Uncertainty-Incorporated Ice and Open Water Detection on Dual-Polarized SAR Sea Ice Imagery

Xinwei Chen, K. Andrea Scott, *Member, IEEE*, Linlin Xu, *Member, IEEE*, Mingzhe Jiang, *Graduate Student Member, IEEE*, Yuan Fang, and David A. Clausi, *Senior Member, IEEE*

***Abstract*— Algorithms designed for ice–water classification of synthetic aperture radar (SAR) sea ice imagery produce only binary (ice and water) output typically using manually labeled samples for assessment. This is limiting because only a small subset of labeled samples are used, which, given the nonstationary nature of the ice and water classes, will likely not reflect the full scene. To address this, we implement a binary ice–water classification in a more informative manner considering the uncertainty associated with each pixel in the scene. To accomplish this, we have implemented a Bayesian convolutional neural network (CNN) with variational inference to produce both aleatoric (data-based) and epistemic (model-based) uncertainty. This valuable information provides feedback as to regions that have pixels more likely to be misclassified and provides improved scene interpretation. Testing was performed on a set of 21 RADARSAT-2 dual-polarization SAR scenes covering a region in the Beaufort Sea captured regularly from April to December. The model is validated by demonstrating: 1) a positive correlation between misclassification rate and model uncertainty and 2) a higher uncertainty during the melt and freeze-up transition periods, which are more challenging to classify. By incorporating the iterative region growing with semantics (IRGS) segmentation algorithm and an uncertainty value-based thresholding algorithm, the Bayesian CNN classification outputs are improved significantly via both numerical analysis and visual inspection.**

***Index Terms*— Classification, ice maps, iterative region growing with semantics (IRGS), segmentation, uncertainty maps.**

I. Introduction

THE monitoring of Arctic sea ice plays a crucial role in ship navigation, the safety of transportation in northern communities, and understanding climate change. Over the past few decades, different types of satellite remote sensing data have been employed for sea ice monitoring, such as optical [1], infrared [2], and microwave (passive and active) [3], [4]. Among these, the spaceborne synthetic aperture radar (SAR) is advantageous due to its high spatial resolution, polarimetric capability, and flexible imaging modes [5], [6]. Since the microwave energy radiated and received by SAR is able to pass through clouds, the measurements are highly independent of atmospheric moisture [7]. In addition, as the imaging mechanism is triggered by surface roughness and subsurface physical properties, different ice types can be distinguished. Starting from 1978, more than a dozen satellite-based SAR observing systems have been developed by different countries that are useful for sea ice monitoring.

Among different sea ice monitoring tasks (e.g., ice concentration measurements and classification of sea ice types), the discrimination of open water from sea ice provides key information to both marine forecasting on synoptic scales and longer climate-scale simulations [8]. In the past two decades, multiple machine learning (ML) and deep learning (DL)-based methods, which can map high-dimensional data to an array of outputs, have been proposed for automated ice–water classification using dual-polarized SAR imagery. Methods include neural networks (NNs) [9], Markov random fields (MRFs) [10], support vector machines (SVMs) [11], [12], random forests [13], [14], [15], conditional random fields (CRFs) [16], and convolutional NNs (CNNs) with different structures [17], [18], [19], [20], [21], [22]. Besides, in recent years, CNNs have also been widely applied to sea ice concentration (SIC) estimation [23], [24], [25], [26]. Although relatively high classification accuracies have been reported in these studies, the predictions are not always dependable due to the ambiguous correlations between radar backscatter and ice/water features, especially during the melt season where ice and water can be difficult to distinguish based on visual interpretation. In fact, only three of the studies mentioned above [11], [16], [21] have considered data obtained during the melt season. Wind roughened area of water bodies can often be misclassified as ice as well [27]. In addition, the performance evaluation of the model on the whole scene is normally limited to visual inspection and comparison with the sea ice chart. Therefore, to obtain more reliable and in-depth interpretation of classification results, investigating the uncertainty in the model prediction is warranted.

In Bayesian modeling, predictive uncertainty can be decomposed into two different sources, namely, aleatoric and epistemic uncertainties. Aleatoric uncertainty (also known as data uncertainty) arises from the inherent variability of the data distribution, which is irreducible [28]. In contrast, epistemic uncertainty (also known as model uncertainty) refers to uncertainty caused by the model, which can in principle be reduced based on additional information [29]. In general, the uncertainty decomposition and quantification help us locate errors

Manuscript received 10 September 2022; revised 3 December 2022; accepted 29 December 2022. Date of publication 3 January 2023; date of current version 13 January 2023. This work was supported in part by NSERC Discovery Grant and in part by NSERC Alliance Funding. *(Corresponding author: Xinwei Chen.)*

The authors are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: xinwei.chen@uwaterloo.ca; ka3scott@uwaterloo.ca; l44xu@uwaterloo.ca; m63jiang@uwaterloo.ca; yuan.fang@uwaterloo.ca; dclausi@uwaterloo.ca).

Digital Object Identifier 10.1109/TGRS.2022.3233871

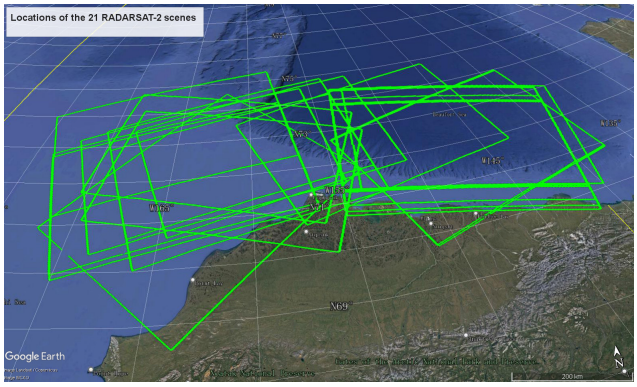

Fig. 1. Locations of the 21 RADARSAT-2 scenes used in this study.

inherent in the observations and evaluate the confidence of the model when making predictions for different scenarios, which has been applied in multiple remote sensing tasks [30], [31]. Uncertainties are also especially important when Earth observation (EO)-derived estimates are assimilated to numerical ice or weather models. For operational sea ice mapping, the uncertainties may be useful to flag regions in the predictions that should be checked manually by an analyst and reduce misclassification rates [32].

