Semi-Supervised Multi-Source Sea Ice Classification in Small-Data Regime

Samira Alkaee Taleghan
College of Engineering, Design and Computing
University of Colorado Denver
Denver, USA
samira.alkaeetaleghan@ucdenver.edu

Morteza Karimzadeh
Department of Geography,
University of Colorado Boulder
Boulder, USA
0000-0002-6498-1763

Andrew P.Barrett
National Snow and Ice Data Center (NSIDC)
CIRES, University of Colorado Boulder
Boulder, USA
andrew.barrett@colorado.edu

Walter N.Meier
National Snow and Ice Data Center (NSIDC)
CIRES, University of Colorado Boulder
Boulder, USA
walt@colorado.edu

Farnoush Banaei-Kashani
College of Engineering, Design and Computing
University of Colorado Denver
Denver, USA
farnoush.banaei-kashani@ucdenver.edu

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825475

***Abstract*—Sea ice type classification is essential for climate change research and maritime safety. Traditionally, this process relies on manual ice charting, which is time-consuming, expensive, and requires expert knowledge, making it difficult to scale up for current demands. Automating sea ice type classification is essential to keep pace with rapidly changing sea ice conditions. However, two main challenges limit the development of effective automated classifiers. First, while ice charts provide valuable labeled data, they only offer large-area (polygon) annotations rather than pixel-level labels, leading to a lack of precise training data. Second, although there are additional datasets with useful sea ice information, effectively combining these different data sources remains difficult. To tackle the first challenge, we employed co-training and label propagation, two semi-supervised learning methods, to learn from a small amount of labeled data and a large pool of unlabeled data, thereby improving the accuracy of sea ice classifiers despite limited labeled data. To address the second challenge, we leveraged co-training's built-in ability to integrate multiple data sources during the training process for the small labeled data. Additionally, we further enhanced data integration by using an ensemble of these co-trained models after training. Our approach demonstrates significant improvements over traditional supervised methods, showcasing the potential of semi-supervised learning methods in addressing two major challenges in developing automated sea ice classification solutions. Our study shows that semi-supervised learning improved F1 scores by 17% for SAR data and 33% for AMSR2 with limited labels, compared to supervised methods, while ensembling further boosted accuracy by 33%.**

***Index Terms*—Sea Ice Classification, Semi-Supervised Learning, Co-training, Label Propagation, Data Integration**

I. Introduction

Classification of sea ice types critical in understanding the climate change and ensuring safe maritime navigation. Sea ice serves as a indicator of environmental shifts, reflecting changes in temperature, ocean currents, and atmospheric conditions [1]. In this context, ice charts play a crucial role. These charts are comprehensive maps that display the distribution, concentration, and types of sea ice in a given area. Traditionally, ice charts are created by interpreting satellite imagery, particularly synthetic aperture radar (SAR) imagery, and manually labeling areas, typically in the form of polygons [2], [3]. However, manual creation of ice charts presents several challenges. It is a time-consuming process that requires extensive expert knowledge, limiting the frequency and coverage of updates.

Furthermore, the manually created ice charts provide labels at the polygon level rather than at the pixel level. While polygon labels offer valuable information about different ice conditions over larger areas, they lack the fine-grained detail that pixel-level labels provide. Pixel-level labels enable more precise ice type classification, capturing variations within each polygon. However, acquiring pixel-level labels is significantly more challenging than creating polygon-level ice charts. It demands not only expert knowledge but also considerable time and resources, making it impractical to obtain pixel-level labels for large areas or frequent updates. The limited availability of pixel-level labeled data presents a significant challenge for developing and training accurate automated sea ice classification, particularly those based on machine learning approaches that typically require large amounts of labeled data. This limited pixel-level labeled data motivates the need for methods that can effectively utilize the limited available pixel-level labels while leveraging the more abundant unlabeled data to improve the accuracy. Due to the difficulty in acquiring pixel-level labels, many researchers attempting to automate sea ice classification have resorted to using the polygon-level labels from ice charts to create pseudo-labels for pixels. This approach involves assigning the dominant ice type within each polygon to all pixels within that polygon, effectively creating pseudo pixel-level labels from the coarser polygon-level information. Supervised machine learning algorithms trained on pseudo labeled sea ice imagery have shown promising results in identifying ice types [4], [5]; however, naturally they fall short of their potential highest accuracy due to in accuracy of the pseudo-labeled data.

In addition to addressing the challenge of limited true pixel-

979-8-3503-6248-0/24/$31.00 ©2024 IEEE
Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

level labeled data, sea ice classification can benefit significantly from the integration of complementary data sources. Synthetic Aperture Radar (SAR) data, primarily from Sentinel-1 instrument, plays a vital role in sea ice monitoring due to its all-weather, day-and-night capabilities and high resolution [6], [7]. Complementing SAR, passive microwave sensors like the Advanced Microwave Scanning Radiometer-2 (AMSR-2) provide valuable information on ice concentration and extent. While previous studies have achieved notable accuracy using these sources individually, they fall short by not integrating these complementary data sources [8], [9]. This paper addresses these gaps by integrating data during training and post-training, especially with limited labeled data. Combining SAR and AMSR2 data can significantly enhance classification accuracy, especially in distinguishing visually similar ice types. SAR offers high-resolution surface patterns, while AMSR2 provides data on ice thickness and concentration. Integrating these complementary sources is a promising approach for improving automated sea ice classification, particularly with limited labeled data.

