

Deep Semisupervised Teacher–Student Model Based on Label Propagation for Sea Ice Classification

Salman Khaleghian , Habib Ullah, Thomas Kræmer , Torbjørn Eltoft , *Member, IEEE*,
and Andrea Marinoni , *Senior Member, IEEE*

Abstract—In this article, we propose a novel teacher–student-based label propagation deep semisupervised learning (TSLP-SSL) method for sea ice classification based on Sentinel-1 synthetic aperture radar data. For sea ice classification, labeling the data precisely is very time consuming and requires expert knowledge. Our method efficiently learns sea ice characteristics from a limited number of labeled samples and a relatively large number of unlabeled samples. Therefore, our method addresses the key challenge of using a limited number of precisely labeled samples to achieve generalization capability by discovering the underlying sea ice characteristics also from unlabeled data. We perform experimental analysis considering a standard dataset consisting of properly labeled sea ice data spanning over different time slots of the year. Both qualitative and quantitative results obtained on this dataset show that our proposed TSLP-SSL method outperforms deep supervised and semisupervised reference methods.

Index Terms—Deep learning, earth observation, scarce training data, sea ice classification, semisupervised learning (SSL).

I. INTRODUCTION

ARCTIC sea ice keeps the northern polar regions cool and thereby helps to moderate the global climate. It is a key component of the Arctic environment [1] that substantially affects the polar physical environment and its ecosystems. The Arctic has faced severe environmental impacts over the past few decades. These changes have transformed its environment, ecology, and meteorology and caused unsteady variations in the weather and sea ice conditions, which pose new challenges to maritime industries, including but not limited to aquaculture, natural energy resources, and travel exploration operating in the high north areas [2], [3]. Therefore, proper monitoring

of the sea ice conditions and how it changes with time is important [4], [5].

For high-resolution sea ice analysis, researchers and ice centers around the world are using synthetic aperture radar (SAR) data [6], [7]. These data are not restricted by weather conditions and polar darkness [8]. An important part of sea ice analysis includes sea ice classification. Sea ice classification based on SAR data [9] is carried out by classical statistical classification methods, traditional machine learning (TML) methods, and deep-learning-based methods (DLMs). Statistical and TML methods rely on handcrafted features, which may not properly encapsulate the challenging sea ice characteristics [10]. Therefore, their generalization capabilities and their abilities to find efficient features that can be considered to various geographic areas and time frames are limited [10]. DLMs, when properly trained on large training datasets, have shown excellent generalization capabilities in many research fields, including several remote sensing applications such as food security monitoring [11], hybrid data-driven Earth observation modeling [12], and flood mapping from high-resolution optical data [13]. We consider these achievements in the aforementioned fields and believe that deep neural networks (DNNs) may also show performance improvement in automatic sea ice classification [14], [15]. However, scarce training data is the most challenging issue in sea ice data analysis. This problem is particularly challenging in the Arctic, where gathering of precise true observations is expensive, time driven, and sometimes not feasible [16]. For sea ice classification, archived ice charts are available rendering huge labeled data. Nonetheless, these charts are very coarsely labeled and do not have the quality and details needed to train a DLM effectively [17].

To extract accurate information from large-scale datasets, when limited amount of labeled data are available, semisupervised learning (SSL) has been introduced in the technical literature [18]. These methods aim to combine labeled data with unlabeled records. In the past few years, semisupervised models have presented performance improvement in various fields of remote sensing research, such as despeckling of SAR images [19], change detection in heterogeneous remote sensing images [20], and hyperspectral image classification [21]. Considering these successes, we anticipate that deep SSL methodologies could also be favorable in sea ice classification and potentially lead to significant improvements by overcoming the specific challenge of few labeled samples. In fact, a deep SSL technique is halfway

Manuscript received July 1, 2021; revised September 11, 2021 and September 22, 2021; accepted September 23, 2021. Date of publication October 14, 2021; date of current version November 2, 2021. This work was supported in part by the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations and the Research Council of Norway under Grant 237906, in part by the European Union's Horizon 2020 Research and Innovation Program ExtremeEarth Project under Grant 825258, and in part by the Fram Center under the Automised Large-scale Sea Ice Mapping “Polhavet” Flagship Project. (*Corresponding author: Salman Khaleghian.*)

Salman Khaleghian, Thomas Kræmer, Torbjørn Eltoft, and Andrea Marinoni are with the Faculty of Science and Technology, University of Tromsø—The Arctic University of Norway, 9019 Tromsø, Norway (e-mail: salman.khaleghian@uit.no; thomas.kramer@uit.no; torbjorn.eltoft@uit.no; andrea.marinoni@uit.no).

Habib Ullah is with the Faculty of Science and Technology, Norwegian University of Life Sciences, 1430 Ås, Norway (e-mail: habib.ullah@nmbu.no).
Digital Object Identifier 10.1109/JSTARS.2021.3119485

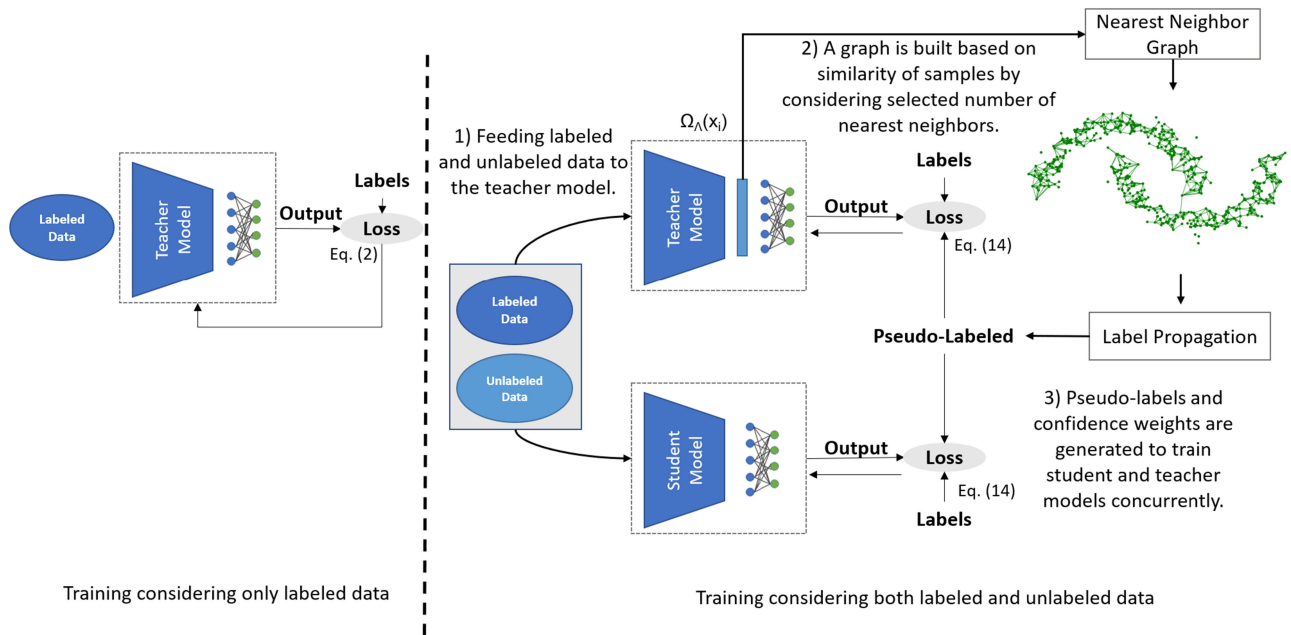


Fig. 1. TSLP-SSL method. We have two models, namely, teacher and student models. The teacher model is trained on labeled data during the first stage, and then, both models are trained on labeled and unlabeled data during the second stage of the training.

between supervised and unsupervised learning. This technique exploits multiple layers to progressively extract higher level features from the raw input data considering both labeled and unlabeled data.

We propose a *teacher–student*-based label propagation deep semisupervised learning (TSLP-SSL) method. Our architecture consists of two models, namely, a teacher model and a student model. The teacher model is trained in a two-step procedure. Initially, we trained the teacher model in a supervised fashion utilizing only the labeled data. We then feed both the labeled and unlabeled samples to the trained teacher model and consider the feature space embedding to engender pseudo-labels for the unlabeled data through a label propagation procedure [22]–[24]. The original and the pseudo-labels are in the next step used to train the student model, which is subsequently used during the inference stage. The purpose of using the student model is to avoid the problem of the teacher model being biased toward the labeled data, which is like in case of a small training set. Our proposed method, hence, effectively exploits a relatively large amount of unlabeled data to improve the final classification performance. The training methodology is depicted in Fig. 1 and is more thoroughly described in Section III. The summary of our contributions is as follows.