In this article, a model based on a Bayesian CNN is proposed to classify sea ice and water in dual-polarized SAR imagery while providing uncertainty maps along with predictions. Iterative region growing with semantics (IRGS) algorithm and an uncertainty value thresholding algorithm are later incorporated to further improve the classification results. The rest of this article is organized as follows. The SAR imagery dataset used in this work is introduced in Section II. Section III illustrates the Bayesian CNN model for predictions and uncertainty quantification, together with the IRGS algorithm for image segmentation. Experimental results with analysis, as well as the uncertainty value thresholding algorithm for reducing misclassification rates, are presented in Section IV. Finally, conclusion and future work of this research are given in Section V.

II. Data Overview

The dataset used in this study for model training and testing is provided by MacDonald, Dettwiler and Associates (MDA) Ltd. It consists of 21 scenes obtained from the C-band SAR satellite RADARSAT-2 in the Beaufort Sea, north of Alaska, as shown in Fig. 1. They were all captured in the ScanSAR wide mode with HH- and HV-polarizations being provided and an incidence angle ranging from 20° to 49°. Since the images were collected from April to December in the year 2010, the model and results can be evaluated under a variety of conditions (e.g., different seasons and ice concentrations). The pixel spacing is 50 m and the size of each image is around 10 000 × 10 000 pixels. To improve computational efficiency, the original images are downsampled using 4 × 4 block averaging. The 4 × 4 pixel block size is selected because it reaches an optimal balance between mapping resolution and operational efficiency. Nevertheless, a block with a similar size (e.g., 3 × 3 or 5 × 5) can also achieve similar results. The

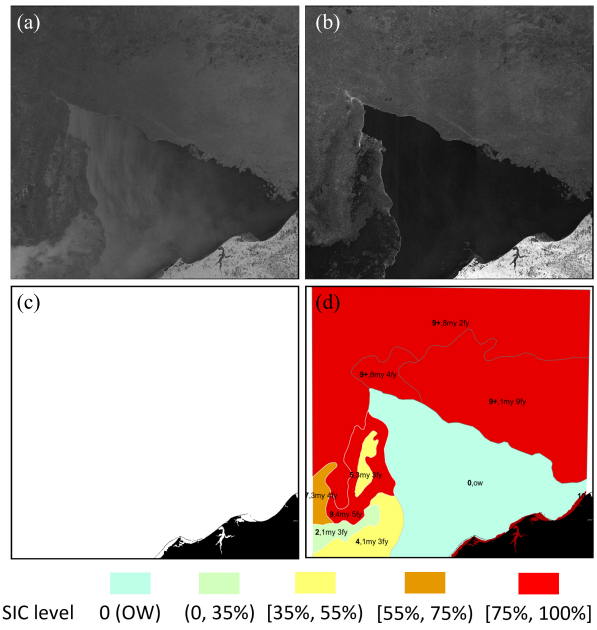

Fig. 2. Example of the RADARSAT-2 SAR scene (scene of data: 20100623) in (a) HH and (b) HV channels. (c) Land mask of the scene. (d) Image analysis chart of the scene. Different colors indicate different SIC levels. For example, the light blue and red represent open water and 90%+ SIC, respectively.

image analysis charts that cover the locations of SAR imagery provided by a Canada Ice Service (CIS)-trained analyst are used to obtain a reference for the labeling of training samples. For each scene, a land mask is provided to exclude pixels of the land from the study. An example of an SAR scene in HH and HV channels with its land mask and ice chart is shown in Fig. 2. More information about the dataset can be found in prior publications [21], [33], [34].

III. Methodology

The flowchart in Fig. 3 gives an overview of the proposed method. First, the pixel-based predictions from the Bayesian CNN are combined with the segmentation results performed by the IRGS algorithm to produce polygon-based ice–water classification results via the majority voting scheme. Then, by analyzing the uncertainty maps, an uncertainty value thresholding algorithm is proposed to reduce misclassification rates and provide more informative scene predictions. In this section, the Bayesian CNN and the IRGS algorithm will be introduced, while the uncertainty value thresholding algorithm will be described in Section IV after the uncertainty analysis.

A. Bayesian CNN With Variational Inference

Previous studies have shown that given an NN, model uncertainty can be obtained by representing the network weights probabilistically. These models are called Bayesian NNs (BNNs) [35]. To make these models operationally practical, several techniques have been incorporated into the training of a BNN, such as variational inference (VI) [36], local reparameterization trick (LRT) [37], and Monte Carlo (MC) estimates. Based on these techniques, the Bayesian CNN was first developed by Gal and Ghahramani [38], which models a

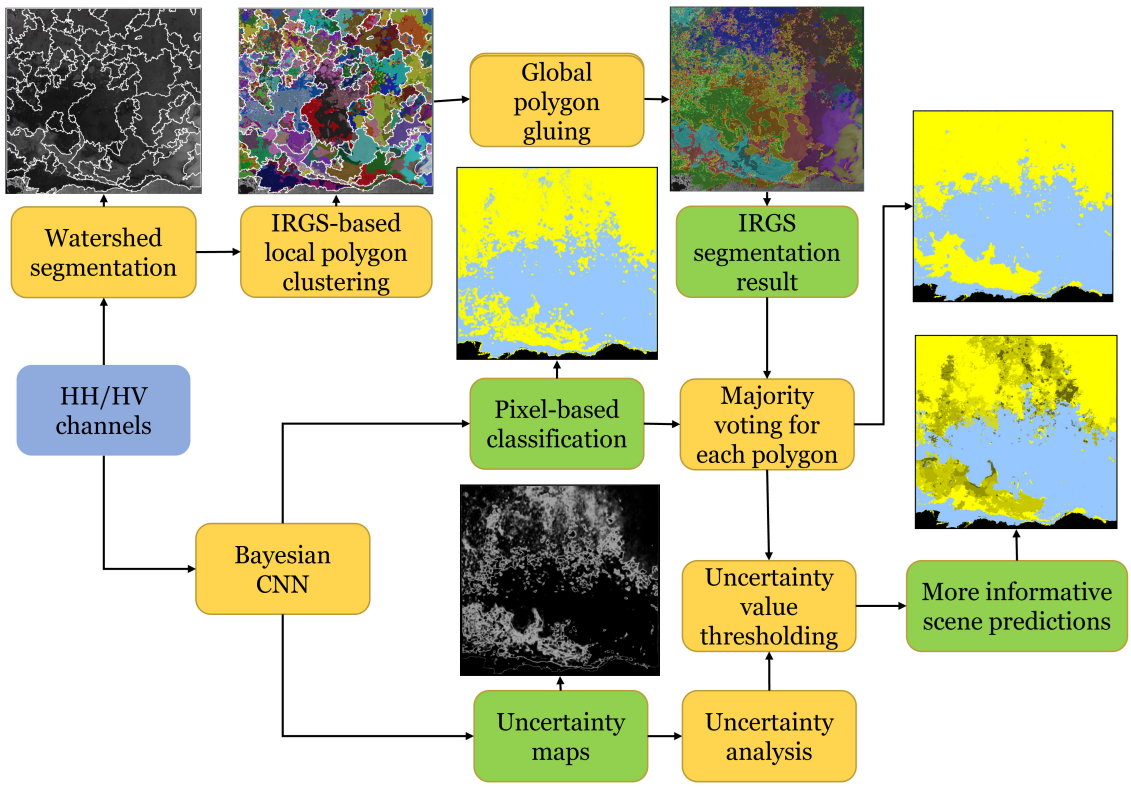

Fig. 3. Flowchart of the proposed method for uncertainty-incorporated ice and water detection on SAR imagery.

distribution over each kernel. The LRT was then introduced to Bayesian CNN by Shridhar et al. [39] to make it more computable. Here, the main techniques and workflow of our classification model based on the Bayesian CNN will be described.