This paper makes two main contributions to sea ice classification. First, we address limited labeled data using semi-supervised methods: label propagation, which spreads labels across similar data points using graph-based techniques, and co-training, where two models trained on different data views iteratively improve by leveraging each other's confident prediction [10], [11]. Our co-training approach trains separate models on different data views, allowing them to reinforce each other's learning. Label propagation further spreads the sparse labeled data influence across the dataset.Notably, our semi-supervised strategies boost F1 scores by 17% for SAR and 33% for AMSR2 with 48 or fewer labeled samples. We evaluated our approach across different geographical locations to assess its spatial generalizability in various limited-data scenarios. Additionally, we conducted a parameter sensitivity analysis to fine-tune our models for optimal performance under varying conditions of limited pixel-level labeled data availability.

Second, we enable data integration by introducing a two stage process. We propose a two-stage data integration process specifically designed to maximize the utility of limited pixel-level labeled data with the two stages of during training (DT stage) and after training (AT stage). In the DT stage, we use co-training to develop models on different data views, such as SAR and AMSR2, enabling them to learn complementary features. With AT stage, we enhance classification accuracy by ensembling the co-trained models, which result in a 33% improvement in the F1 score. This approach allows us to exploit the underlying structure and relationships within the data to iteratively improve model performance.

By combining semi-supervised learning techniques with data integration strategies for limited labeled data, our approach significantly improves sea ice classification performance while minimizing the need for extensive manual pixel-level labels.

The remainder of this paper is organized as follows. Section II reviews the related work. In Section III, we detail the methodologies employed in our study. Section IV presents our experimental evaluation, including a thorough parameter sensitivity analysis and model performance assessment, highlighting key factors that influence the results. Finally, Section V concludes the paper with a summary of our findings and discusses potential directions for future research in this domain.

II. Related Work

This section reviews three relevant areas for sea ice classification: deep learning methods, which enhance sea ice classification accuracy but require extensive labeled data; semi-supervised methods, which leverage both labeled and unlabeled data; and data integration techniques, which combine sources to enhance classification accuracy and robustness.

A. Deep learning Methods for Sea Ice Classification

Advancements in machine learning, particularly deep learning, have significantly improved sea ice classification. Convolutional neural networks (CNNs) have become a powerful tool, capturing spatial and textural information from SAR and passive microwave imagery, leading to highly accurate sea ice type classification.

A sea ice classification method using Sentinel-1 SAR data was proposed, employing a CNN trained on expert-labeled ice charts to achieve computational efficiency and noise robustness [12]. The Sea Ice Residual Convolutional Network (SI-ResNet) with ensemble learning was developed to classify ice types from SAR imagery, surpassing traditional methods [13]. A hierarchical CNN pipeline was introduced for SAR-based sea ice mapping, improving boundary delineation and classification accuracy with limited training data [5]. CNN performance on Gaofen-3 images was enhanced by training with larger patch assemblies [14]. While these methods show impressive results, they rely on polygon-level labels from ice charts, using them as pseudo-labels for pixel-level classification, which introduces uncertainty due to labeling inaccuracies. Our approach uses a semi-supervised framework to reduce dependence on limited pixel-level labels, better leveraging both labeled and unlabeled data, and improving generalization. While many sea ice classification approaches exist, our focus is on addressing the challenge of limited labeled data. To evaluate our method, we compare it with a representative CNN-based classifier [12].

B. Semi-Supervised Learning Methods for Sea Ice Classification

Semi-supervised learning (SSL) has proven effective for sea ice classification, where pixel-level labeling is costly and scarce. SSL leverages limited labeled data with abundant unlabeled data, bridging supervised and unsupervised methods [15]. Though SSL is popular in various fields, its use in sea ice classification is limited. The Teacher-Student Label Propagation (TSLP) method was proposed for binary sea ice classification, combining teacher-student models with label propagation

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

to enhance accuracy [16]. The CFATSVM framework was introduced, integrating active learning and semi-supervised learning (SSL) for hyperspectral sea ice classification [17]. Semi-supervised GANs were also utilized to classify icebergs, ocean waves, and sea ice [18]. In [19], self-training IRGS (ST-IRGS) is introduced, a method that merges iterative region growing using semantics (IRGS) algorithm [20] with SSL for SAR-based ice-water classification, improving accuracy with minimal labeled data.

However, the existing approaches in semi-supervised work often focus on specific limitations: some use only SAR data, others tackle binary classification such as ice-water classification, or involve human experts in the labeling process. Our approach targets multi-class sea ice classification involving various ice types. We integrate multiple semi-supervised learning techniques, including co-training and label propagation, specifically designed for multi-class sea ice classification utilizing both SAR and AMSR2 data.

C. Data Integration for Sea Ice Classification

Data integration is frequently used in remote sensing to combine complementary information from multiple sensors in order to improve data quality and interpretation. This can occur at pixel-level (combining data per pixel), feature-level (merging extracted features), or decision-level (integrating separate analyses) [21]. For example, [22] used early (pixel-level), deep (feature-level), and late (decision-level) integration of Sentinel-1 SAR and AMSR-2 data for sea ice classification. In [23], SAR and optical data were fused using multi-scale SAR features and optical features via Improved Spatial Pyramid Pooling (ISPP) and Path Aggregation Network (PANet). In [24], feature-level integration was applied by first processing SAR data with a CNN and then incorporating AMSR2 data at a deeper layer to enhance accuracy. Previous work on data integration has focused mainly on raw data integration during training, often struggling with limited labeled data. We enhance the robustness and accuracy of classification by integrating data both during and after training, while addressing the lack of labeled data.