- 1) We propose a novel TSLP-SSL method. One of the major attractions of our proposed method is its capability to deal with a small number of labeled samples. This is a favorable property in the case of sea ice classification using SAR data, where the availability of a large amount of reliable labeled data is scarce.
- 2) We consider sea ice datasets to train and analyze the generalization capabilities of our proposed method. We compare our method with a supervised method and three

state-of-the-art semisupervised methods. Our results show that our proposed method performs better than all the reference methods, especially in cases with a small number of labeled samples.

- 3) Additionally, we present a comprehensive literature review covering both the probabilistic learning method and the DLM.

The rest of this article is organized as follows. Related work is described in Section II. We present our proposed deep models and training approaches in Section III. Section IV depicts the experimental analysis considering a set of SAR images. Finally, Section V concludes this article and presents future work.

II. RELATED WORKS

In general, sea ice classification can be divided into two major classes: TML/probabilistic methods and DLMs [25]. The approaches in the latter class fall into two subclasses, namely, supervised deep learning and semisupervised deep learning methods. The literature is very limited in the case of semisupervised DLMs since methods in this subcategory are quite recent and still under development.

A. Probabilistic Methods for Sea Ice Classification

The literature on TML/probabilistic methods is very rich, and we will restrict ourselves to only including a few recent publications. Statistical algorithms often combine probabilistic models and classical classification methods with texture or polarimetric features to perform sea-ice-type maps. An extensive survey is given in [26].

Some specific studies in this category are highlighted below. Examples of machine learning algorithms include the use

of standard multilayer perceptrons, as in [14], support vector machines, as in [7], or decision tree methods [15], as in [15]. Statistical and shallow machine learning methods often rely on having extracted the input features in a preoperation prior to the classification. Karvonen [27] and Dinnessen [28] used probabilistic and statistical features for estimating sea ice concentration from SAR imagery. Johansson *et al.* [29] used statistical entropy and horizontal–vertical (HV) polarization computations to isolate sea ice from open water and thicker sea ice. Furthermore, Fors *et al.* [30] investigated the potential of *C*- and *X*-band multipolarization SAR features for sea ice segmentation during late summer. Daboor *et al.* [31] analyzed a set of compact polarimetric parameters for classifying newly formed ice and multiyear ice. Hong and Yang [32] used the statistical coefficient, incidence angle, environment temperature, and speed of wind to improve the sea ice and water classification. Johansson *et al.* [33] used a statistical mixture model to isolate open water from sea ice. Their method is based on the semiautomatic segmentation technique. They applied the algorithm to explore the sea ice characteristics in Svalbard. Aldenhoff *et al.* [34] demonstrated that *C*-band SAR can reliably generate the layout of the ice boundary, whereas the *L*-band shows effectiveness considering thin ice and water regions.

B. DLMs for Sea Ice Classification

Deep-learning-based approaches have been widely exploited for addressing the challenge of sea ice classification. Malmgren-Hansen *et al.* [17] applied a convolutional neural network (CNN) model to predict Arctic sea ice by fusing data from two different satellites. They found that the CNNs are showing good performance for multisensor data integration. It is worth noting that they used archived ice chart data for both training and validation. However, these data are coarsely labeled, hence leading to undesired effects in the training of the CNN model. Wang *et al.* [10], [35], [36] exploited CNNs for ice concentration estimation. Tom *et al.* [37] proposed an ice monitoring model based on Sentinel-1 data with a deep learning approach. Boulze *et al.* [38] introduced a CNN for detecting different kinds of sea ice [39] using SAR data. They trained the CNN considering the archived ice chart data. They performed comparison with a random forest classifier using texture features.

SSL methods are proposed for classification when only scarce training data or a limited number of training samples are available. The idea of SSL relies on the assumption that unlabeled samples provide essential information and clues on how the data are distributed. Therefore, a DLM can be trained by considering this distribution. In this sense, different approaches such as teacher–student models [40], graph-based methods [41], pseudo-labeling [42], consistency regularization [43], and generative models (i.e., generative adversarial networks—GANs) [44] have been introduced. Shin [40] proposed a multiteacher single-student method to solve the visual attribute prediction problem. His method learnt task-specific domain experts called teacher networks and a student network by forcing a model to imitate the distributions learned by domain experts. Xie *et al.* [45] proposed a noisy student method for

generating pseudo-labels to train a model in an iterative way. The output of the trained model based on the labeled samples is exploited to produce pseudo-labels for the unlabeled samples, which are subsequently used to train another model. They used the teacher–student model to train a larger student model by incorporating noise, considering data augmentation (DA), dropout, and stochastic depth. Tarvainen and Valpola [46] proposed a mean teacher method that averages model weights instead of label predictions. Their method improves test accuracy and enables training with fewer labeled samples. Salimans *et al.* [47] trained the semisupervised generative adversarial network (semi-GAN) as a generative model. Kingma *et al.* [48] exploited a variational autoencoder in the form of a semisupervised model. In their method, a classifier is trained on top of a latent representation to predict the labels. Iscen *et al.* [24] proposed a transductive label propagation model for deep SSL. This model is trained in an iterative two-step procedure. In the first phase, a CNN is trained using the labeled part of the dataset in a supervised manner. In the second phase, based on a manifold assumption in the feature space of the CNN, pseudo-labels are produced for the unlabeled data through a label propagation procedure using a nearest neighbor graph. The pseudo-labels are considered to extend the set of labeled samples in the second stage to train the CNN model. Berthelot *et al.* [49] used an augmentation technique to introduce an SSL approach. They assumed that the distribution of a classifier should remain the same considering unlabeled data. They used average prediction to produce pseudo-labels for the unlabeled samples.

C. SSL Methods for Sea Ice Classification

The aforesaid cases show that the development of SSL methods is a hot topic in the data analysis community. However, it is also true that the application of SSL architectures to sea ice classification is very limited. For example, Han *et al.* [50] investigated an approach for sea ice classification based on active learning (AL) and SSL. They acquired the most informative data examples considering AL. They exploited these informative examples in training the SSL method. Staccone [51] introduced an SSL approach based on GANs for sea ice classification. In this work, both labeled and unlabeled data were considered to achieve more accurate results by exploiting the knowledge from both data sources. Li *et al.* [52] presented an SSL method for ice and water classification based on self-training. Their method combined a contextual model and the self-training approach into a unified framework.

Our proposed method falls into the subcategory of SSL methods. We propose a teacher–student model considering the feature space using the label propagation method, which is summarized in the following section.

III. TEACHER–STUDENT-BASED LABEL PROPAGATION METHOD

As mentioned above, labeled sea ice samples are difficult to acquire, making the training of sea ice classification architectures a difficult task. Therefore, we explore a novel TSLP-SSL method for this application. We adequately utilize a limited

number of labeled samples and a comparatively much large number of unlabeled samples to train a deep CNN architecture for extracting sea ice information. Our proposed TSLP-SSL method consists of a teacher model and a student model, which are cooperatively trained in an iterative way during two training stages. Our method is different from the teacher–student models presented in [45] and [46] in two major aspects. First, in our case, features generated by the trained teacher model are extracted before the final classification layer and used in the label propagation process to produce pseudo-labels for the unlabeled samples using a k -nearest neighbor approach. Hence, label propagation is performed in feature space, and not in output label space. Second, the pseudo-labels from the teacher model are exploited, together with the original labels, to train the student model in order to find an optimal decision boundary during a second iterative training stage. Our proposed method is also different from the deep SSL model in [24] in the way it aims to avoid the model to be biased toward the labeled data. In fact, the method in [24] is based on a single model, which is trained on only the labeled data, making it susceptible to be biased toward these data samples. The biasing problem may be even more significant in the sea ice classification task, considering the small amount of labeled data and noting the fact that texture features are important for discriminating between different ice types.