Suppose that the weights in our Bayesian CNN model are denoted as $\mathbf{w}$, where w_j represents each individual weight in $\mathbf{w}$. The modeling of distributions over $\mathbf{w}$ refers to the learning of the posterior probability distribution $p(\mathbf{w}|\mathcal{D})$, where $\mathcal{D}$ represents a set of training examples with image patches and their corresponding labels. However, due to the large number of weights, $p(\mathbf{w}|\mathcal{D})$ is intractable. Therefore, the VI technique is applied, which uses an approximate distribution $q_\theta(\mathbf{w}|\mathcal{D})$ to be as similar as possible to $p(\mathbf{w}|\mathcal{D})$ by minimizing their Kullback–Leibler (KL) divergence. In our Bayesian CNN model, the distributions of $\mathbf{w}$ are assumed to conform to Gaussian distributions, which means that $q_\theta(\mathbf{w}|\mathcal{D}) = \prod_j \mathcal{N}(w_j|\mu, \sigma^2)$. Finding the optimized parameters θ_{opt} (i.e., consisting of μ_{opt} and σ^2_{opt}) can be approximated by minimizing the following tractable cost function [40]:

$$\begin{aligned}\mathcal{F}(\mathcal{D},\theta) &\approx \sum_{i=1}^{n} \log q_\theta(\mathbf{w}^{(i)}|\mathcal{D}) - \log p(\mathbf{w}^{(i)}) - \log p(\mathcal{D}|\mathbf{w}^{(i)}) \\ &\approx \sum_{i=1}^{n} \log \prod_j \mathcal{N}\left(w_j^{(i)}|\mu,\sigma^2\right) + \log \prod_j \mathcal{N}\left(w_j^{(i)}|0,\sigma_p^2\right) \\ &\quad - \log p(\mathcal{D}|\mathbf{w}^{(i)}) \end{aligned} \tag{1}$$

where $\mathbf{w}^{(i)}$ denotes the ith MC sample of weights drawn from the variational posterior $q_\theta(\mathbf{w}|\mathcal{D})$ and n is the number of draws. As illustrated in [40], to introduce $L2$ regularization, a Gaussian prior is placed on $p(\mathbf{w}^{(i)})$ and the distribution of each weight can be expressed as $p(w_j^{(i)}) \sim \mathcal{N}(0, \sigma_0^2)$. The final term is the likelihood term. Following the research by Shridhar et al. [41] to calculate the activations of each layer in the Bayesian CNN, the LRT proposed by Kingma and Welling [42] is implemented to sample layer activations directly, which is more computationally efficient than sampling on the weights. Thus, the convolution operation is represented in a probabilistic form as

$$b = A * W_\mu + \epsilon \odot \sqrt{A^2 * W_\sigma^2} \tag{2}$$

where A represents a certain input receptive field and b represents the corresponding convolution layer activations. The means and variances of the weights in a certain kernel filter are denoted as W_μ and W_σ^2, respectively. The term "$*$" indicates convolution operation, $\epsilon \sim \mathcal{N}(0,1)$, and $\odot$ represents elementwise multiplication. Similarly, the output of the fully connected layer can be sampled by introducing ϵ as well. To ensure that the variance is positive, an unconstrained parameter ρ is used to generate σ as $\sigma = \text{softplus}(\rho)$ [41]. The softplus function is defined as

$$\text{softplus}(\rho) = \frac{1}{\beta} \times \log(1 + \exp(\beta \times \rho)) \tag{3}$$

where β is set to 1 by default.

In this research, the network architecture of the Bayesian CNN used is based on AlexNet proposed by Krizhevsky et al. [43] as it is able to produce accurate classification results with a relatively simple architecture (as shown in Table III).

B. Uncertainty Estimation

Since the output from the Bayesian CNN after the fully connected layer is probabilistic, the ice–water classification output is therefore a predictive distribution. Given a certain sample $\mathbf{w}^{(i)}$, the corresponding predictive distribution can be expressed as $p_{\mathbf{w}^{(i)}}(y^*|x^*)$, where x^* is a new image patch data and y^* is its predicted class. To estimate the expectation of the predictive distribution from all $\mathbf{w}$ drawn from $q_\theta(\mathbf{w}|\mathcal{D})$, $p_{\mathbf{w}}(y^*|x^*)$ should be integrated over the infinite combinations of $\mathbf{w}$. This can be approximated by taking an average on a finite number of samplings, i.e., $\mathbf{w}^{(i)}$. This can be summarized by the following expression:

$$\begin{aligned}\mathbb{E}_{q_\theta(\mathbf{w}|\mathcal{D})}[p_{\mathbf{w}}(y^*|x^*)] &= \int q_\theta(\mathbf{w}|\mathcal{D})\, p_{\mathbf{w}}(y^*|x^*) dw \\ &\approx \frac{1}{T}\sum_{t=1}^{T} p_{\mathbf{w}^{(t)}}(y^*|x^*) \end{aligned} \tag{4}$$

where T is the predefined number of samples. In classification tasks, the overall predictive uncertainty can be obtained from predictive variance $\mathrm{Var}_q(p(y^*|x^*))$. To decompose the overall predictive uncertainty into aleatoric and epistemic uncertainty, two main methods have been proposed [44], [45]. Here, the decomposition method proposed by Kwon et al. [45] is used due to its direct computation from predictive probabilities and direct interpretation of the predictive variance. In contrast, the method proposed by Kendall and Gal [44] requires the perturbation of the logits to compute two types of uncertainty, which may lead to numerical instability [46]. The decomposition of predictive variance [45] can be expressed as