III. Methods

This section addresses challenges in sea ice classification, focusing on multi-class classification of open water and five ice types. We use semi-supervised methods, like co-training and label propagation, to leverage both labeled and unlabeled data, and integrate data during and after training to boost model accuracy and robustness.

A. Addressing Lack of Pixel-Level Labeled Data

One of the primary challenges in sea ice classification is the shortage of pixel-level labeled data, which is essential for training robust machine learning models. To address this, we utilized two semi-supervised learning techniques: co-training and label propagation. Co-training exploits multiple data views, while label propagation utilizes data point similarities, making them complementary techniques for enhancing classification performance in this domain.

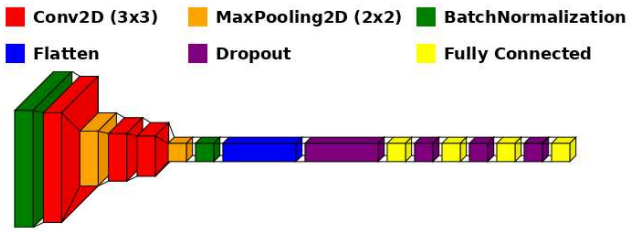

Fig. 1. CNN Architecture Layers and Filters

1) Co-training: Co-training initially proposed by Blum and Mitchell, has been widely studied and applied in various domains, including and image recognition [10]. Our co-training method involves training two separate CNN models on different views of the data: one model on dual-polarized SAR images and the other on AMSR2 data. We adopted a CNN architecture inspired by [12], consisting of Conv2D layers with Batch Normalization, followed by MaxPooling2D operations. The extracted features are processed through fully connected layers with Dropout regularization, as shown in Fig. 1. The co-training process is as follows: a) Initialization: Two classifiers are trained on separate views, SAR and AMSR2, using a small labeled set. b) Self-Labeling: Each classifier generates pseudo-labels for unlabeled data. c) Confidence-based Selection: Samples with high prediction confidence are added to the other classifier's labeled set. d) Peer Learning: The classifiers are re-trained with the expanded sets, boosting accuracy. e) Iteration: Steps (a) to (d) repeat until a stopping criterion, like a set iteration count or performance convergence, is met. As an extension of co-training, we set a high confidence threshold to select unlabeled samples for incorporation into the training dataset. If no samples meet this threshold in an iteration, it is gradually lowered to include more samples in subsequent cycles. This approach allows each model to iteratively incorporate highly confident unlabeled samples, enhancing the training process over multiple iterations. This co-training approach enables the CNN models to benefit from each other's predictions, effectively utilizing and integrating the diverse information from SAR and AMSR2 data during model training to enhance classification accuracy.

2) Label Propagation: Label propagation, originally introduced in [11], is a powerful yet straightforward iterative algorithm designed for semi-supervised learning, aiming to propagate labels through datasets along high-density areas defined by unlabeled data. The methodology assumes similar data points have similar labels. he algorithm constructs a graph G with nodes as data points and edges weighted by W, a similarity metric (e.g., Euclidean distance) [15]. Label propagation infers labels for unlabeled samples by leveraging data structure, using a K-nearest neighbor (KNN) kernel to find the closest neighbors and propagate their labels. This approach uses a sparse matrix to efficiently capture key relationships, minimizing memory and computation. We fine-tuned the neighborhood size parameter K for label propaga-

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

tion, optimizing it separately for SAR and AMSR2 datasets to account for their distinct characteristics. Data were also flattened to ensure compatibility with the label propagation algorithm.

B. Data Integration

The second challenge is effective data integration. Sea ice classification benefits from combining diverse sources, like SAR and AMSR2, which capture different ice characteristics. We integrated data in two stages: during training (DT) and after training (AT).

1) Data Integration During Training (DT Stage): The data integration process during the co-training phase involves training separate models on SAR and AMSR2 datasets simultaneously within a shared learning framework. Each model processes its respective dataset, learning distinct features specific to SAR and AMSR2. Through the co-training mechanism, these models exchange valuable insights and update their learning strategies based on the feedback from each other. This method allows each model to benefit from the unique strengths of the other dataset, leading to a more comprehensive and robust learning outcome as they integrate the diverse characteristics of both SAR and AMSR2 data throughout the training process.

2) Data Integration After Training (AT Stage): For after training integration, we utilize a stacking ensemble technique [25], which also functions as a method of data integration. When a single CNN is trained on concatenated SAR and AMSR2 data, it often struggles to effectively learn and integrate the distinct features from both data types. The increased complexity of the task can prevent the model from fully exploiting the unique characteristics of each dataset. Conversely, by training separate CNNs on SAR and AMSR2 data, each model can specialize in learning features specific to its respective dataset. The SAR-specific CNN focuses on high-resolution spatial features, while the AMSR2-specific CNN interprets the physical properties captured by microwave radiation. After the co-training phase (i.e., DT stage), we integrate the predictions from these models through ensembling. Specifically, we take the predicted probabilities from both the SAR-trained and AMSR2-trained models and combine them as input features for a logistic regression model, which is trained on a validation set. This stacking approach allows the logistic regression model to learn how best to combine the information from the two co-trained models, effectively integrating the diverse insights captured by each data source. The trained logistic regression model is then used to generate ensemble predictions on the test set by combining the predicted probabilities from the SAR and AMSR2 models. This final ensemble prediction reflects a more comprehensive understanding of the data, as it integrates the strengths of both SAR and AMSR2 models.