In our proposed method, both models are represented by a CNN constructed of a 13-layer architecture [24]. During the first training stage, the teacher model is trained on the labeled data only. During the second stage, the teacher model generates pseudo-labels for the unlabeled data. These pseudo-labels, combined with the labeled samples, are used to train the student model. The motivation for considering an additional student model is to handle the problem of the teacher model being biased toward the labeled data, as discussed above [53]. To further elaborate on this issue, the teacher model formulates a decision boundary considering a small set of labeled data. However, this decision boundary may not be the best boundary when also considering the unlabeled data during the second stage, especially if the teacher model gets overfitted to the labeled data because of the limited number of samples [54]. The idea is that the student model should discover a more appropriate decision boundary, as illustrated in Fig. 2. Fig. 2 displays a simplified case, in which the triangles represent samples from one arbitrary class and the circles show samples from another class. Hence, the red and blue symbols represent labeled data from the two classes, respectively, and the black symbols represent unlabeled data from both classes. Since the teacher model is trained using labeled data only, the decision boundary shown as a blue solid line in Fig. 2 could be a solution. A better decision boundary is discovered by repeatedly training the student model from scratch with both pseudo-labeled and labeled data. In this way, the student model would end up with the decision boundary defined by the green-dashed line, which properly separates both the labeled and unlabeled data from both the two classes. It is worth noting that this example shows the advantage of using label propagation based on nearest neighbors instead of using the network output as pseudo-labels.

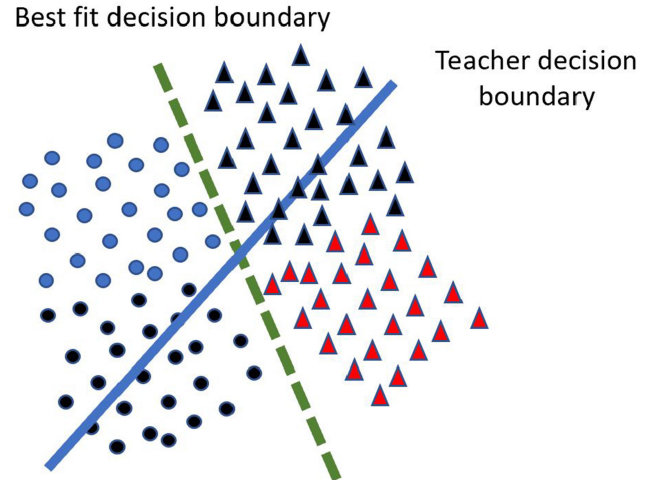


Fig. 2. Complexity of tuning the teacher decision boundary to also take into account the unlabeled data. We show two-class labeled data with red triangles and blue circles. The black markers represent the unlabeled data.

During the second stage of our training, the teacher model generates predictions for the entire dataset. The feature space embedding is subsequently used to construct a nearest neighbor graph and an adjacency matrix, from which we assign pseudo-labels to the unlabeled samples in a transductive label propagation procedure [24].

A. Formulation for the Learning Process

To clearly provide the details of the process of label propagation for our teacher model, we present the affiliated notations in this section. In this, we will largely follow the outline in [24]. We consider a set of n samples denoted by $X := (x_1, \dots, x_s, x_{s+1}, \dots, x_n)$ with $x_i \in X$, where s samples x_i for $i \in S := \{1, \dots, s\}$, represented by X_S , are labeled according to $Y_S := (y_1, \dots, y_s)$. Each element in Y_S is $y_i \in G$, where $G := \{1, \dots, g\}$ is a discrete label set of g classes. The rest of the $e := n - s$ samples x_i for $i \in E := \{s + 1, \dots, n\}$, represented by X_E , are unlabeled. We consider all samples in X and labels in Y_S to train the CNN to assign class labels to the previously unseen samples. The CNN takes an input sample x_i from X and builds a vector of class probabilities $f_\Lambda(x_i)$, $f_\Lambda : X \rightarrow \mathbb{R}^g$, where Λ represents the hyperparameters of our deep model. In this process, the feature extraction stage is represented by the function $\Omega_\Lambda : X \rightarrow \mathbb{R}^d$, which maps the input data to a d -dimensional feature vector, where the i th sample is represented by $d_i := \Omega_\Lambda(x_i)$. In the next stage, a vector of class probabilities is built by the softmax on top of the fully connected layer considering Ω_Λ . The prediction of the CNN for the i th sample is the class of the highest probability, i.e.,

$$\hat{y}_i = \operatorname{argmax}_j f_\Lambda(x_i)_j \quad (1)$$

where j is the j th dimension of the vector. In supervised learning, the loss function in (2) is minimized to train the CNN

$$\xi_{\text{sup}}(X_S, Y_S; \Lambda) = \sum_{i=1}^s \varepsilon_{\text{sup}}(f_\Lambda(x_i), y_i). \quad (2)$$

Equation (2) applies only to the labeled samples, i.e., $x_i \in X_S$. In fact, (2) shows one term of the loss function in SSL. In classification problems, the *cross-entropy* loss function is generally used for ε_{sup} , which for a given sample x_i is defined as

$$\varepsilon_{\text{sup}}(f_\Lambda(x_i), y_i) = - \sum_{k=1}^g y'_k \log(f_\Lambda(x_i))_k \quad (3)$$

where y'_k is the k th component of the one-hot encoding of $y_i \in G$. Pseudo-labeling finds a pseudo-label \hat{y}_i for each sample x_i for $i \in E$. The pseudo-labels for unlabeled samples in X_E are represented by $\hat{Y}_E = \{\hat{y}_{s+1}, \dots, \hat{y}_n\}$, and they form an additional loss term formulated as

$$\xi_{\text{pseu}}(X_E, \hat{Y}_E; \Lambda) = \sum_{i=s+1}^n \varepsilon_{\text{pseu}}(f_\Lambda(x_i), \hat{y}_i). \quad (4)$$

B. Pseudo-Label Generation and Learning Process

In our method, the CNN is represented by the parameters Λ , and we formulate the descriptor set as $D := (d_1, \dots, d_s, d_{s+1}, \dots, d_n)$, where $d_i := \Omega_\Lambda(x_i)$. We build a sparse affinity matrix $\Delta \in \mathbb{R}^{n \times n}$, where its elements are represented by

$$\nu_{ij} = \begin{cases} [d_i^T, d_j]_+^\gamma, & \text{if } i \neq j \wedge d_i \in N_k(d_j) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where N_k represents the set of k -nearest neighbors in X , and γ is a hyperparameter. It is worth noticing that building the sparse affinity matrix is computationally efficient even if we have a very large number of samples. We then build a symmetric adjacency matrix $\Theta = \Delta + \Delta^T$ such that $\Theta \in \mathbb{R}^{n \times n}$. The diagonal of the matrix Θ consists of zeroes. The rest of the elements of Θ are nonnegative pairwise similarities between d_i and d_j for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, n$. We formulate the symmetrically normalized counterpart of Θ as

$$\Xi = \Gamma^{-\frac{1}{2}} \Theta \Gamma^{-\frac{1}{2}} \quad (6)$$

where $\Gamma = (\Theta \mathbf{1}_n)$ is the degree matrix and $\mathbf{1}_n$ is an n -dimensional vector with all elements set to 1. We formulate a label matrix Y of size $n \times g$ consisting of the elements

$$Y_{ij} = \begin{cases} 1, & \text{if } i \in S \wedge y_i = j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The rows of the matrix Y represent one-hot encoded labels for the labeled samples. Subsequently, the diffusion amounts to formulating an $n \times g$ matrix ψ such that

$$\psi = (I - \alpha \Xi)^{-1} Y \quad (8)$$

where $\alpha \in [0, 1)$ is a parameter. The elements of ψ are represented by δ_{ij} . In fact, calculating matrix ψ , according to (8), is impractical for large n because the inverse matrix $(I - \alpha \Xi)^{-1}$ is not sparse. Therefore, we use the conjugate gradient method to solve the linear system

$$(I - \alpha \Xi)\psi = Y. \quad (9)$$

Equation (9) is fast and valid since the matrix $(I - \alpha \Xi)$ is positive definite. We find the pseudo-labels $\hat{Y}_E = \{\hat{y}_{s+1}, \dots, \hat{y}_n\}$