$$\begin{aligned}\mathrm{Var}_q(p(y^*|x^*)) &= \mathbb{E}_q[y^* y^{*T}] - \mathbb{E}_q[y^*]\mathbb{E}_q[y^*]^T \\ &= \underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathrm{diag}(\hat{p}_t) - \hat{p}_t\hat{p}_t^T}_{\text{aleatoric}} \\ &\quad + \underbrace{\frac{1}{T}\sum_{t=1}^{T}(\hat{p}_t - \bar{p})(\hat{p}_t - \bar{p})^T}_{\text{epistemic}} \end{aligned} \tag{5}$$

where $\hat{p}_t = \mathrm{Softmax}(f_{\mathbf{w}^{(t)}}(x^*))$ ($f_{\mathbf{w}^{(t)}}(x^*)$ is the output before the softmax layer) and $\bar{p} = (1/T)\sum_{t=1}^{T}\hat{p}_t$. Let $\mathbf{z} = f_{\mathbf{w}^{(t)}}(x^*)$, and the softmax function can be defined as

$$\mathrm{Softmax}(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{6}$$

where $i = 1, \ldots, K$ and $\mathbf{z} = [z_1, \ldots, z_K]$. K is the number of classes, which is 2 here.

The detailed derivation of the decomposition can be found in the Appendix of Kwon et al. [45]. It can be observed from (5) that each uncertainty can be interpreted meaningfully. For example, in our binary ice–water classification task, if the output probability of being classified as ice and the probability of being classified as water are close to each other, which means that the uncertainty mainly comes from the data itself and will result in relatively large values in the aleatoric part of (5). In contrast, if the model is uncertain about the classification result, its predictive probabilities in each sampling may have a relatively large variance, which leads to relatively large element values in the epistemic part of (5). In addition, to generate uncertainty quantification result for each pixel, the diagonal elements in each part are summed together (nondiagonal terms are either close to zero or negative), which results in one value for aleatoric uncertainty and a separate value for epistemic uncertainty.

C. Combination With IRGS Algorithm

Although the pixel-based classification results can be obtained directly from the Bayesian CNN, they are susceptible to local noise and do not preserve the ice–water boundary information well. Hence, similar to previous implementations [11], [21], [47], the unsupervised IRGS segmentation algorithm proposed by Yu and Clausi [48] and Yu et al. [49] is performed to identify homogeneous regions using an unsupervised hierarchical approach, which will be illustrated in the following. Then, each region is labeled as either ice or water class by taking a majority voting scheme on the Bayesian CNN classification results of all pixels within that region. As shown in the top of Fig. 3, first, the watershed segmentation proposed by Vincent and Soille [50] is applied to the HV channel, which segments the image into a number of autopolygons. For each autopolygon, the IRGS algorithm is then performed on a local scale and clusters the region into six arbitrary classes. As illustrated in [48], IRGS is an extension of the traditional MRF spatial context model in which the objective functions are formulated by gradually increased edge penalty (GIEP) functions. Those objective functions can be optimized by a novel region growing (merging) scheme. Finally, the region merging scheme is conducted on a global scale to the whole scene, which first converts autopolygon boundaries into region boundaries and then merges regions in different autopolygons into 12 arbitrary classes. The implementation of IRGS on both local and global scales can be regarded as a "glocal" segmentation scheme. Nevertheless, since whether each class belongs to either ice or water is still unknown, the Bayesian CNN pixel-based ice–water classification results of the whole scene are introduced to label each glocal autopolygon based on a majority voting scheme, which produces the polygon-based classification results. In this study, the IRGS algorithm is implemented on the MAp-Guided Ice Classification (MAGIC) system developed by Clausi et al. [51].

D. Incorporating Uncertainty Maps for Misclassification Reduction

Given the uncertainty values, it would be desirable to further improve the classification accuracy by utilizing that uncertainty information. It has been observed that most of the misclassifications correspond to the misclassification of water polygons as ice possibly caused by wind-roughened water texture and the blurry boundaries of melting ice. Besides, given the fact that misclassification rates increase as uncertainty values increase, a thresholding-based method is proposed to correct those misclassified polygons based on uncertainty values, as summarized in Table I. In sum, for a given uncertainty interval, if the range of ice pixel percentage is within the right

TABLE I
SUMMARIZATION OF THE THRESHOLDING-BASED ALGORITHM

Range of mean aleatoric uncertainty value	Range of mean epistemic uncertainty value	Range of ice pixel percentage
[0.1, 0.15)	[0.01, 0.015)	[0, 70%)
[0.15, 0.2)	[0.015, 0.02)	[0, 80%)
[0.2, 0.25)	[0.02, 0.025)	[0, 90%)
[0.25, 0.3)	[0.025, 0.03)	[0, 95%)
>=0.3	>=0.03	[0, 100%]

column, the region is given a label of water. For example, as indicated by the second row of the table, for a certain region, if the percentage of pixels that are classified as ice by Bayesian CNN is below 70% and the mean aleatoric and epistemic values of all pixels are within the given ranges, the region will be corrected as water; otherwise, the region is still classified as ice by following the majority voting scheme. Hence, as the values of average uncertainty increase, the polygon is more likely to be classified as water, which is manifested by the following rows (criteria) in Table I. Finally, when the average aleatoric and epistemic uncertainty are no smaller than 0.3 and 0.03 (which are very large values in terms of uncertainty), all polygons will be corrected as water regardless of the classification results. Note that those thresholds of uncertainty values and ice pixel percentages are determined empirically based on the uncertainty analysis (which are illustrated in Section IV-C) as there is no rule to find optimal values.

IV. EXPERIMENTAL RESULTS

A. CNN Model Training

Based on sea ice charts and visual observation, we label certain numbers of data points in each scene manually for training and validating the Bayesian CNN model. Specifically, instead of simply labeling on certain regions, scribble-based annotations are drawn in both ice and water regions. The 10th point of every ten data points is included in the training set, which minimizes labeled samples from being highly correlated with each other due to spatial proximity. The number of labeled points for each scene is given in Table II. Similar to Leigh et al. [11], the leave-one-out training scheme is adopted for model training and testing. Specifically, to test the performance of the model on the labeled samples in a certain scene, all samples in the other 20 scenes are used for model training, with a percentage of 80% and 20% samples for training and validation, respectively. In this way, a total of 21 CNN models are trained. The Adam optimizer proposed by Kingma and Ba [52] is used to optimize all the parameters. In particular, the variational parameters θ are optimized by taking the gradient of the cost function with respect to μ and σ. The input patch size of the network is set to be 33, indicating that each patch covers an area of around 6.6 $\times$ 6.6 km. The label of each patch corresponds to the label of the data point located at its center. Therefore, to make pixel-based predictions on the whole SAR scene, the input patches are overlapping with each other with a stride of 1.