IV. Experimental Evaluation

This section evaluates our sea ice classification methods, emphasizing semi-supervised approaches and data integration.

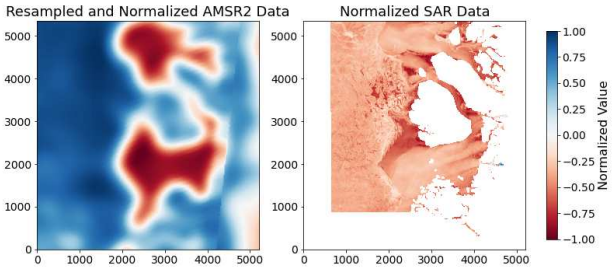

Fig. 2. Comparison of AMSR2 and SAR Data Representations of File 20180122T205424

We compare our method with supervised models through extensive experiments, detailing our methodology and results.

A. Experimental Methodology

1) Dataset: In our study, we utilize the AI4Arctic Sea Ice Challenge dataset, the largest and most comprehensive dataset available for this sea ice classification, which covers a wide geographical area and includes diverse data sources [26]. Each file in the dataset comprises two key modalities: Sentinel-1 SAR data and corresponding AMSR2 passive microwave data, supplemented by detailed ice charts based on that Sentinel-1 image, which provide polygon-level labels for sea ice types. SAR and AMSR-2 are both microwave-based remote sensing technologies, but they differ in function. SAR, an active sensor, emits its own signals and measures the reflections, enabling high-resolution imaging of surface features regardless of weather or light conditions. The dataset includes dual-polarized (HH and HV) Sentinel-1 Extra Wide Swath (EW) images, which cover areas of 400×400 square kilometers with a pixel spacing of 40×40 meters. These images have been noise-corrected using the NERSC algorithm [27], capture essential details about sea ice surface features. Complementing the SAR data, AMSR2 is a passive microwave radiometer aboard the JAXA GCOM-W satellite that detects naturally emitted microwave radiation. This sensor provides broad, low-resolution data, useful for monitoring environmental parameters such as sea ice concentration and soil moisture. AMSR2 records brightness temperatures at various frequencies in both horizontal and vertical polarizations, enriching the dataset. Fig. 2 shows SAR and AMSR2 data features side by side.

The ready-to-train (RTT) version of the AI4Arctic dataset preprocessed for deep learning. Compared to the raw data, this RTT dataset includes downsampling of SAR and ice chart data from 40-meter ($10,000 \times 10,000$ pixels) to 80-meter resolution ($5,000 \times 5,000$ pixels). Downsampling was achieved with a 2×2 averaging kernel for SAR and a 2×2 max kernel for ice charts, aligning masks (nan-values) accordingly. Data was then standardized using the mean and standard deviation from the training set. The AI4Arctic dataset spans diverse locations and periods. For our study, we used an RTT

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

subset of 11 files from winter seasons between 2018 and 2021, with dual-polarized SAR (HH, HV) and AMSR2 data at 18.7 GHz, incorporating both horizontal and vertical polarizations. SAR and AMSR2 swaths were matched within a seven-hour window, and AMSR2 data was resampled to SAR coordinates using Gaussian weighted interpolation [26].

For label preparation, the ice chart data was used to prepare labels for the curated 2-modal SAR-AMSR2 dataset described above. The dataset comprises six ice type classes/labels used in ice charts: open water (0), new ice (1), young ice (2), thin first-year ice (3), thick first-year ice (4), and old ice which is more than 1 year old (5). These types derive from ice chart data where each polygon on the chart denotes a specific sea ice type along with its partial ice concentration. Partial ice concentration within a polygon refers to the proportion of the polygon's area that is covered by a specific type of ice. This is crucial as it impacts the dominant ice type designation for the polygon. For instance, if a specific ice type's concentration exceeds 65% of the polygon's area, it becomes the dominant type and assigns the polygon-level label.

Finally, based on curated and labeled polygon-level data described above, we generate 2-modal pixel-level labeled data to be used for our experimentation. During the rasterization of these polygon-level labels to a pixel-based label, each pixel within the bounds of a polygon adopts the polygon's dominant label as an approximation. It is important to note that while we have access to abundant polygon-level labels, true pixel-level labels are absent. As an alternative, instead we use pseudo labeled pixels that are labeled based on the dominant ice type within the polygon that contains them. To reduce approximation errors, we use polygons where the dominant ice type covers over 65% of the area and limit our experiments to a subset of these pseudo-labeled pixels, simulating limited true pixel labeled data. Despite these limitations, our approach shows promising results for sea ice classification

For each file in the dataset containing SAR and AMSR2 data, we generated 32×32 patches with a stride of 32, focusing on patches within polygons representing a single ice type or open water, yielding approximately 98,605 patches. These were split into training (80%), validation (10%), and test (10%) sets, resulting in around 79,869, 8,874, and 9,860 samples, respectively. The training set was further divided into labeled and unlabeled subsets, with a minimal number of labeled samples per class selected randomly to ensure representation, while the remainder was used as unlabeled data to simulate limited labeling

Additionally, for cross-location co-training experiment, We focus on two regions, SouthEast and Qaanaaq, selected for their comprehensive coverage of all six sea ice classes. The dataset is location-tagged within each filename. For data preparation, we utilized only the HH channel of the SAR imagery to generate data patches, following the procedure previously described. Data patches are generated at a size of 32×32 pixels to maintain uniformity across analyses. The data include 60,665 samples from SouthEast and 106,885 from Qaanaaq. The combined test set from both locations includes about 15,080 samples.

