SOD, SIC, and FLOE. Alternatively, it may reflect real-world conditions if both ice types are represented in a region but only the majority ice type (e.g., thin FYI) is assigned as the polygon label.

The model continues to outperform AutoICE on the open water class, supporting findings that IST measurements improve classification where SAR and PM channels yield ambiguous or inaccurate surface measurements. In addition, cases such as Fig. 4 demonstrate improvements in the prediction of land-fast ice, where PM measurements are inaccurate. While both models are poor predictors of new ice, which is the least represented class, DEU-Net-V still increases correct classification of this ice type.

3) *Floe Size*: DEU-Net-V improves class-wise accuracy scores for open water, as well as medium, big, and vast floes (see Fig. 8), resulting in the performance improvements seen in Table VI. The AutoICE baseline does outperform this architecture for small and medium floes; this finding is attributed to the spatial resolution of the VIIRS product. Specifically, small floes are 20–100 m across, while medium floes range from 100 to 500 m—both of which require a finer resolution than the IST product [12]. While alternative fusion architectures (e.g., DEU-Net-SV) may improve scores for these classes, this architecture still achieves the highest accuracy across all parameters. In addition, because small floes are largely misclassified as medium floes (72% for AutoICE versus 75% for DEU-Net-V), this issue may be partially attributed to low dataset representation and learned relationships between parameters, similar to thin FYI.

C. Applications of SIC Accuracy

Comparison of SIC predictions using the categories of open water, marginal ice, and consolidated ice (see Section IV-D), provides both an effective visual tool for comparing model predictions, and captures misclassifications outside of the label’s category of ice formation. As demonstrated in Fig. 6 with consolidated ice, use of these thresholds improves the ability to identify and interpret classification errors. This method can be easily applied to other studies, because the thresholds are applied to the predicted concentrations following model evaluation.

In addition, we found that not penalizing predictions within the marginal ice category is useful for evaluating models when datasets have limited representation of these classes. Ultimately, misclassifications within underrepresented categories minimally impact scores. Comparing to existing approaches, we argue that these broad categories are of greater interest to analysts, because $\pm 10\%$ does not result in a decrease in accuracy if the misclassification is an adjacent class, but different category of ice formation.

Notably, the metric reports lower test accuracy (see Table VI) than R^2 , highlighting misclassifications of open water as marginal ice, and marginal ice as consolidated ice. These misclassifications minimally impact R^2 , because classifications

is adjacent and nearby classes results in lower penalties. Conversely, the validation set reports a higher accuracy; from analysis of the train-validation set, this is attributed to evaluation over a larger dataset, featuring a smaller percentage of scenes where input channel data results in misclassification of open water. Further, this metric does not focus on misclassifications within the MIZ; variability in mid-range concentration estimates by ice analysts [54], and use of polygon labels at the pixel-level, may both contribute to why models less accurately segment these polygons (e.g., Fig. 7). This forgiveness may be useful for considering real-world, pixel-level conditions. Specifically, ice chart polygons are assigned a label based on the majority ice type, and overestimation of concentration is often linked to older/vaster SOD and FLOE. Thus, the ice type associated with the dominant concentration may skew SIC estimates to be higher than the total concentration, which is the sum of partial concentrations. Ultimately, SIC accuracy is a useful metric for evaluation, which should be used in combination with the R^2 score to interpret predictions, and highlight recurring classification errors.

D. Considerations for Incorporating Thermal-Infrared Imagery

This study highlights the potential benefits of incorporating thermal-infrared imagery in deep learning-based sea ice mapping approaches. An 8 h acquisition window found ample overlap between VIIRS, Sentinel-1, and AMSR2 coverage, with models generating accurate pixel-level ice chart predictions in combination with downscaled SAR imagery. In addition, improvements in open-water predictions, where this data type offers the most benefit, are less impacted by ice drift velocity. Ultimately, the primary limitation for using this data type is pixel cloud contamination, where filtering at the defined 60% cloud cover threshold restricted dataset size and spatial-temporal coverage. This issue is compounded by VIIRS cloud masking, which is more conservative at night in response to increased uncertainty [56].