B. Bayesian CNN Classification Results

The testing accuracy of labeled samples using different combinations of models is presented in Table III. Results obtained from the typical deterministic CNN with the same network architecture are also presented for comparison. As shown in the second and third columns of Table III, both the deterministic and the Bayesian CNNs are able to perform accurate classification on most of the labeled samples with average testing accuracy very close to each other (i.e., 97% and 97.2%, respectively). Note that there are still three scenes with classification accuracy lower than 90%, which will be discussed in the following. The classification rates are still similar for the deterministic and Bayesian CNN after introducing IRGS, as presented in the fourth and fifth columns of Table III. Although the overall testing accuracy only improves around 1%, many misclassified samples obtained in the melting seasons have been corrected, especially in scene 20100730, where the classification accuracy is improved by 7.8%. While a testing accuracy of 98.1% is only around 1.7 percentage points higher than that obtained by Leigh et al. [11] using the same dataset, one scene (scene of data 20100816) was removed from the study in this previous work due to the challenges associated with classifying this scene. In this work, this scene is still retained since it was obtained in the melt season with complex ice–water boundaries and would be important for the study concerning uncertainty.

Nevertheless, since the labeled data points only consist of a small portion of all the pixels in the scenes, it is necessary to further evaluate the performance of the model by visually inspecting the classification results of the whole scene. A few examples of scene classification results consisting of both ice and water samples are shown in Fig. 4. It can be observed from the first and second scenes that the classification results produced by the Bayesian CNN (i.e., the second row of Fig. 4) do not delineate the complex ice–water boundaries and surrounding details (e.g., small floes and wispy looking ice) very well.

TABLE II
NUMBERS OF LABELED SAMPLES FOR EACH SCENE

Scene of data	# of ice	# of water
20100418_163315	0	191617
20100426_040439	0	247420
20100510_035620	15357	274494
20100524_034756	46537	253900
20100605_163323	6941	262942
20100623_041255	107845	211812
20100629_163326	32819	130951
20100712_031834	25360	123510
20100721_173208	50010	63192
20100730_162908	49596	42753
20100807_173610	45462	13400
20100816_163329	64329	53444
20100907_035614	121003	0
20100909_163321	115469	7990
20101003_163324	93027	42070
20101021_041325	125690	42770
20101027_025726	66151	71672
20101114_041304	41096	67983
20101120_163324	0	239720
20101206_015139	0	239252
20101214_025725	0	203292
Total	1006692	2784184

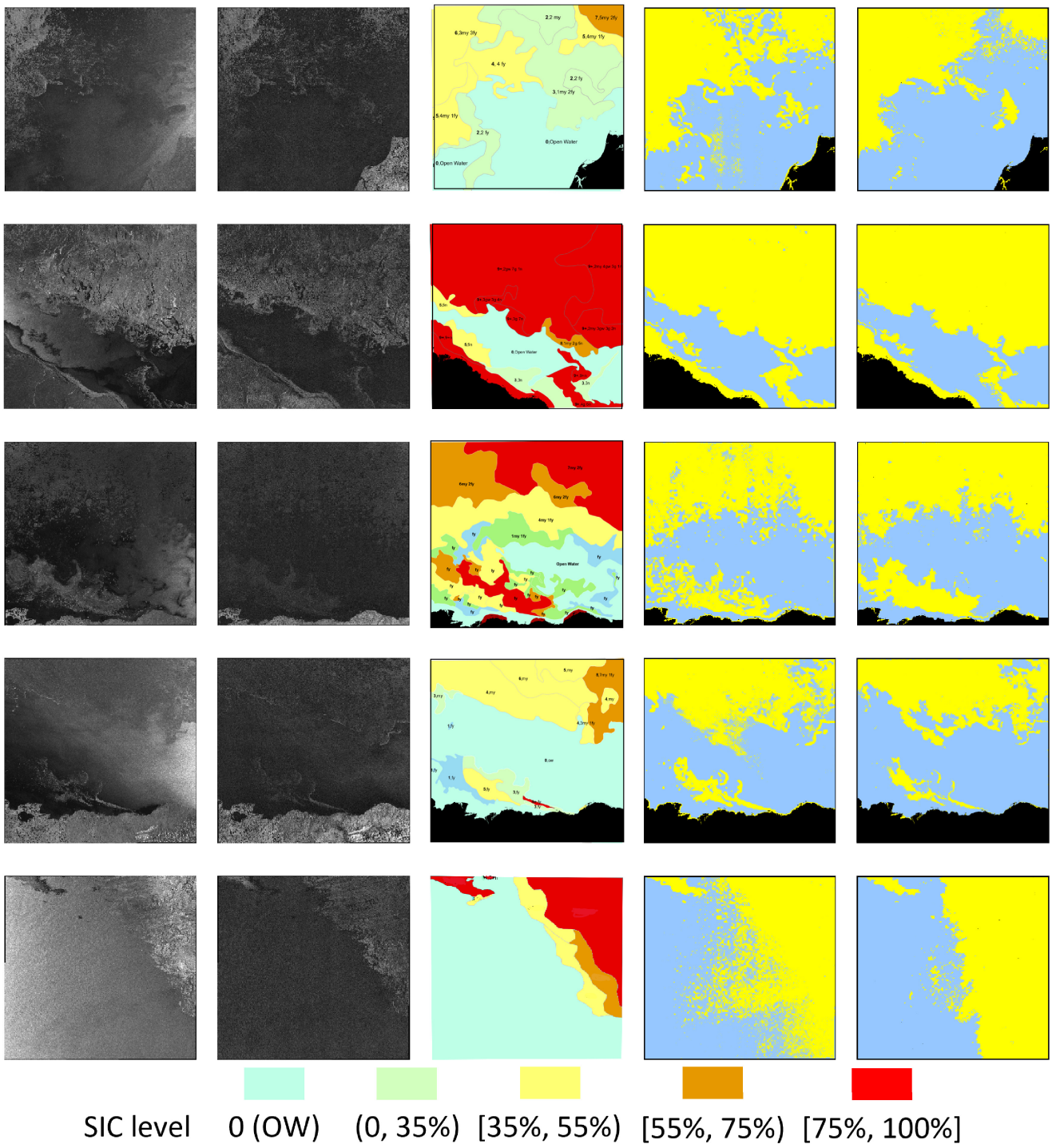

Fig. 4. First and second columns: examples of SAR scenes (scene of data from top to bottom: 20100807, 20101027, 20100730, 20100816, and 20101021) in HH and HV channels, respectively. The contrast of the HV images has been enhanced for better observation. Third column: image analysis charts for each scene. Fourth column: classification results of the scenes on the same rows using Bayesian CNNs, with ice and water indicated by yellow and blue, respectively. Fifth column: polygon-based classification results after combining IRGS algorithm with Bayesian CNN outputs. The classification accuracy of the labeled samples is listed on each scene.