for unlabeled samples as

$$\hat{y}_i = \operatorname{argmax}_j \delta_{ij} \quad (10)$$

where δ_{ij} is the (i, j) th element of matrix ψ . It is worth noting that finding pseudo-labels from matrix ψ in this way has some unwanted causes. For example, we assign pseudo-labels to all unlabeled samples; however, we are clearly not confident about the same certainty for all generated pseudo-labels. Moreover, pseudo-labels may not represent the same number of samples for each class, which will affect the performance of the learning process. To handle the former problem, we affiliate a weight representing the certainty of the prediction to each pseudo-label. For this purpose, we consider the entropy Υ to compute the level of uncertainty and provide a weight ω_i to sample x_i formulated as

$$\omega_i = 1 - \frac{\Upsilon(\hat{\delta}_i)}{\log(g)} \quad (11)$$

where $\Upsilon : \mathbb{R}^g \rightarrow \mathbb{R}$ is the entropy function, and the weight ω_i is normalized in $[0, 1]$ because $\log(g)$ is the maximum possible entropy in \mathbb{R}^g [when all datapoints are equally distributed to the clusters, the maximum entropy for g classes is $H = -\sum_{c=1}^g 1/g \log(1/g) = \log(g)$]. $\hat{\delta}_i$ is a g -dimensional vector of the i th rowwise normalized counterpart of δ_i with components formulated as

$$\hat{\delta}_{ij} = \frac{\delta_{ij}}{\sum_k \delta_{ik}}. \quad (12)$$

To cope with the issue of the situation when we have different number of samples for each class, we provide weight v_j to class j that is inversely related to class size, formulated as

$$v_j = (|S_j| + |E_j|)^{-1} \quad (13)$$

where $|S_j|$ is the number of labeled samples and $|E_j|$ is the number of pseudo-labeled samples in class j . To this end, we formulated per-sample and per-class weights. We relate the weighted loss to the labeled and pseudo-labeled samples as follows:

$$\xi_w(X, Y_S, \hat{Y}_E; \Lambda) = \sum_{i=1}^s v_{y_i} \varepsilon_{\text{sup}}(f_\Lambda(x_i), y_i) + \sum_{i=s+1}^n \omega_i v_{\hat{y}_i} \varepsilon_{\text{pseu}}(f_\Lambda(x_i), \hat{y}_i). \quad (14)$$

In fact, (14) is the sum of weighted versions of ξ_{sup} and ξ_{pseu} in (2) and (4), respectively. Iscen *et al.* [24] used one CNN model to produce the pseudo-labels and then used these labels to train the same model. On the contrary of this approach, we are using two CNN models in the form of a teacher model and a student model. The teacher model generates the pseudo-labels, which are combined with the labeled samples to train the student model. Therefore, the trained student model is not biased toward the labeled data. To this end, the student and teacher models are trained in parallel, according to (14), in which \hat{y}_i in the student model comes from the teacher model.

In summary, considering the nearest neighbor graph definition in the form of affinity matrix, label propagation, sample

TABLE I
DIFFERENT WATER AND ICE CLASSES

WMO code	Classes
02	Open Water/ Leads with Water
01–02	Brash/Pancake Ice
83	Young Ice (YI)
86–89	Level first-year ice FYI
95	Old/deformed Ice

and class weights, and label and pseudo-label loss terms, our semisupervised method follows a repetitive procedure. Initially, we randomly initialize all the parameters. We then train the teacher model using the s labeled samples in X_S , considering the supervised loss term. We use the trained teacher model to extract descriptors D for the complete training set X . We then find the k -nearest neighbors of all samples to build the adjacency matrix Θ and carry out label propagation by computing (9). We then assign pseudo-labels to the unlabeled samples in X_E by considering (10). Subsequently, we train both the teacher and student models for one epoch on the complete training set X using the weighted loss function in (14). This process is repeated for T' epochs.

IV. EXPERIMENTAL ANALYSIS

A. SAR-Based Sea Ice Dataset

We have trained our proposed method considering 31 Sentinel-1 images. The images are acquired from the North of Svalbard with $40\text{ m} \times 40\text{ m}$ pixel resolution. They are pre-processed using the ESA SNAP software by applying thermal noise removal, calibration using the σ_0 lookup table, and multilooking using a 3×3 boxcar filter. After converting the intensity images to dB values, they are clipped and scaled linearly in the range $[0, 1]$ considering individual channels. The range in dB for horizontal–horizontal (HH) polarization and HV polarization are $[\text{min}: -30, \text{max}: 0]$ and $[\text{min}: -35, \text{max}: -5]$, respectively.

To create a suitable dataset for sea ice classification, we used labeled polygons generated from 31 Sentinel-1 EW scenes from the North of Svalbard. These polygons were carefully labeled manually according to coregistered optical images with as small as possible time gaps. We used these images for training our proposed method. More details can be found in [39]. The dataset consists of five classes, as shown in Table I.

Nonetheless, to perform sea ice classification and create a proper dataset [55] for deep learning, we extracted patches with size equal to 32×32 pixels, corresponding to a spatial resolution of 1280 m^2 , from inside the polygons, with a stride of 10 pixels. This dataset can be accessed from the link [55]. It is worth mentioning that we analyzed the effect of different patch sizes in a previous work [9]. We found that the validation results got better by increasing the patch size. However, this improvement comes at the cost of a lower spatial resolution as larger patches cover wider areas of the surface. For instance, a larger patch will be classified as water if the majority of the pixels represent water. This would become a significant issue at ice edges as classification based on larger patches would lead to coarser or nonsmooth edges. Therefore, there is a tradeoff

between accuracy and resolution. To compensate for this, in our proposed work, we consider a patch size equal to 32×32 pixels. We extracted two channel patches consisting of HH and HV intensities. It is also worth mentioning that we also analyzed the effect of different channel composition (HH, HV, and incidence angle) in our previous work [9]. We found that adding the HV channel to the HH gives large improvement. However, the improvement resulting from also adding the incidence angle is quite small. In the current work, we do not include the incidence angle as this also enables more proper comparison with other SSL methods [48], [49]. These reference SSL methods largely apply different DA techniques, and the inclusion of the incidence angle is not feasible because of the DA techniques. Therefore, the patches in our work consist of only HH and HV intensities to maintain consistency. In Table I, we provide ice type codes, following the definitions of the World Meteorological Organization [56] and a brief description of each class. We consider binary sea ice classification. The first class, namely, the *water class*, consists of open water and leads with water, and the various ice types are grouped together as the *ice class*. The total number of patches for water is 9317, and for ice, it is 5433. We provided the dataset online [55]. For now, we are interested in analyzing the performance of DNNs for binary classification. Our consideration based on our experience with sea ice classification is that if DNNs can perform well in the binary classification case, they may also classify multiple sea ice types properly.

For validation, we consider some other Sentinel-1 scenes provided by the Norwegian Meteorological Institute [57] from the Danmarkshavn area on the Northeastern coast of Greenland and extract 1516 water patches and 1324 ice patches, mostly from challenging areas. In the first experiment, we consider the training dataset from the North of Svalbard and split it into labeled X_S and unlabeled X_E samples. In the next experiment, to show the capability of the proposed method in classifying real unlabeled data, we consider 5000 random patches picked from the Norwegian Meteorological Institute dataset as the unlabeled dataset X_E and use all samples in the training set as the labeled dataset X_S . We insert a different number of labeled datasets for each class, i.e., 15, 30, 60, 100, 500, and 1000. For the inference results, we apply SAR images from the Norwegian Meteorological Institute dataset [57], which were collected in 2018.