Because this study sought to evaluate the impacts of IST data, scenes with extensive cloud cover were discarded; however, expansion of the dataset in future work can allow a higher cloud threshold. In this case, the model would function similarly to current products (e.g., Ludwig et al. [9]), which use data from passive instruments where available. Further, upcoming satellite missions offer promising advancements in thermal-infrared products, which may reduce these limitations. For example, the TRISHNA satellite, set to launch in 2026, will provide a spatial resolution of 60 m and global revisit period of three days [59]. This mission provides an alternative, high-resolution product to the existing SDGSAT-1, which has a spatial resolution of 30 m, but is limited by its 11 d revisit period [60]. Higher-resolution thermal-infrared data could enhance sea ice parameter estimation near land boundaries, within inland features such as tributaries, and capture small and medium floes. Further, improved spatial resolution may also increase data retrieval within cloudy regions. Although these products will not overlap with the AI4Arctic dataset, they could complement existing

algorithms for estimating SIC, such as those by Ludwig et al. [9] and König et al. [10]. In addition, combining higher-resolution products with SAR imagery can refine pixel-level labels of ice conditions, enabling advancements beyond ice chart labels and supporting the exploration of semi-supervised or transfer learning approaches [61], [62].

Finally, previous research reports variability in multitask predictions for the open-water class, which should be consistent across SIC, SOD, and FLOE [63]. Building on the DEU-Net-V architecture, use of a second decoder, which accomplishes an additional ice–water classification task, is one potential method for improving model accuracy. A dedicated decoder would allow the model to focus on distinguishing between the two classes. Further, in learning shared features that are useful for both ice–water classification and sea ice parameter estimation, the model may learn better feature representations that improve model accuracy and generalization. Alternatively, the second decoder could classify open-water, marginal ice, and consolidated ice, in line with the SIC accuracy classes. In addition to improving ice–water classification, this approach has the potential to reduce misclassification of marginal ice as consolidated ice. Exploration of these alternative architectures have the capability to further leverage the improvements seen from incorporating IST measurements.

VII. CONCLUSION

This study investigated the impacts of applying thermal infrared data to multitask sea ice parameter estimation, using input- and feature-level fusion approaches. Ultimately, DEU-Net-V, a novel dual-encoder U-Net architecture, was found to reduce classification errors resulting from ambiguous or inaccurate surface measurements from SAR and PM sensors. Learning the salient features of VIIRS IST measurements separately from Sentinel-1 SAR, AMSR2 PM, and ERA5 reanalysis data, was found to improve predictive accuracy relative to networks learning the combined features of all inputs, or grouping input channels based on resolution.

Analysis leveraged evaluation metrics and predicted sea ice maps of SIC, SOD, and floe size for an 84-scene subset of the AI4Arctic dataset. These maps were compared to ice charts from the Canadian and Greenland Ice Services, acting as labels. Four models were evaluated against the winning AI4EO AutoICE Challenge architecture, which acts as an AI4Arctic baseline. A SIC accuracy metric was also introduced to provide an additional assessment of model performance.

The study found that while learning common features between SAR and IST channels did improve model predictions, regions where SAR data is ambiguous and IST measurements are masked due to cloud contamination negated the predictive gains. In addition, comparison of architectures leveraging shared versus task specific decoders demonstrated that architectures with a single decoder, such as DEU-Net-V, benefited from improved contextual awareness across parameters. Conversely, multidecoder architectures inconsistently predict shared classes, but can effectively reconstruct task-specific features. Results demonstrate that models incorporating IST consistently reduce

classification errors across all three tasks, particularly when identifying open water under conditions with low-incidence angle (SAR), high atmospheric moisture (PM), and wind roughening (SAR and PM). Further, inference, which evaluated models on the months of November through May, demonstrated these improvements under a range of freeze-up and break-up/early melt conditions. These findings support the continued development of deep learning approaches incorporating IST, and highlight the potential for next-generation thermal–infrared instruments to further improve automated sea ice mapping.

ACKNOWLEDGMENT

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Waterloo for the support.