In addition, the misclassification of small clusters of pixels can be found in the regions of both ice and open water. For example, for the scene in the first row of Fig. 4, the misclassification of water pixels as ice with patterns similar to salt and pepper noise can be clearly observed. In contrast, after introducing IRGS segmentation results (as shown in the fifth column of Fig. 4), the detailed ice–water boundaries with wispy ice and small floes surrounded are well preserved. Most of the noise-like misclassifications have been eliminated. Therefore, it can be concluded that combining the IRGS algorithm with Bayesian CNN significantly improves the classification results visually.

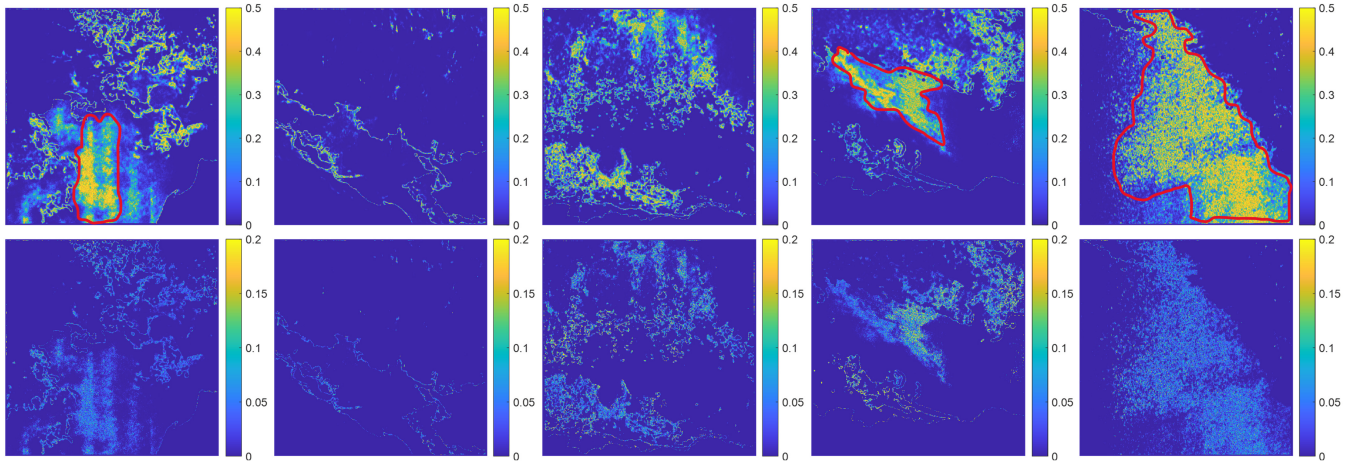

Fig. 5. Aleatoric (top row) and epistemic (bottom row) uncertainty maps of the scenes presented in Fig. 4 calculated from Bayesian CNN predictions. First to fifth columns: uncertainty maps of scene of data 20100807, 20101027, 20100730, 20100816, and 20101021, respectively. Regions with high uncertainty that are misclassified are outlined in red. Note that the color scales of aleatoric uncertainty and epistemic uncertainty are different.

However, limitations still exist. For better illustration, three scenes with the lowest testing accuracy are presented in the third, fourth, and fifth rows of Fig. 4. First, for misclassified pixels within large homogeneous regions, which can be observed in the scenes in the last two rows of Fig. 4, introducing IRGS cannot correct all those misclassifications effectively. In particular, the misclassification of open water as ice in the scene in the fourth row (scene of data 20100816) is probably caused by the blurry ice–water boundaries. In contrast, the open water region of the scene in the last row (scene of data 20101021) consists of rough textures caused by high wind speeds, which looks totally different from the relatively smooth water surface in the other scenes and eventually leads to the worst classification results among the whole dataset. However, classification results can be improved significantly if some data points of water are added to model training. The second limitation that cannot be lifted by combining IRGS is that for scenes obtained in the melt season, it is difficult to distinguish between ice and water samples in regions with intermediate SICs (e.g., regions in the middle top of the scene of data: 20100807), which makes it unreliable to evaluate and compare the performance of the proposed models only based on labeled points in either open water (regions with 0 SIC) or regions with 90%+ SIC.

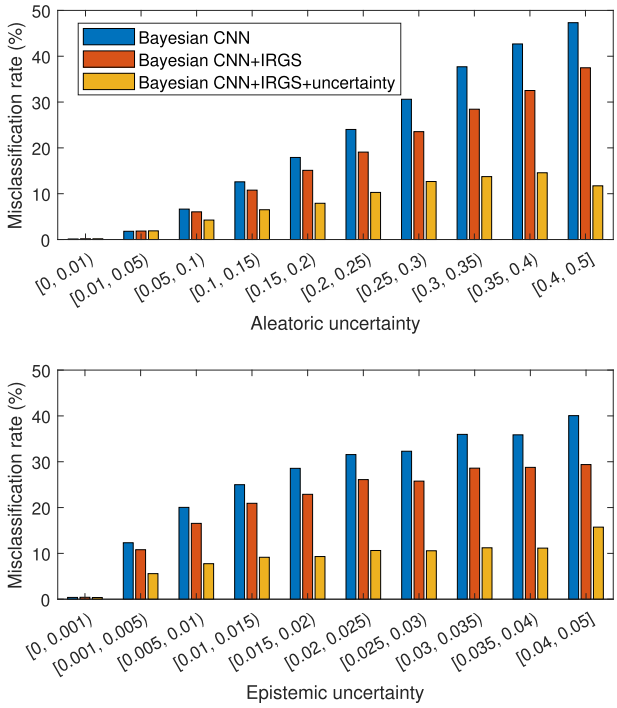

Fig. 6. Relationship between aleatoric/epistemic uncertainty and misclassification rate calculated from the labeled data points using different combinations of models.