B. Our Model Configurations

We exploit the same network models for the teacher and student models. Similar to [24], we use the network architecture defined in [46] and shown in Table II. We trained the teacher model for 100 epochs in the first training step. In the second step, we trained the teacher model for 200 epochs based on the label propagation to produce pseudo-labels. These labels are then exploited to train the student model concurrently. The learning rate for the teacher model is 0.0008 for the first step and 0.0001 for the second step. The learning rate for the student model during the second step is 0.002. For DA, we used only rotation in both steps to keep the same physical meaning of all

TABLE II
BASE CNN ARCHITECTURE

Layer	Hyperparameters
Input	32×32 patches
Convolutional	128 filters, 3×3
Convolutional	128 filters, 3×3
Convolutional	128 filters, 3×3
Pooling	Max-pooling, 3×3
Dropout	$p = 0.5$
Convolutional	256 filters, 3×3
Convolutional	256 filters, 3×3
Convolutional	256 filters, 3×3
Pooling	Max-pooling, 3×3
Dropout	$p = 0.5$
Convolutional	256 filters, 3×3
Convolutional	256 filters, 3×3
Convolutional	256 filters, 3×3
Pooling	Average pool (6×6 to 1×1)
Softmax	Fully connected 128 to 2

TABLE III
VALIDATION ACCURACY FOR DIFFERENT AMOUNT OF LABELED DATA AND UNLABELED DATA FROM THE TRAINING DATASET

	15	30	40	60	100	500	1000
Fully supervised	39.60	52	55.67	70.35	88.72	91.50	92.06
semi-GANs [48]	-	-	71.96	88.14	89.23	90.5	90.04
MixMatch [50]	-	-	-	86.28	85.88	88.19	91.02
LP-SSL [24]	55.62	75.34	75.23	89.97	90.42	91.73	91.24
TSLP-SSL	88.03	86.96	90.87	90.47	91.21	91.07	91.94

the channels of the SAR data and considered the same values for the hyperparameters as used in the previous studies [24], [58] in all experiments. We run the experiments on a single NVIDIA Quadro RTX 5000 with 16-GB memory. The code is available.¹

C. Results and Discussion

We trained our models with a distinct number of labeled data to assess the performance of our proposed method in comparison with four reference methods. For this purpose, we consider both a supervised CNN model and three semisupervised methods, namely, semi-GANs [47], MixMatch [49], and label propagation model (LP-SSL) [24]. In the supervised CNN model, we consider the same CNN architecture that we use for both our teacher and student models. We present the validation results in Table III in terms of accuracy for both our proposed TSLP-SSL method and the reference methods. In the first experiment, we use our training data and split it into labeled, i.e., X_S , Y_S , and unlabeled datasets, (X_E). For the validation, we use the validation data that were mentioned previously (see Section IV-A). As can be seen in Table III, our proposed method outperforms the fully supervised CNN architecture considering 15, 30, 40, 60, and 100 labeled samples. Similarly, our method also outperforms the semisupervised methods semi-GANs [47], MixMatch [49], and LP-SSL [24] considering different number of labeled datasets except in case of 500 labeled samples. For comprehensive analysis, we also consider other performance metrics, namely, average precision, average recall, and average F1-score, for both the classes: water and ice. We present the results in Table IV. As

can be seen, we also outperform in most cases considering both the supervised and semisupervised methods. In fact, our method learns more information from the unlabeled data, especially when a very limited number of samples are available. In fact, the student model in our approach has the potential to remedy the problem of overfitting of the teacher model when only few samples are available, and it presents comparable validation accuracy when considering 500 and 1000 labeled datasets. However, when the number of labeled datasets increases, the amount of information extracted from the unlabeled data does not significantly improve the results. It is worth noticing that the *good samples* of the labeled data can significantly impact the results in the second step. This can be seen when comparing the results of using 15 and 30 labeled samples in Tables III and IV.

In fact, our proposed method can learn from the unlabeled data and, thus, improves its performance. It even achieves better validation accuracy than the supervised and LP-SSL models considering 15, 30, 40, 60, and 100 labeled samples. In order to explain the behavior of our method considering 500 and 1000 labeled samples, we compute the accuracy of the pseudo-labels from the teacher model during the second step of the training process. This can be done since the ground-truth labels of the unlabeled data can be extracted from the training dataset. We consider the comparison of our proposed method with the fully supervised CNN architecture. When both the methods are trained on 500 and 1000 labeled datasets, the accuracy on the pseudo-labels reaches more than 99%, but at the same time, the validation accuracy does not increase, as shown in Table III. This means that there is no more information in the unlabeled data to further improve the validation accuracy considering this particular dataset. We investigate this by training the supervised model with all the data in the training dataset, and it reached a validation accuracy of 91.57%.

We also investigated the inference results on a single-image SAR scene from Danmarkshavn considering 30, 60, and 100 labeled datasets in our proposed TSLP-SSL model. The results of this experiment are reported in Fig. 3, where the first row shows results using the supervised model and the second row shows results using our proposed method. Blue color indicates the water and white color indicates the ice class. As can be seen, our method presents improvement compared to the supervised model, especially in the noisy areas.

D. Feature Separability of Our Proposed Method

Furthermore, we illustrate the capability of the label propagation step that we use to generate the pseudo-labels for training the student model. In fact, label propagation is characterized by consolidated feature separability, which helps generate meaningful pseudo-labels for training the student model. To explain this visually, we extract the feature vector output from the last convolution layer. The dimension of the feature vectors is 128. We transform the feature vectors into three components based on the principal component analysis (PCA), considering both labeled and unlabeled data, to visually understand the feature space. These components are shown in Fig. 4. Fig. 4(a) and (c) shows the feature space when training the teacher model in the

¹<https://github.com/sakh251/TSLP-SSL>

TABLE IV
AVERAGE OF PRECISION, RECALL, AND F1-SCORE FOR DIFFERENT AMOUNT OF LABELED DATA AND UNLABELED DATA FROM THE TRAINING DATASET

	15			30			40			60			100			500			1000		
	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
Fully supervised	.5623	.5981	.3645	.6064	.7100	.4948	.5991	.7016	.5123	.6259	.7452	.6163	.7812	.6977	.7283	.8860	.8036	.8336	.8955	.8154	.8480
semi-GANs [48]	-	-	-	-	-	-	.4278	.4295	.4286	.7951	.6709	.7086	.7914	.7192	.7474	.8957	.7602	.8059	.8506	.7825	.8103
MixMatch [50]	-	-	-	-	-	-	-	-	-	.9160	.5263	.5041	.9137	.5108	.4740	.8804	.6848	.7319	.9089	.7665	.8143
LP-SSL [24]	.5963	.6922	.4921	.6392	.7474	.6493	.6611	.8038	.6674	.7971	.7738	.7847	.8041	.7949	.7990	.9154	.7866	.8324	.9074	.7751	.8211
TSLP-SSL	.7591	.7927	.7741	.7314	.7345	.7329	.8299	.7640	.7914	.8599	.7546	.7925	.8452	.7609	.7990	.8674	.8070	.8326	.9007	.8062	.8432

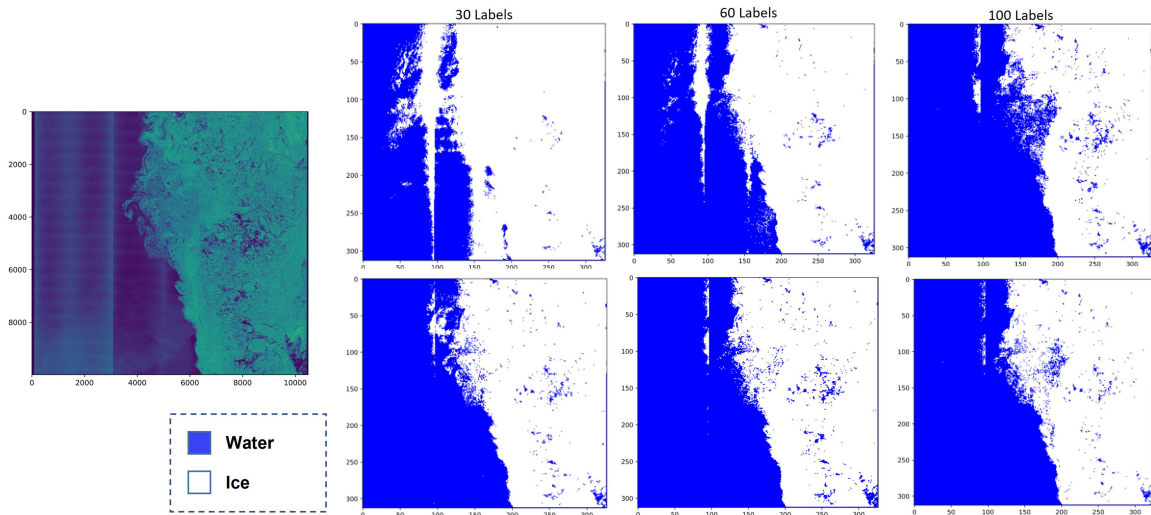


Fig. 3. Inference results. We present qualitative results of a single input image. The first row depicts the results considering supervised deep learning, and the second row depicts the results using our proposed TSLP-SSL model.