REFERENCES

- [1] W. Meier et al., “Sea ice: Arctic report card 2023,” National Oceanic and Atmospheric Administration, Tech. Rep., 2023.
- [2] G. Boutin, T. Williams, P. Rampal, E. Olason, and C. Lique, “Impact of wave-induced sea ice fragmentation on sea ice dynamics in the MIZ,” Copernicus GmbH, Tech. Rep., 2020.
- [3] N. Melia, K. Haines, and E. Hawkins, “Sea ice decline and 21st century trans-arctic shipping routes,” *Geophysical Res. Lett.*, vol. 43, no. 18, pp. 9720–9728, 2016.
- [4] L. Beaulieu et al., “Indigenous self-determination in cryospheric science: The Inuit-led Sikumik Qaujimajjutit (“tools to know how the ice is”) program in Inuit Nunangat, Canada,” *Front. Earth Sci.*, vol. 11, 2023. [Online]. Available: <https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2023.1076774>
- [5] M. Rantanen et al., “The arctic has warmed nearly four times faster than the globe since 1979,” *Commun. Earth Environ.*, vol. 3, 2022.
- [6] J. Eamer et al., “Life linked to ice: A guide to sea-ice-associated biodiversity in this time of rapid change,” Conservation of Arctic Flora Fauna, Iceland, Tech. Rep. 10, 2013.
- [7] Y. Liu, J. R. Key, X. Wang, and M. Tschudi, “Multidecadal Arctic sea ice thickness and volume derived from ice age,” *Cryosphere*, vol. 14, pp. 1325–1345, 2020.
- [8] C. Melsheimer and G. Spreen, “AMSR2 ASI Sea Ice Concentration Data, Arctic, version 5.4 (NetCDF) (july 2012 - december 2018),” 2019.
- [9] V. Ludwig, G. Spreen, and L. T. Pedersen, “Evaluation of a new merged sea-ice concentration dataset at 1 km resolution from thermal infrared and passive microwave satellite data in the arctic,” *Remote Sens.*, vol. 12, no. 19, 2020.
- [10] C. König, T. König, S. Singha, A. Frost, and S. Jacobsen, “Combined use of space borne optical and SAR data to improve knowledge about sea ice for shipping,” *Remote Sens.*, vol. 13, no. 23, 2021.
- [11] A. S. Komarov, A. Caya, L. Pogson, and M. Buehner, “Assimilation of RCM data in the canadian ice concentration analysis system,” *Remote Sens. Environ.*, vol. 306, 2024, Art. no. 114113. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003442572400124X>
- [12] Environment Canada, “Manual of standard procedures for observing and reporting ice conditions (MANICE),” 2005, issuing authority: Assistant Deputy Minister, Meteorological Service of Canada. [Online]. Available: <https://www.Canada.ca/en/environment-climate-change/services/weather-manuals-documentation/manice-manual-of-ice.html>
- [13] D. Malmgren-Hansen et al., “A convolutional neural network architecture for sentinel-1 and AMSR2 data fusion,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1890–1902, Mar. 2021.
- [14] H. Boulze, A. Korosov, and J. Brajard, “Classification of sea ice types in sentinel-1 SAR data using convolutional neural networks,” *Remote Sens.*, vol. 12, no. 13, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/13/2165>
- [15] K. Radhakrishnan, K. A. Scott, and D. A. Clausi, “Sea ice concentration estimation: Using passive microwave and SAR data with a U-net and curriculum learning,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5339–5351, 2021.