2) Evaluation Metric: The F1 score, a commonly used metric, was employed to assess models performance. F1 score considers both precision, which measures the ratio of correctly predicted positive observations to the total predicted positives, and recall, which calculates the ratio of correctly predicted positive observations to all actual positives. F1 score is computed using the formula:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{1}$$

3) Configuration of the Models: To evaluate our co-training approach, two CNN models were trained separately on SAR and AMSR2 datasets. Fig. 1 shows the architecture and layer specifications. Key parameters included a batch size of 16, 30 epochs, the Adam optimizer (learning rate 0.001), and categorical cross-entropy loss, ensuring effective training and validation.

Additionally, A 95% confidence threshold was initially set to select pseudo-labeled samples for training, decreasing by 0.05 if no samples met it in an iteration. Up to 100 highly confident pseudo-labeled samples per model were added each cycle, repeating for up to 30 cycles. Training stopped if accuracy did not improve over three iterations. Performance was assessed using the F1 score, comparing co-trained models with a supervised CNN.

In experiments on co-training across various locations, we used uniform CNN model parameters (Fig. 1) consistent with the co-training framework. Each location's CNN model was trained using this co-training setup. For comparison, we trained identical CNN models in a supervised manner, enabling us to assess whether co-training across locations improves performance over location-specific supervised models.

The label propagation approach used a k-nearest neighbor kernel with K=6, optimized for both datasets after testing values from 1 to 10. The model ran for a maximum of 100 iterations, refining label assignments to improve classification. Results, evaluated with the F1 score, were compared to a supervised CNN.

In the after-training (AT) stage, we used the same parameters as in co-training, employing a stacking ensemble with logistic regression to combine predictions from both models. The logistic regression model used L2 regularization (C =1.0) and the 'lbfgs' solver, with up to 100 iterations for convergence. This stacked model's performance was compared on test data with a supervised CNN using early integration, where SAR and AMSR2 data were concatenated.

Each model was run five times, and the F1 score values represent the average of these runs. The models were developed using TensorFlow and Keras, with computations performed on a single NVIDIA RTX A6000 GPU.

B. Experimental Results

Our experiments evaluated our proposed SSL methods and data integration techniques for sea ice classification outperformed traditional supervised methods, effectively addressing

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

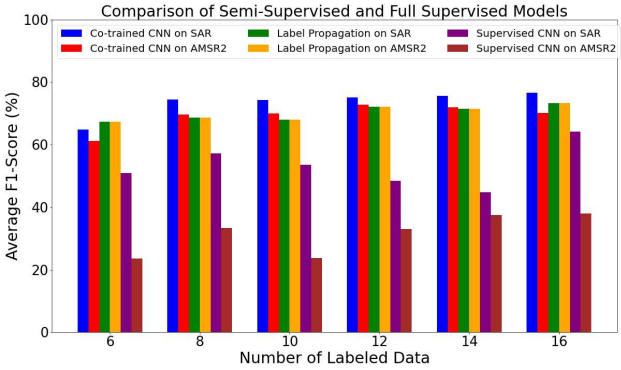

Fig. 3. Comparison of Average F1-Score (%) on Test Set Between Semi-Supervised and Supervised Models

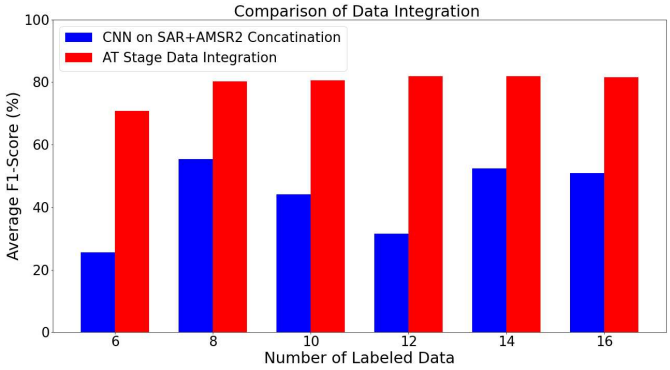

Fig. 4. Average F1-Score (%) on Test Set: CNN with SAR+AMSR2 Concatenation vs. Data Integration at AT Stage

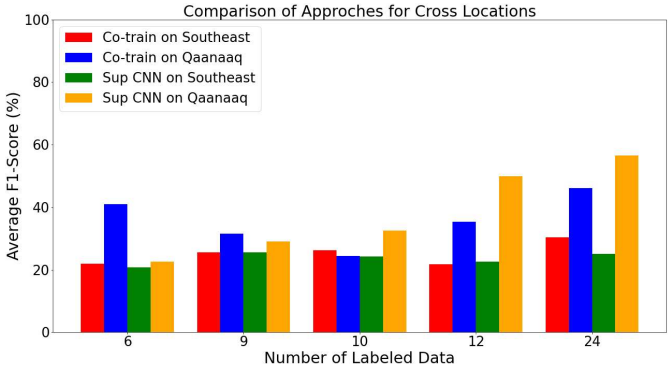

Fig. 5. Average F1-score (%) on Test Set for Cross-Location Co-training and Supervised Models

the lack of pixel-level labels. Co-training, label propagation, and data integration significantly improved classification accuracy and robustness, proving effective even with limited labeled data.