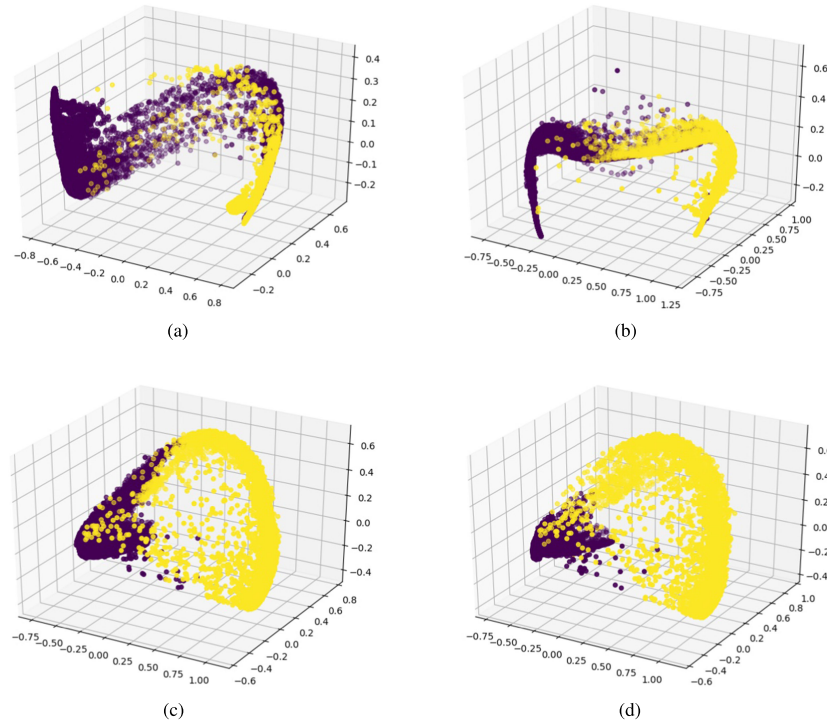


Fig. 4. Three PCA components' visualization of extracted features (flattened vector after convolution layers with 128 values) from labeled and unlabeled data. The yellow color represents water and the purple color represents ice. (a) and (c) show the supervised feature space from first step with 60 and 1000 labeled data, respectively. (b) and (d) show the best feature space of second step with 60 and 1000 labeled data, respectively.

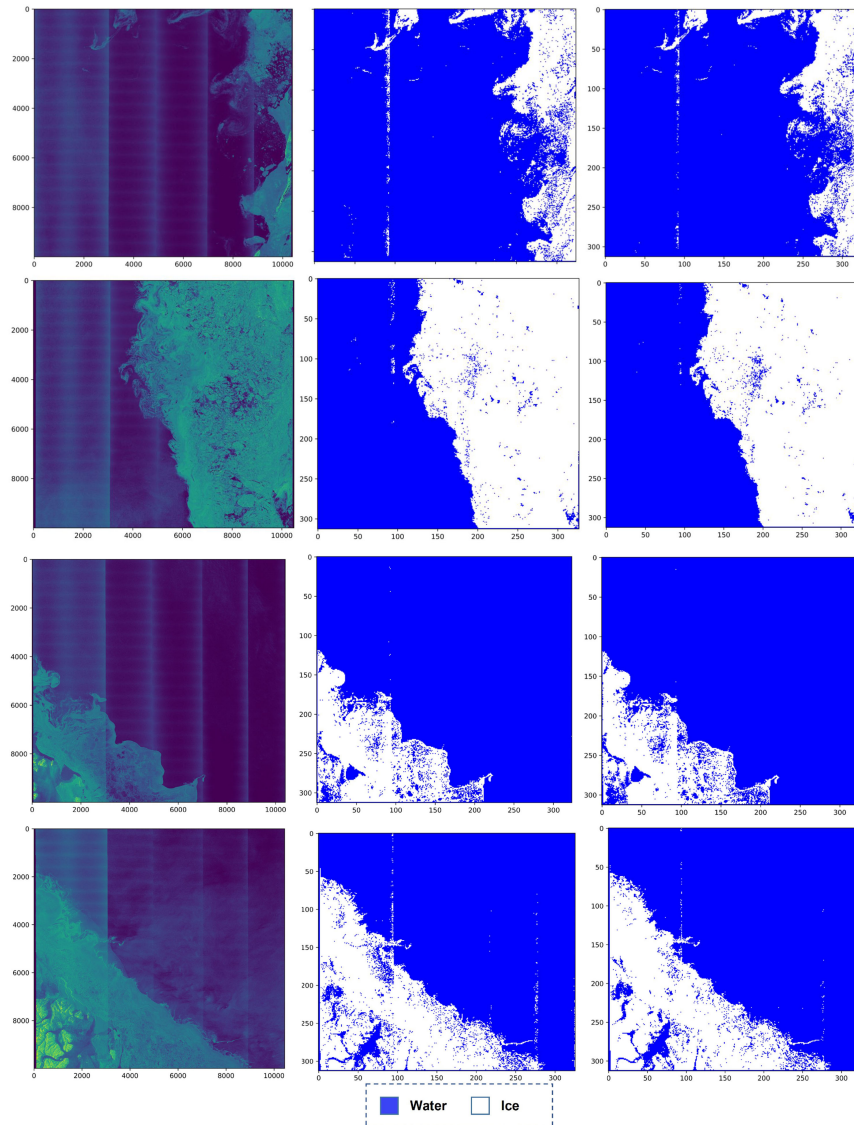


Fig. 5. Inference results. The first column shows input images, the second column shows the results obtained with supervised deep learning, and the third row shows results obtained with our TSLP-SSL model, which is trained by also taking into account unlabeled data from other images.

first step considering 60 and 1000 labeled samples, respectively. Fig. 4(b) and (d) shows the feature space representation after label propagation is applied in the second step. The yellow circles represent water and the purple circles represent the ice class. As can be seen, label propagation leads to more separable classes in the feature space, especially when 1000 labeled samples are considered. Therefore, through label propagation, the unlabeled data help to build a more class-separable feature space and generate more meaningful and informative pseudo-labels to train the student model.

E. Extended Unlabeled Data

To elaborate a bit more on the capability of our proposed method, we conduct another experiment. We evaluate the validation accuracy of the proposed method by considering 1000 data samples from the training dataset as labeled data (i.e., considering it as an element of X_S) and adding unlabeled data

TABLE V
VALIDATION ACCURACY, AVERAGE PRECISION, AVERAGE RECALL, AND AVERAGE F1-SCORE CONSIDERING ADDITIONAL REAL UNLABELED DATA

	Acc.	Pre.	Rec.	F1
Fully supervised	92.06	0.8955	0.8154	0.8480
semi-GANs [48]	89.22	0.8730	0.7264	0.7716
MixMatch [50]	89.55	0.8788	0.7345	0.7800
LP-SSL [24]	91.19	0.9243	0.7626	0.8144
TSLP-SSL	92.93	0.9291	0.8182	0.8606

not contained in the training dataset. For this purpose, we extract 5000 random patches from the Danmarkshavn data and add to the training process in the second step X_E . We present the performance of all the methods in Table V in terms of accuracy, average precision, average recall, and average F1-scores. As can be seen, our method performs better than the fully supervised CNN method and three semisupervised methods: semi-GANs [47], MixMatch [49], and LP-SSL [24]. These results demonstrate

that our proposed method can extract and use relevant information from real unlabeled data and learn new information from unseen and unlabeled data. This is a useful and powerful capability that can be beneficial in sea ice classification, where the amount of available training data is limited.

We also present inference results using four different images from the Danmarkshavn data considering the student model trained on 1000 labeled datasets and extended with unlabeled data. In Fig. 5, the left column depicts the original SAR images, the middle column presents the inference results obtained with the supervised learning model, and the last column shows the results obtained with our proposed TSLP-SSL method. Water is highlighted in blue color and ice is highlighted in white color. These inference results again show the capability of our proposed semisupervised method in using the information of unlabeled data.