C. Uncertainty Analysis

Uncertainty maps of the five scenes in Fig. 4 calculated from the Bayesian CNN outputs are presented in Fig. 5 for analysis. Note that the color scales of aleatoric uncertainty and epistemic uncertainty are different. The sampling number T in (5) is set to be 5. Although a larger T can be more accurate in approximating the real predictive variance, it will be very time-consuming in computation. We also find that by choosing a larger T, the uncertainty values stay basically the same. Misclassified pixels/regions (e.g., regions outlined in red in Fig. 5) tend to have high uncertainty values in both aleatoric and epistemic maps. High uncertainty can also be observed in regions with intermediate SICs where it is difficult to determine the exact class label. Besides, most of the ice–water boundaries can be well delineated in uncertainty maps as pixels on and beside the boundaries tend to have brighter intensities. Even in the scenes with large areas of misclassifications (e.g., scene of data 20100816 and 20101021), a strong contrast of uncertainty values on the two sides of the "true" ice–water boundaries helps users know where the model makes either confident or probably erroneous predictions.

To study the relationship between uncertainty values and misclassification rates, Fig. 6 shows the numerical relationship between aleatoric/epistemic uncertainty and misclassification rates calculated from labeled data points. Also, the misclassification rates using different combinations of algorithms are compared with each other. For Bayesian CNN outputs, the

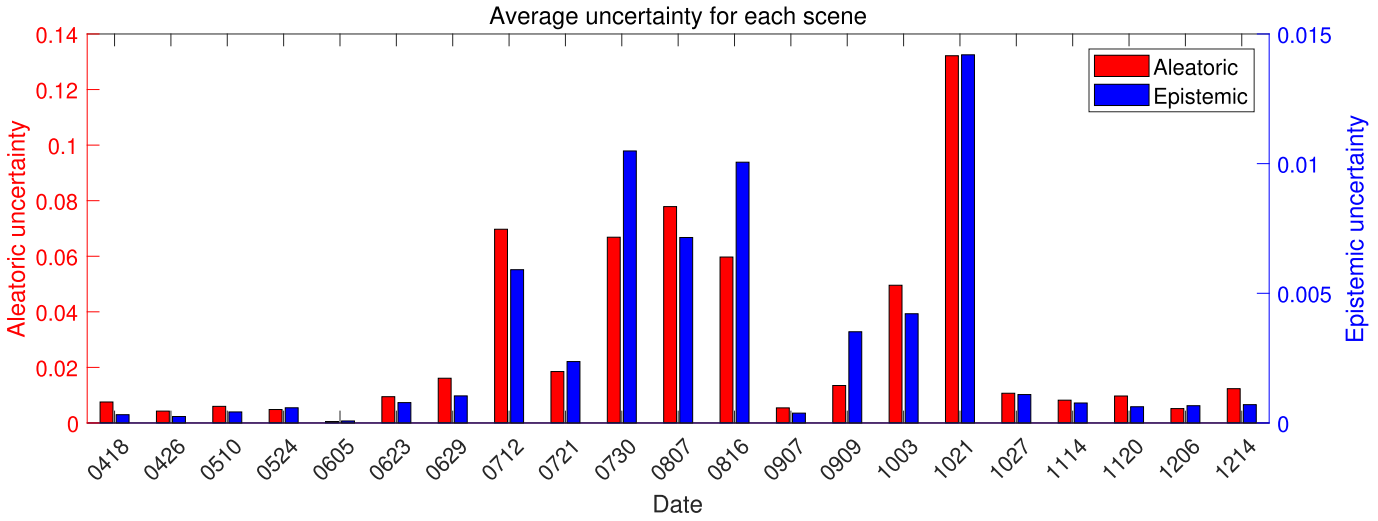

Fig. 7. Average uncertainty values for each scene. Note that the y-axis scales of aleatoric uncertainty and epistemic uncertainty are different.

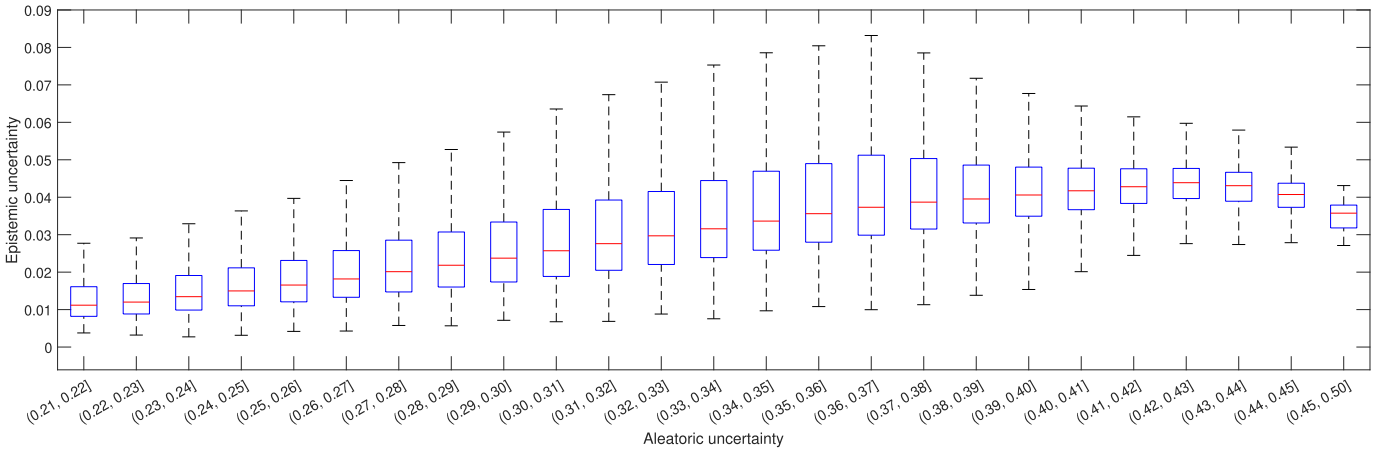

Fig. 8. Bivariate box plot for aleatoric and epistemic uncertainties among the top 5% data points in terms of the aleatoric uncertainty.

misclassification rates (i.e., the blue bars) increase linearly with aleatoric uncertainty. In contrast, there is a significant increase in the misclassification rate when epistemic uncertainty exceeds 0.001, followed by a linear increase, gradually flattening out after around 0.03. Although similar patterns can be observed after combining the IRGS algorithm, the misclassification rates have been decreased effectively, as indicated by the orange bars. For example, when aleatoric uncertainty is higher than 0.25, the misclassification rates decrease by around 10%. Therefore, compared to the slight improvement in average classification accuracy, the IRGS algorithm is crucial in reducing misclassification rates in regions with high uncertainty.