- [16] Y. Wang and X. Li, "Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning," *Earth Syst. Sci. Data*, vol. 13, no. 6, pp. 2723–2742, 2021.
- [17] A. Stokholm, T. Wulff, A. Kucik, R. Saldo, J. Buus-Hinkler, and S. M. Hvidegaard, "AI4SeaIce: Toward solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [18] X. Chen et al., "MMSeaIce: Multi-task mapping of sea ice parameters from AI4Arctic sea ice challenge dataset," *EGUphere*, vol. 2023, pp. 1–17, 2023.
- [19] M. Rogers, M. Fox, A. Fleming, L. van Zeeland, J. Wilkinson, and J. Hosking, "Sea ice detection using concurrent multispectral and synthetic aperture radar imagery," *Remote Sens. Environ.*, vol. 305, 2024, Art. no. 114073.
- [20] J. Buus-Hinkler et al., "Ready-to-train AI4Arctic sea ice challenge dataset," 2022.
- [21] S. C. Deodoro, R. de Andrade Moral, R. Fealy, T. McCarthy, and R. Fealy, "Using the surface scattering mechanism from dual-pol SAR data to estimate topsoil particle-sizefractions," *Int. J. Appl. Earth Observation Geoinformation*, vol. 128, 2024, Art. no. 103742. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1569843224000967>
- [22] J. Chi, H. cheol Kim, S. Lee, and M. M. Crawford, "Deep learning based retrieval algorithm for arctic sea ice concentration from AMSR2 passive microwave and MODIS optical data," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111204.
- [23] T. Feng, X. Liu, and R. Li, "Super-resolution-aided sea ice concentration estimation from AMSR2 images by encoder–decoder networks with atrous convolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 962–973, 2023.
- [24] Q. Liu, R. Zhang, Y. Wang, H. Yan, J. Xu, and Y. Guo, "Short-term forecasting of sea ice thickness based on PredRNN," *J. Physics: Conf. Ser.*, vol. 2486, no. 1, May 2023, Art. no. 0 12017.
- [25] J. C. Comiso, "Sea ice effective microwave emissivities from satellite passive microwave and infrared observations," *J. Geophysical Research: Oceans*, vol. 88, no. C12, pp. 7686–7704, 1983. [Online]. Available: <https://agupubs.onlinelibrary-wiley-com.proxy.lib.uwaterloo.ca/doi/abs/10.1029/JC088iC12p07686>
- [26] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorological Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [27] J. C. Comiso, W. N. Meier, and R. Gersten, "Variability and trends in the arctic sea ice cover: Results from different techniques," *J. Geophysical Res.: Oceans*, vol. 122, no. 8, pp. 6883–6900, 2017.
- [28] C. Strong, D. Foster, E. Cherkaev, I. Eisenman, and K. M. Golden, "On the definition of marginal ice zone width," *J. Atmospheric Ocean. Technol.*, vol. 34, no. 7, pp. 1565–1584, 2017.
- [29] A. R. Stokholm et al., "The autoice challenge," *EGUphere*, vol. 2023, pp. 1–28, 2023.
- [30] A. W. Bateson et al., "Sea ice floe size: Its impact on pan-arctic and local ice mass and required model complexity," *Cryosphere*, vol. 16, no. 6, pp. 2565–2593, 2022. [Online]. Available: <https://tc.copernicus.org/articles/16/2565/2022/>
- [31] A. W. Bateson, D. L. Feltham, D. Schröder, L. Hosekova, J. K. Ridley, and Y. Aksenov, "Impact of sea ice floe size distribution on seasonal fragmentation and melt of arctic sea ice," *Cryosphere*, vol. 14, no. 2, pp. 403–428, 2020. [Online]. Available: <https://tc.copernicus.org/articles/14/403/2020/>
- [32] K. A. Scott, M. Buehner, A. Caya, and T. Carrieres, "Direct assimilation of amsr-e brightness temperatures for estimating sea ice concentration," *Monthly Weather Rev.*, vol. 140, no. 3, pp. 997–1013, 2012. [Online]. Available: <https://journals.ametsoc.org/view/journals/mwre/140/3/mwr-d-11-00014.1.xml>
- [33] G. J. Macdonald, R. K. Scharien, K. Duncan, S. L. Farrell, P. Rezania, and A. Tavri, "Arctic sea ice topography information from RADARSAT constellation mission (RCM) synthetic aperture radar (SAR) backscatter," *Geophysical Res. Lett.*, vol. 51, no. 4, 2024, Art. no. e2023GL107261.
- [34] L. Pogson, T. Geldsetzer, M. Buehner, T. Carrieres, M. Ross, and K. A. Scott, "Collecting empirically derived SAR characteristic values over one year of sea ice environments for use in data assimilation," *Monthly Weather Rev.*, vol. 145, no. 1, pp. 323–334, 2017. [Online]. Available: <https://journals.ametsoc.org/view/journals/mwre/145/1/mwr-d-16-0110.1.xml>
- [35] W. N. Meier and J. S. Stewart, "Assessing the potential of enhanced resolution gridded passive microwave brightness temperatures for retrieval of sea ice parameters," *Remote Sens.*, vol. 12, no. 16, 2020.
- [36] N. Ivanova et al., "Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges and optimal approach using passive microwave observations," *Cryosphere*, vol. 9, no. 5, pp. 1797–1817, 2015.
- [37] K. Naoki, J. Ukita, F. Nishio, M. Nakayama, J. C. Comiso, and A. Gasiewski, "Thin sea ice thickness as inferred from passive microwave and in situ observations," *J. Geophysical Res.: Oceans*, vol. 113, no. C2, 2008.
- [38] X. Chen, R. Valencia, A. Soleymani, and K. A. Scott, "Predicting sea ice concentration with uncertainty quantification using passive microwave and reanalysis data: A case study in baffin bay," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, 2023.
- [39] M. Leppäranta and J. E. Lewis, "Observations of ice surface temperature and thickness in the baltic sea," *Int. J. Remote Sens.*, vol. 28, no. 17, pp. 3963–3977, 2007. [Online]. Available: <https://doi.org/10.1080/01431160601075616>
- [40] T. Zeng, L. Shi, M. Mäkinen, and J. Zhang, "Sea ice thickness analyses for the Bohai sea using MODIS thermal infrared imagery," *Acta Oceanologica Sinica*, vol. 35, no. 6, pp. 96–104, May 2016. [Online]. Available: <https://doi.org/10.1007/s13131-016-0908-8>
- [41] D. K. Hall and G. A. Riggs, "MODIS/aqua sea ice extent 5-min L2 swath 1 km, version 61," 2021.
- [42] G. Riggs, M. Tschudi, and D. K. Hall, "VIIRS/NPP ice surface temperature 6-min L2 swath 750 m, version 1," 2017.
- [43] L. de Loë, L. Xu, and K. A. Scott, "Fusing ice surface temperature with the AI4Arctic dataset for enhanced sea ice concentration estimation: A preliminary assessment," in *Proc. OCEANS 2024-Halifax*. IEEE, 2024, pp. 1–9.
- [44] L. Wang, K. A. Scott, L. Xu, and D. A. Clausi, "Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4524–4533, Aug. 2016.
- [45] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.
- [46] A. Kucik and A. Stokholm, "AI4SeaIce: Comparing loss representations for SAR sea ice concentration charting," in *Proc. ICLR Workshop AI Earth Sci.*, 2022. [Online]. Available: <https://ai4earthscience.github.io/iclr-2022-workshop/accepted>
- [47] A. Kucik and A. Stokholm, "AI4SeaIce: Selecting loss functions for automated SAR sea ice concentration charting," *Sci. Rep.*, vol. 13, 2023.
- [48] A. Stokholm, A. Kucik, N. Longépé, and S. M. Hvidegaard, "AI4SeaIce: Task separation and multistage inference CNNs for automatic sea ice concentration charting," *EGUphere*, vol. 2023, pp. 1–25, 2023.
- [49] A. Mahbod, G. Schaefer, G. Dorffner, S. Hatamikia, R. Ecker, and I. Ellinger, "A dual decoder U-net-based model for nuclei instance segmentation in hematoxylin and eosin-stained histological images," *Front. Med.*, vol. 9, 2022.
- [50] L. M. Pereira, A. Salazar, and L. Vergara, "A comparative study on recent automatic data fusion methods," *Computers*, vol. 13, no. 1, 2024.
- [51] L. Wang, K. A. Scott, and D. A. Clausi, "Improved sea ice concentration estimation through fusing classified SAR imagery and AMSR-E data," *Can. J. Remote Sens.*, vol. 42, no. 1, pp. 41–52, 2016. [Online]. Available: <https://doi.org/10.1080/07038992.2016.1152547>
- [52] A. Korosov, D. Demchev, N. Miranda, N. Franceschi, and J.-W. Park, "Thermal denoising of cross-polarized sentinel-1 data in interferometric and extra wide swath modes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.
- [53] C. Soriot, C. Prigent, C. Jimenez, and F. Frappart, "Arctic sea ice thickness estimation from passive microwave satellite observations between 1.4 and 36 GHz," *Earth Space Sci.*, vol. 10, no. 2, 2023, Art. no. e2022EA002542.
- [54] A. Cheng, B. Casati, A. Tivy, T. Zagon, J.-F. Lemieux, and L. B. Tremblay, "Accuracy and inter-analyst agreement of visually estimated sea ice concentrations in canadian ice service ice charts using single-polarization RADARSAT-2," *Cryosphere*, vol. 14, no. 4, pp. 1289–1310, 2020. [Online]. Available: <https://tc.copernicus.org/articles/14/1289/2020/>
- [55] J. Key, J. A. Maslanik, T. Papakyriakou, M. C. Serreze, and A. J. Schweiger, "On the validation of satellite-derived sea ice surface temperature," *Arctic*, vol. 47, no. 3, pp. 280–287, 1994. [Online]. Available: <http://www.jstor.org/stable/40511576>
- [56] M. A. Tschudi, G. A. Riggs, D. K. Hall, and M. O. Román, "Suomi-NPP VIIRS ice surface temperature algorithm theoretical basis document (ATBD)," NASA Goddard Space Flight Center, Greenbelt, MD, Tech. Rep., 2016. [Online]. Available: https://viirsland.gsfc.nasa.gov/PDF/VIIRS_IST_ATBD_V1.pdf
- [57] S. Andersen, R. Tonboe, L. Kaleschke, G. Heygster, and L. T. Pedersen, "Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration arctic sea ice," *J. Geophysical Res.: Oceans*, vol. 112, no. C8, 2007. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JC003543>