V. CONCLUSION

In this article, we proposed a teacher–student-based label propagation method for sea ice classification. The teacher model and the student model were trained in an iterative way during the training stage. The teacher model produced features that were extracted before the final classification layer. These features were used during the label propagation process. Considering the unlabeled data, the labels were propagated to produce pseudo-labels. Subsequently, the pseudo-labels from the teacher models were fed to the student model during the training to find an unbiased decision boundary. Our method outperformed the supervised CNN and the semisupervised LP-SSL models. We presented both qualitative and quantitative results for our proposed method and the reference methods. Our proposed method considered a very limited number of labeled samples starting from 15 samples and unlabeled samples to train the models efficiently. In fact, our proposed method was characterized by the ability to learn useful information from both labeled and unlabeled data. Our method reduced the dependence on labeled samples, which is very time consuming and costly to collect for sea ice analysis. Therefore, this property of our method makes it a good fit for the community of sea ice analysis, where limited labeled data are available. We have also shown that by adding more unlabeled samples, the performance of the inference results has improved. Considering the semisupervised aspect, our method can be extended to other problem areas, where a very limited number of labeled samples are available since we coped with the biasing and dependence issues related to the labeled samples.

The dataset we collected consists of different ice types. However, the number of samples for each ice type is limited. Considering the promising performance of our proposed method for binary sea ice classification, in our future work, we would adopt and extend our method to ice type classification.

ACKNOWLEDGMENT

This work is funded in part by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and

the Research Council of Norway (RCN Grant no. 237906), the European Union’s Horizon 2020 research and innovation programme ExtremeEarth project, grant agreement no. 825258 (<http://earthanalytics.eu/>) and by the Fram Center under the Automised Large-scale Sea Ice Mapping (ALSIM) “Polhavet” flagship project.

REFERENCES

- [1] L. P. Bobylev and M. W. Miles, “Sea ice in the Arctic paleoenvironments,” in *Sea Ice in the Arctic*. Berlin, Germany: Springer, 2020, pp. 9–56.
- [2] T. Vihma, “Effects of Arctic Sea ice decline on weather and climate: A review,” *Surv. Geophys.*, vol. 35, no. 5, pp. 1175–1214, 2014.
- [3] M. R. Najafi, F. W. Zwiers, and N. P. Gillett, “Attribution of Arctic temperature change to greenhouse-gas and aerosol influences,” *Nat. Climate Change*, vol. 5, no. 3, pp. 246–249, 2015.
- [4] J. C. Stroeve, M. C. Serreze, M. M. Holland, J. E. Kay, J. Malanik, and A. P. Barrett, “The Arctic’s rapidly shrinking sea ice cover: A research synthesis,” *Climatic Change*, vol. 110, no. 3, pp. 1005–1027, Feb. 2012.
- [5] S. Haykin, E. O. Lewis, R. K. Raney, and J. R. Rossiter, *Remote Sensing of Sea Ice and Icebergs*, vol. 13. Hoboken, NJ, USA: Wiley, 1994.
- [6] A. Cristea, J. van Houtte, and A. P. Doulgeris, “Integrating incidence angle dependencies into the clustering-based segmentation of SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2925–2939, 2020.
- [7] M. Ghanbari, D. A. Clausi, L. Xu, and M. Jiang, “Contextual classification of sea-ice types using compact polarimetric SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7476–7491, Oct. 2019.
- [8] J. L. Awange and J. B. K. Kiema, “Microwave remote sensing,” in *Environmental Geoinformatics*. Berlin, Germany: Springer, 2013, pp. 133–144.
- [9] 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 Sens.*, vol. 13, no. 9, 2021, Art. no. 1734.
- [10] L. Wang, K. Scott, and D. Clausi, “Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network,” *Remote Sens.*, vol. 9, no. 5, 2017, Art. no. 408.
- [11] W. Wen, J. Timmermans, Q. Chen, and P. M. van Bodegom, “A review of remote sensing challenges for food security with respect to salinity and drought threats,” *Remote Sens.*, vol. 13, no. 1, 2021, Art. no. 6.
- [12] D. H. Svendsen, M. Piles, J. Muñoz-Marí, D. Luengo, L. Martino, and G. Camps-Valls, “Integrating domain knowledge in data-driven Earth observation with process convolutions,” *IEEE Trans. Geosci. Remote Sens.*, 2021.
- [13] L. Hashemi-Beni and A. A. Gebrehiwot, “Flood extent mapping: An integrated method using deep learning and region growing using UAV optical data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2127–2135, Jan. 2021.
- [14] N. Asadi, K. A. Scott, A. S. Komarov, M. Buehner, and D. A. Clausi, “Evaluation of a neural network with uncertainty for detection of ice and water in SAR imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 247–259, Jan. 2021.
- [15] J. Lohse, A. P. Doulgeris, and W. Dierking, “An optimal decision-tree design strategy and its application to sea ice classification from SAR imagery,” *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1574.
- [16] J. Lohse, A. Doulgeris, and W. Dierking, “Mapping sea-ice types from Sentinel-1 considering surface-type dependent effect of incidence angle,” *Ann. Glaciol.*, vol. 61, no. 83, pp. 260–270, 2020.
- [17] 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.
- [18] G.-J. Qi and J. Luo, “Small data challenges in Big Data Era: A survey of recent progress on unsupervised and semi-supervised methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: [10.1109/TPAMI.2020.3031898](https://doi.org/10.1109/TPAMI.2020.3031898).
- [19] E. Dalsasso, L. Denis, and F. Tupin, “SAR2SAR: A semi-supervised despeckling algorithm for SAR images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4321–4329, 2021.
- [20] X. Jiang, G. Li, X.-P. Zhang, and Y. He, “A semisupervised Siamese network for efficient change detection in heterogeneous remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2021.3061686](https://doi.org/10.1109/TGRS.2021.3061686).

- [21] Y. Ding, X. Zhao, Z. Zhang, W. Cai, N. Yang, and Y. Zhan, "Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3100578.
- [22] M. Douze, A. Szlam, B. Hariharan, and H. Jégou, "Low-shot learning with large-scale diffusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3349–3358.
- [23] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2004, pp. 321–328.
- [24] A. Iscen, G. Toliás, Y. Avrithis, and O. Chum, "Label propagation for deep semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5065–5074.
- [25] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Sea ice classification based on neural networks method using Sentinel-1 data," *Int. Multidisciplinary Sci. GeoConf.*, vol. 19, no. 2.2, pp. 617–623, 2019.
- [26] N. Zakhvatkina, V. Smirnov, and I. Bychkova, "Satellite SAR data-based sea ice classification: An overview," *Geosciences*, vol. 9, no. 4, 2019, Art. no. 152.
- [27] J. Karvonen, "Baltic sea ice concentration estimation using SENTINEL-1 SAR and AMSR2 microwave radiometer data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2871–2883, May 2017.
- [28] F. Dinessen, "Operational multisensor sea ice concentration algorithm utilizing SENTINEL-1 and AMSR2 data," in *Proc. 19th EGU General Assembly*, 2017, Art. no. 19037.
- [29] A. M. Johansson, C. Brekke, G. Spreen, and J. A. King, "X-, C-, and L-band SAR signatures of newly formed sea ice in Arctic leads during winter and spring," *Remote Sens. Environ.*, vol. 204, pp. 162–180, 2018.
- [30] A. S. Fors, C. Brekke, A. P. Doulgeris, T. Eltoft, A. H. Renner, and S. Gerland, "Late-summer sea ice segmentation with multi-polarisation SAR features in C and X band," *Cryosphere*, vol. 10, no. 1, pp. 401–415, 2016.
- [31] M. Dabboor, B. Montpetit, and S. Howell, "Assessment of the high resolution SAR mode of the RADARSAT constellation mission for first year ice and multiyear ice characterization," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 594.
- [32] D.-B. Hong and C.-S. Yang, "Automatic discrimination approach of sea ice in the arctic ocean using SENTINEL-1 extra wide swath dual-polarized SAR data," *Int. J. Remote Sens.*, vol. 39, no. 13, pp. 4469–4483, 2018.
- [33] A. M. Johansson *et al.*, "Consistent ice and open water classification combining historical synthetic aperture radar satellite images from ERS-1/2, ENVISAT ASAR, RADARSAT-2 and sentinel-1A/B," *Ann. Glaciology*, vol. 61, no. 82, pp. 40–50, 2020.
- [34] W. Aldenhoff, C. Heuzé, and L. E. Eriksson, "Comparison of ice/water classification in Fram Strait from C-and L-band SAR imagery," *Ann. Glaciol.*, vol. 59, no. 76pt2, pp. 112–123, 2018.
- [35] 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.
- [36] Y. Gao, F. Gao, J. Dong, and S. Wang, "Transferred deep learning for sea ice change detection from synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1655–1659, Oct. 2019.
- [37] M. Tom, R. Aguilar, P. Imhof, S. Leinss, E. Baltsavias, and K. Schindler, "Lake ice detection from SENTINEL-1 SAR with deep learning," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 409–416, 2020.
- [38] 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, Art. no. 2165.
- [39] J. Lohse, A. P. Doulgeris, and W. Dierking, "Mapping sea-ice types from SENTINEL-1 considering the surface-type dependent effect of incidence angle," *Ann. Glaciol.*, vol. 61, no. 83, pp. 260–270, 2020.
- [40] M. Shin, "Semi-supervised learning with a teacher-student network for generalized attribute prediction," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 509–525.
- [41] Q. She, J. Zou, M. Meng, Y. Fan, and Z. Luo, "Balanced graph-based regularized semi-supervised extreme learning machine for EEG classification," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 4, pp. 903–916, 2021.
- [42] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.
- [43] K. Yu, H. Ma, T. R. Lin, and X. Li, "A consistency regularization based semi-supervised learning approach for intelligent fault diagnosis of rolling bearing," *Measurement*, vol. 165, 2020, Art. no. 107987.
- [44] J. Gordon and J. M. Hernández-Lobato, "Combining deep generative and discriminative models for Bayesian semi-supervised learning," *Pattern Recognit.*, vol. 100, 2020, Art. no. 107156.
- [45] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves ImageNet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10 687–10 698.
- [46] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.
- [47] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
- [48] D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 3581–3589.
- [49] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Proc. 33rd Conf. Neural Inf. Process. Syst.*, 2019, pp. 5050–5060.
- [50] Y. Han *et al.*, "A cooperative framework based on active and semi-supervised learning for sea ice classification using EO-1 hyperion data," *Trans. Japan Soc. Aeronaut. Space Sci.*, vol. 62, no. 6, pp. 318–330, 2019.
- [51] 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, School Elect. Eng. Comput. Sci., KTH Roy. Inst. Technol., Stockholm, Sweden, 2020.
- [52] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 28–35.
- [53] G. Algan and I. Ulusoy, "Image classification with deep learning in the presence of noisy labels: A survey," *Knowl.-Based Syst.*, vol. 215, 2021, Art. no. 106771.
- [54] M. Ren, W. Zeng, B. Yang, and R. Urtasun, "Learning to reweight examples for robust deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4334–4343.
- [55] S. Khaleghian, J. P. Lohse, and T. Krämer, "Synthetic-aperture radar (SAR) based ice types/ice edge dataset for deep learning analysis," 2020. [Online]. Available: <https://doi.org/10.18710/QAYI40>
- [56] J. Falkingham and V. Smolyanitsky, "Electronic chart systems ice objects catalogue," Version 5.1. draft for approval. Feb. 2012. [Online]. Available: <http://hdl.handle.net/11329/403>
- [57] N. Hughes, "Extremearth polar use case training data," 2020. [Online]. Available: <https://zenodo.org/record/3695276#.X-ytf2j0mUn>
- [58] A. Iscen, G. Toliás, Y. Avrithis, T. Furon, and O. Chum, "Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2077–2086.