For comparison, we used supervised CNN models on SAR and AMSR2 data as baselines, trained for 30 epochs with a batch size of 16, Adam optimizer, and categorical cross-entropy loss. We compared these baselines to individual co-trained and label propagation models, each trained on SAR or AMSR2 data. To assess data integration, we also compared two-stage integration methods against a CNN trained with concatenated SAR and AMSR2 data, allowing a thorough evaluation of SSL approaches and data integration against supervised models.

a) Comparative Analysis of Semi-Supervised vs. Supervised Approaches: Fig. 3 shows the performance of various approaches across different quantities of pixel-level labeled data. Under conditions of limited labeled data, the SSL approach proves particularly advantageous, transforming the challenge of sparse supervision into a strength by adeptly leveraging unlabeled data to substantially enhance classification performance. Particularly, the individual co-trained model trained on AMSR2 data exhibits significant improvement when incorporating information from the SAR model, surpassing the performance of its corresponding supervised counterpart by approximately 30% with fewer than 16 labeled data instances. Similarly, for experiments where the individual co-trained model focuses on SAR data, with labeled data fewer than 16 instances, the mean improvement over the corresponding supervised model is approximately 13%. This improvement underscores the efficacy of leveraging complementary information from the AMSR2 data under small data regime.

b) Comparative Analysis of AT Integration vs. Data Concatenation: Fig. 4 compares the results of a supervised CNN trained on the concatenation of SAR and AMSR2 data with after training integration (AT stage) of co-trained models. The results demonstrate significant improvements in accuracy for the AT stage integration of co-trained CNNs compared to the single CNN trained on concatenated data, across various numbers of pixel-level labeled data points. For instance, with six pixel-level labeled data points, the supervised CNN on concatenation of SAR and AMSR2 achieved an accuracy of 25.56%, while the AT stage of co-trained models reached 70.88%. This trend of superior performance for the AT stage method continues as the number of labeled data points increases.

c) Comparative Analysis of Cross-Location Co-Training vs. Supervised Learning: Fig. 5, which includes curves that are titled Co-trained on SouthEast, Co-trained on Qaanaaq, Supervised CNN on SouthEast, and Supervised CNN on Qaanaaq. These titles correspond to the respective models and locations where their data are used to train the models. Note that in this experiment, when the x-axis shows 6 labeled data points, it means that each co-trained model is trained with 6 pixel-level labeled samples. In contrast, the supervised models specific to each location are trained with 12 labeled samples from that location. This ensures fairness of the comparison while the co-trained models benefit from the diverse information provided by both locations. We tested with labeled sample sizes of 6, 9, 12, and 24 for each co-trained model. The results show that in the Southeast location, the co-trained models perform as well as or better than the supervised models, especially with 24 labeled samples. In Qaanaaq, the co-trained models outperform the supervised models at 6 and 9 labeled samples,

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

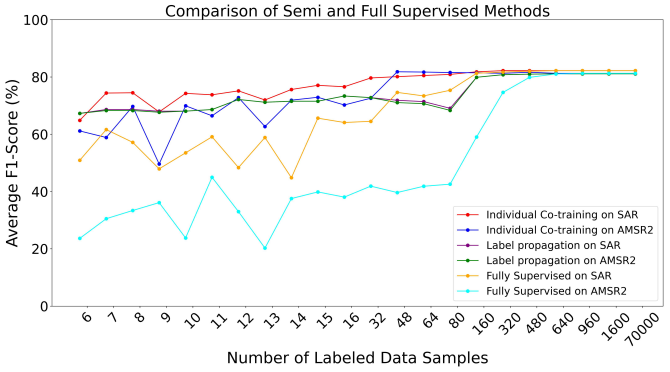

Fig. 6. Average F1-score (%) on test set: Comparison of different approaches for different number of labeled data.

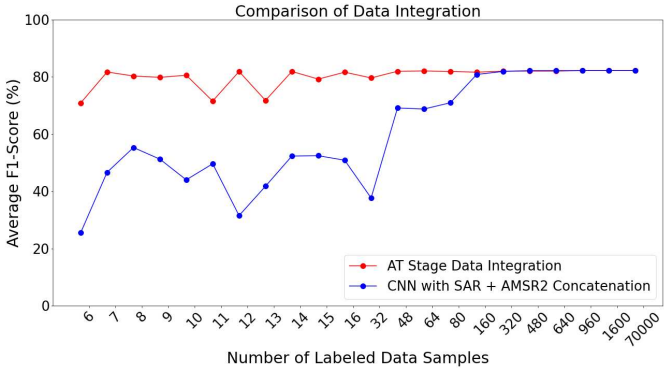

Fig. 7. Average F1-Score (%) on Test Set: CNN with SAR+AMSR2 Concatenation vs. Data Integration at AT Stage Across Different Numbers of Labeled Data

but as the number of labeled samples increases, the supervised models begin to perform better as expected.