- [58] C. Horvat, E. Blanchard-Wrigglesworth, and A. Petty, "Observing waves in sea ice with ICESat-2," *Geophysical Res. Lett.*, vol. 47, no. 10, 2020.
- [59] T. H. Vidal, P. Gamet, A. Olioso, and F. Jacob, "Optimizing TRISHNA TIR channels configuration for improved land surface temperature and emissivity measurements," *Remote Sens. Environ.*, vol. 272, 2022, Art. no. 112939. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425722000530>
- [60] H. Guo et al., "The sdgsat-1 mission and its role in monitoring SDG indicators," *Remote Sens. Environ.*, vol. 328, 2025, Art. no. 114885. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425725002895>
- [61] M. Jiang, X. Chen, L. Xu, and D. A. Clausi, "Semi-supervised sea ice classification of SAR imagery based on graph convolutional network," in *IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 1031–1034.
- [62] Y. Liu, Z. Huang, and J. Han, "Aleatoric uncertainty embedded transfer learning for sea-ice classification in SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 4980–4983.
- [63] F. J. Pena Cantu et al., "A hierarchical multitask u-net for automated sea ice mapping from AI4Arctic sea ice challenge dataset," in *Proc. OCEANS 2023 - MTS/IEEE U. S. Gulf Coast*, 2023, pp. 1–7.