Salman Khaleghian received the bachelor's degree in applied mathematics from Shahed University, Tehran, Iran, in 2006, and the M.S. degree in computer software engineering from Science and Research branch of Azad University, Tehran, Iran, in 2010. He is currently working toward the Ph.D. degree in scalable computing for Earth observation with the Center for Integrated Remote Sensing and Forecasting for Arctic Operations, Faculty of Science and Technology, University of Tromsø—The Arctic University of Norway, Tromsø, Norway, and the SIR-IUS Lab, Department of Informatics, University of Oslo, Oslo, Norway.

His research interests include machine learning, deep learning, scalable deep learning, and computer vision.



Habib Ullah received the M.S. degree in electronics and computer engineering from Hanyang University, Seoul, South Korea, in 2009, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, in 2015.

From 2015 to 2016, he was an Assistant Professor of Electrical Engineering with COMSATS University, Islamabad, Pakistan. From 2016 to 2020, he was an Assistant Professor with the College of Computer Science and Engineering, University of Ha'il, Ha'il, Saudi Arabia. In 2020, he was a Postdoctoral Researcher with the University of Tromsø—The Arctic University of Norway, Tromsø, Norway. He is currently an Associate Professor with the Norwegian University of Life Sciences, Ås, Norway. His research interests include computer vision and machine learning.



Thomas Kræmer received the M.Sc. degree in data analysis and sensor technology in 2011 from the University of Tromsø—the Arctic University of Norway, Tromsø, Norway, where he is currently working toward the Ph.D. degree.

Since 2016, he has been the Head Engineer with the Earth Observation Laboratory, University of Tromsø—The Arctic University of Norway. His research interests include algorithms for automated analysis of synthetic aperture radar images for sea ice applications.



Torbjørn Eltoft (Member, IEEE) received the M.Sc. and Ph.D. degrees in physics from University of Tromsø, Norway, in 1981 and 1984, respectively.

In 1988, he joined the Faculty of Science and Technology, University of Tromsø (UiT)—The Arctic University of Norway, Tromsø, Norway, where he is currently a Professor of Remote Sensing with the Department of Physics and Technology. He is also the Director of the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations. He has a significant publication record. His research interests

include signal and image analysis, statistical modeling, and machine learning with applications in synthetic aperture radar and ocean color remote sensing.

Dr. Eltoft was the co-recipient of the 2000 Outstanding Paper Award in Neural Networks awarded by the IEEE Neural Networks Council and an honorable mention for the 2003 *Pattern Recognition* Journal Best Paper Award. He was the recipient of the UiT Award for Research and Development in 2017. He was an Associate Editor for *Pattern Recognition* from 2005 to 2011 and a Guest Editor for *Remote Sensing* on the Special Issue for the PolInSAR 2017 Conference.



Andrea Marinoni (Senior Member, IEEE) received the B.S., M.Sc. (*cum laude*), and Ph.D. degrees in electronic engineering from the University of Pavia, Pavia, Italy, in 2005, 2007, and 2011, respectively.

He is currently an Associate Professor with the Earth Observation Group, Centre for Integrated Remote Sensing and Forecasting for Arctic Operations, Department of Physics and Technology, University of Tromsø—The Arctic University of Norway, Tromsø, Norway, and a Visiting Academic Fellow with the Department of Engineering, University of Cambridge, Cambridge, U.K. From 2015 to 2017, he was a Visiting Researcher with the Earth and Planetary Image Facility, Ben-Gurion University of the Negev, Be'er Sheva, Israel; the School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China; the School of Computer Science, Fudan University, Shanghai, China; the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China; and the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal. From 2013 to 2018, he was a Research Fellow with the Telecommunications and Remote Sensing Laboratory, Department of Electrical, Computer and Biomedical Engineering, University of Pavia. In 2009, he was a Visiting Researcher with the Communications Systems Laboratory, Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, USA. In 2020 and 2021, he was a Visiting Professor with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia. His research interests include efficient information extraction from multimodal remote sensing, nonlinear signal processing applied to large-scale heterogeneous records, Earth observation interpretation and Big Data mining, and analysis and management for human–environment interaction assessment.

Dr. Marinoni was the recipient of the two-year “Applied Research Grant,” sponsored by the Region of Lombardy, Italy, and STMicroelectronics N.V., in 2011; the INROAD Grant, sponsored by the University of Pavia and Fondazione Cariplo, Italy, for supporting excellence in design of ERC proposal in 2017; the “Progetto professionalità Ivano Becchi” grant funded by the Fondazione Banco del Monte di Lombardia, Italy, and sponsored by University of Pavia and NASA Jet Propulsion Laboratory, Pasadena, CA, for supporting the development of advanced methods of air pollution analysis by remote sensing data investigation, in 2018; and the Åsgard Research Program and Åsgard Recherche+ Program grants funded by the Institut Français de Norvège, Oslo, Norway, in 2019 and 2020, respectively, for supporting the development of scientific collaborations between French and Norwegian Research Institutes. He is the Founder and the Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Norway Chapter. He is also an Ambassador of the IEEE Region 8 Humanitarian activities, and a research contact point for the Norwegian Artificial Intelligence Research Consortium. He is a topical Associate Editor of machine learning for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor for three special issues on multimodal remote sensing and sustainable development of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He is the Leader of the GR4S Committee of the IEEE GRSS, coordinating the organization of schools and workshops sponsored by the IEEE GRSS worldwide.