1) Parameter Sensitivity Analysis: Next, We conducted a sensitivity analysis to examine how varying the number of pixel-level labels (from 6 to full dataset) and amounts of unlabeled data (from 1 to nearly 6000) impact model performance. This analysis helps optimize the balance between labeled and unlabeled data, fine-tuning the SSL process for best performance under different data conditions.

a) Influence of Labeled Data Ratio on Supervised and Semi-Supervised Learning Performance: Fig. 6 shows the average F1 score on test set for each of the SSL and full supervised models across different numbers of pixel-level labeled data points. The plot shows that as the number of labeled data points increases, the performance of the supervised approach becomes more competitive. Specifically, for SAR data, the co-trained models maintain superior performance until the number of labeled data points reaches 160, with co-training improving performance by an average of 17%. Also, for AMSR2 data, the co-trained model outperforms the supervised CNN until the number of labeled samples exceeds 640, with co-training showing on average 31% improvement. Label propagation approach consistently demonstrates competitive performance, even with limited labeled data, highlighting its efficacy in leveraging unlabeled data. For SAR data, the label propagation approach outperforms the supervised CNN until the number of labeled instances reaches 48, improving performance by an average of 15%. Similarly, for AMSR2 data, label propagation maintains its advantage until there are 640 labeled instances, showing on average 27% improvement. As a result of these findings, it becomes evident how labeled data size differ between label propagation approach and supervised CNN. Despite this, label propagation remains a strong option, especially in small data scenarios, effectively leveraging both labeled and unlabeled data.

b) Influence of Labeled Data Ratios on Data Integration: Fig. 7 illustrates the impact of varying the number of labeled data samples on performance of both CNN models with data concatenation and model with after-training data integration. The figure demonstrates the consistent performance gains achieved by the after-training (AT) data integration approach. The model traning by AT stage integration consistently outperforms the CNN model trained on concatenated SAR and AMSR2 data, particularly excelling with fewer labeled samples. Notably, until the number of labeled data points reaches 160, the AT integration shows an average improvement of 29% over the supervised model. This approach effectively combines the strengths of co-trained SAR and AMSR2 models, yielding more accurate and robust classifications. Even with more labeled samples, the AT integration maintains its advantage, underscoring the importance of data integration during and after training. In conclusion, integrating co-trained CNNs on SAR and AMSR2 data outperforms a single CNN on concatenated data, achieving higher accuracy and robustness by leveraging each dataset's unique characteristics.

c) Evaluating SSL Performance with Varying Unlabeled Data Ratios: Fig. 8 illustrates how the amount of unlabeled data, with a fixed count of 16 pixel-level labeled data points, influences the average F1 score for both label propagation models and co-trained CNNs using SAR and AMSR2 data. The data show a clear trend: as the volume of unlabeled data increases, there is a corresponding improvement in the F1 score. This pattern highlights the value of incorporating unlabeled data to capitalize on the inherent data distribution, which in turn boosts model performance through iterative refinements. The results for the co-training approach reveal significant performance enhancements across individual co-trained models trained on SAR and AMSR2 data with varying quantities of unlabeled data. As the number of unlabeled data points increases, F1 scores consistently improve, highlighting the value of unlabeled data in enhancing model accuracy within an SSL framework, especially when labeled data is limited.

V. Conclusion

In conclusion, this study addresses the challenges of sea ice classification with limited pixel-labeled data by investigating SSL techniques, including co-training, label propagation, and data integration through two starge of during and after training.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.

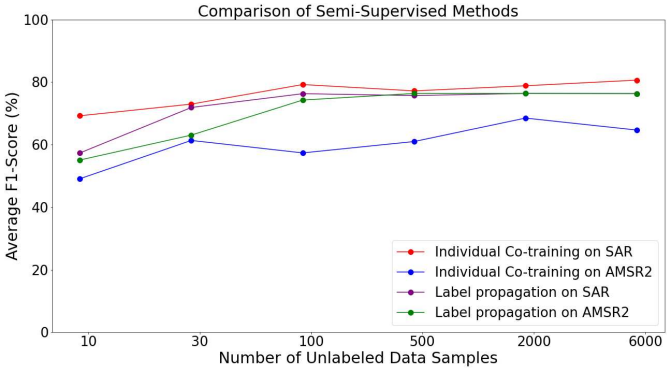

Fig. 8. Average F1-score (%) on test set for label propagation and co-trained CNNs: varying unlabeled data with fixed labeled data (16)

By leveraging SAR and AMSR2 data, we demonstrated that these approaches significantly enhance classification accuracy, particularly in scenarios with limited pixel-level labeled data. The integration of data through ensemble methods further improves model performance, effectively combining the strengths of different data sources both during and after training. Our findings establish the superiority of SSL methods over supervised CNN models, especially when labeled data is scarce, while also highlighting the continued relevance of supervised models as labeled data increases.

For future work, several promising directions emerge. Firstly, exploring advanced co-training strategies, such as incorporating additional modalities, holds potential for further improving classification accuracy. Additionally, extending our study to incorporate temporal and spatial information could enhance the robustness of the classification system, especially in dynamic sea ice environments.

Acknowledgment

The authors would like to acknowledge the support of the U.S. National Science Foundation under Grants No. 2026962 and 2026865, and the Center for Computational Mathematics at the University of Colorado Denver, including access to the Alderaan cluster, funded by NSF award OAC-2019089, for providing computational resources.