David A. Clausi (Senior Member, IEEE) received the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1996.

He is currently a Professor in systems design engineering with the University Research Chair, University of Waterloo (Canada), who specializes in the field of Intelligent Systems. After the Ph.D., he worked in medical imaging with Mitra Imaging (Waterloo). He started his academic career in 1997 as an Assistant Professor in geomatics engineering with the University of Calgary, Calgary, AB, Canada.

Dr. Clausi was the Associate Dean – Research & External Partnerships in the Faculty of Engineering (2018–2024). He has many contributions, conducting research primarily in remote sensing, computer vision, image processing, sports analytics, and algorithm design. He was the Co-chair of IAPR Technical Committee 7 – Remote Sensing. He has authored or coauthored extensively, has been an Associate Editor for leading journals, received many scholarships, paper awards, research awards, teaching excellence awards and his efforts have led to successful commercial implementations, culminating in the creation and sale of a tech company. He is a Professional Engineer (Ontario), Fellow of the Canadian Academy of Engineering, Fellow of the Asia-Pacific Artificial Intelligence Association, and Fellow of the Engineering Institute of Canada.



Lily de Loë (Student Member, IEEE) received the B.A.Sc. degree in engineering physics from Queen's University, Kingston, ON, Canada, in 2023, and the M.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2025.

Her research interests focus on applying deep learning approaches to geospatial data, with an emphasis on sea ice remote sensing.



K. Andrea Scott (Member, IEEE) received the B.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1999, the M.A.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2001, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2008, all in mechanical engineering.

She was a Postdoctoral Researcher with the Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Toronto, ON, Canada, in 2008, where she was part of a team involved in the development of a sea ice data assimilation system. In 2012, she joined the Department of Systems Design Engineering, University of Waterloo, moving to the Department of Mechanical and Mechatronics Engineering in 2023. She works on remote sensing, deep learning, and computational fluids.