References

[1] L. P. Bobylev and M. W. Miles, "Sea ice in the arctic paleoenvironments," in *Sea Ice in the Arctic: Past, Present and Future*, 1st ed. London, U.K.: Springer, 2020, vol. 1, pp. 9–56.

[2] U.S. National Ice Center, "Arctic Sea Ice Charts and Climatologies in Gridded Format, 1972–2007, Version 1," Data set, Boulder, Colorado, USA, 2006, compiled by F. Fetterer and C. Fowler. [Online]. Available: https://doi.org/10.7265/N5X34VDB.

[3] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Satellite sar data-based sea ice classification: An overview," *Geosciences*, vol. 9, no. 4, p. 152, 2019.

[4] S. Khaleghian, H. Ullah, T. Kræmer, N. Hughes, T. Eltoft, and A. Marinoni, "Sea ice classification of sar imagery based on convolution neural networks," *Remote Sensing*, vol. 13, no. 9, p. 1734, 2021.

[5] X. Chen, K. A. Scott, M. Jiang, Y. Fang, L. Xu, and D. A. Clausi, "Sea ice classification with dual-polarized sar imagery: A hierarchical pipeline," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 224–232.

[6] C. R. Jackson and J. R. Apel, *Synthetic Aperture Radar: Marine User's Manual*. Washington, DC, USA: NOAA, 2004.

[7] J. W. Park, A. A. Korosov, M. Babiker, J. S. Won, M. W. Hansen, and H. C. Kim, "Classification of sea ice types in sentinel-1 synthetic aperture radar images," *The Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020.

[8] H. Lyu, W. Huang, and M. Mahdianpari, "Eastern arctic sea ice sensing: First results from the radarsat constellation mission data," *Remote Sensing*, vol. 14, no. 5, p. 1165, 2022.

[9] Y. R. Wang and X. M. Li, "Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning," *Earth System Science Data*, vol. 13, no. 6, pp. 2723–2742, 2021.

[10] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the 11th Annual Conference on Computational Learning Theory*, 1998, pp. 92–100.

[11] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Carnegie Mellon University, Tech. Rep., 2002.

[12] H. Boulze, A. Korosov, and J. Brajard, "Classification of sea ice types in sentinel-1 sar data using convolutional neural networks," *Remote Sensing*, vol. 12, no. 13, p. 2165, 2020.

[13] W. Song, M. Li, Q. He, D. Huang, C. Perra, and A. Liotta, "A residual convolution neural network for sea ice classification with sentinel-1 sar imagery," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 795–802.

[14] J. Li, C. Wang, S. Wang, H. Zhang, Q. Fu, and Y. Wang, "Gaofen-3 sea ice detection based on deep learning," in *2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)*, 2017, pp. 933–939.

[15] X. J. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison, Tech. Rep., 2005.

[16] S. Khaleghian, H. Ullah, T. Kræmer, T. Eltoft, and A. Marinoni, "Deep semisupervised teacher–student model based on label propagation for sea ice classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10 761–10 772, 2021.

[17] Y. Han, Y. Zhao, Y. Zhang, J. Wang, S. Yang, Z. Hong, and S. Cao, "A cooperative framework based on active and semi-supervised learning for sea ice classification using eo-1 hyperion data," *Transactions of the Japan Society for Aeronautical and Space Sciences*, vol. 62, no. 6, pp. 318–330, 2019.

[18] F. Staccone, "Deep learning for sea-ice classification on synthetic aperture radar (sar) images in earth observation: Classification using semi-supervised generative adversarial networks on partially labeled data," Master's thesis, M.S. thesis, 2019.

[19] F. Li, D. A. Clausi, L. Wang, and L. Xu, "A semi-supervised approach for ice-water classification using dual-polarization sar satellite imagery," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015, pp. 28–35.

[20] Q. Yu and D. A. Clausi, "Sar sea-ice image analysis based on iterative region growing using semantics," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3919–3931, 2007.

[21] C. Pohl and J. L. V. Genderen, "Review article: Multisensor image fusion in remote sensing: Concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823–854, 1998.

[22] L. Zhao, T. Xie, W. Perrie, and J. Yang, "Deep-learning-based sea ice classification with sentinel-1 and amsr-2 data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 5514–5525, 2023.

[23] Y. Han, Y. Liu, Z. Hong, Y. Zhang, S. Yang, and J. Wang, "Sea ice image classification based on heterogeneous data fusion and deep learning," *Remote Sensing*, vol. 13, no. 4, p. 592, 2021.

[24] D. Malmgren-Hansen, L. T. Pedersen, A. A. Nielsen, M. B. Kreiner, R. Saldo, H. Skriver, and K. H. Krane, "A convolutional neural network architecture for sentinel-1 and amsr2 data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1890–1902, 2020.

[25] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[26] J. Buus-Hinkler, T. Wulf, A. R. Stokholm, A. Korosov, R. Saldo, L. T. Pedersen, and *et al.*, "Ai4arctic sea ice challenge dataset," Technical University of Denmark, 2022, [Online]. Available: https://doi.org/10.11583/DTU.c.6244065.v2.

[27] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

Authorized licensed use limited to: UNIVERSITY OF COLORADO. Downloaded on December 11,2025 at 05:24:18 UTC from IEEE Xplore. Restrictions apply.