An overview of uncertainty values for all scenes is shown in Fig. 7 to investigate the change of uncertainty values under different ice conditions. It is obvious that the main source of uncertainty comes from the aleatoric part as the values of aleatoric uncertainty are nearly ten times larger than those of epistemic uncertainty, which is consistent with other studies concerning uncertainty quantification in computer vision tasks [44], [45]. This can also be accounted for by the complex backscattering mechanisms on different surface conditions across different incidence angles. The average

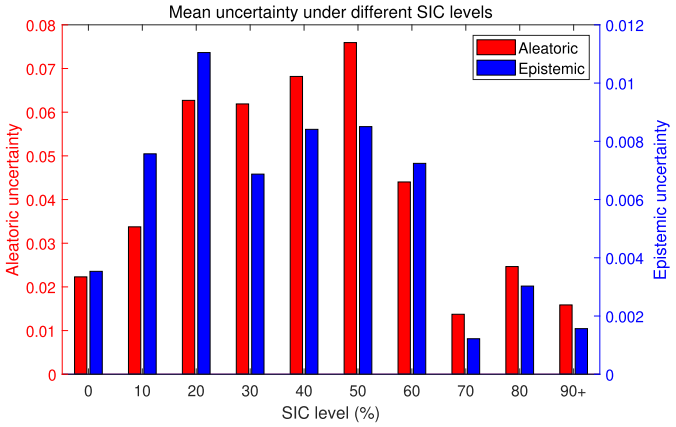

Fig. 9. Mean uncertainty value under different SIC levels. The numbers of different scenes contributed to each SCI level (from 0 to 90+ are: 15, 6, 5, 4, 5, 8, 4, 3, 6, and 16). Note that the y-axis scales of aleatoric uncertainty and epistemic uncertainty are different.

uncertainty values in the melting season are much higher compared to other times of the year, especially for scene of data 20100712, 20100730, 20100807, and 20100816. While the results of the other three scenes have been presented in

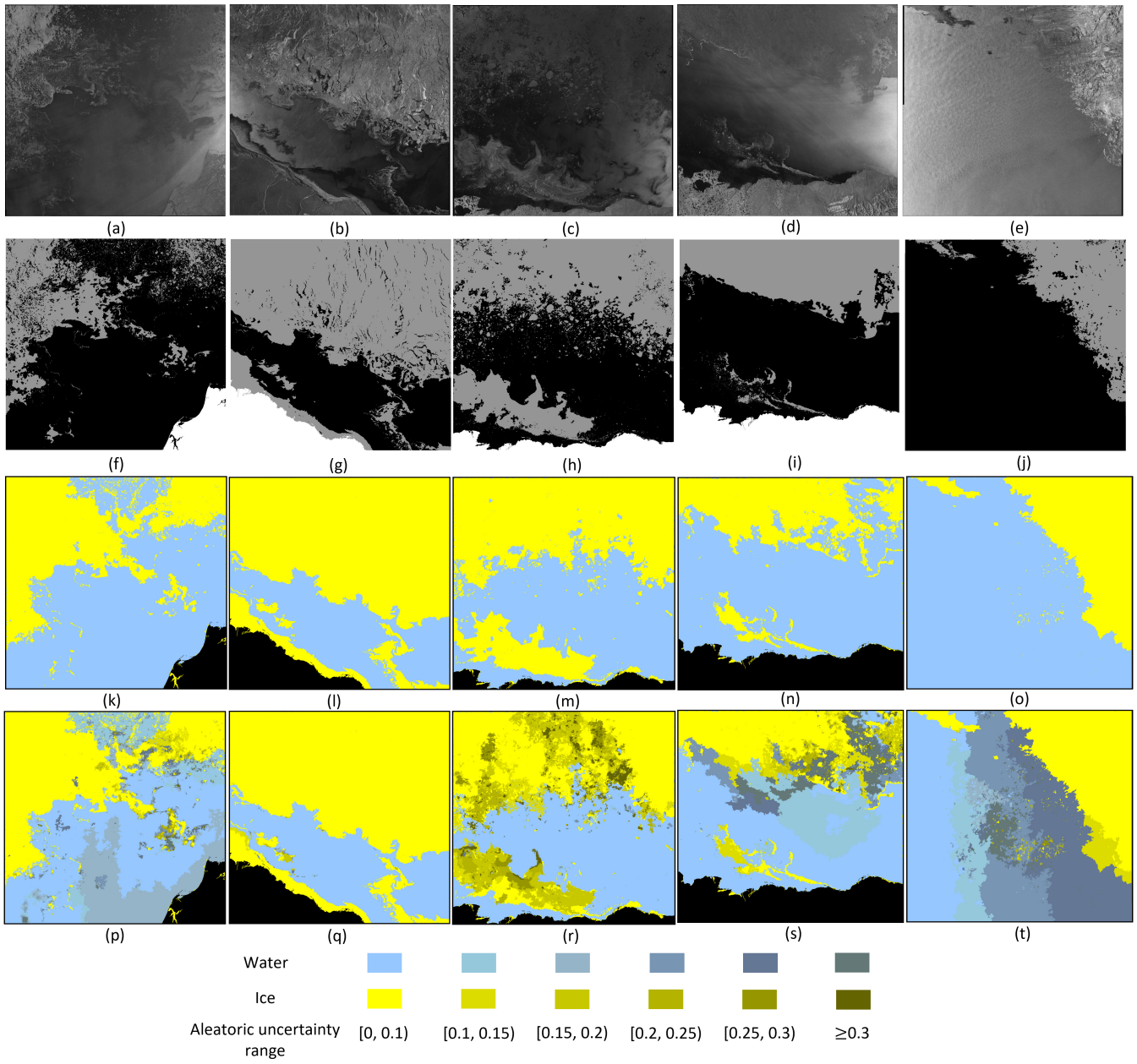

Fig. 10. (Top) (a)–(e) Scene examples presented in Fig. 4 (scene of data from left to right: 20100807, 20101027, 20100730, 20100816, and 20101021) in HH channel. (f)–(j) Second row: corresponding manually labeled label maps for visual evaluation. The gray, black, and white colors represent ice, water, and land, respectively. (k)–(o) Third row: examples of the classification results with uncertainty thresholding algorithm incorporated. (Bottom) (p)–(t) More informative representation of the classification results on the top. Regions in different ranges of average aleatoric uncertainty values are indicated by different colors.

Fig. 5, the high uncertainty values in scene 20100712 are mainly caused by regions with thin ice. Despite the high uncertainty, the model manages to classify most of these regions correctly, as shown in Table III. The low uncertainty values for the scene obtained on scene 20100907 are due to the fact that only pure water is present in that scene. As ice began to freeze up again in fall, the uncertainty values decrease gradually (except for the scene obtained on scene 20101021, which has been explained in Section IV-B), which is plausible as the ice is forming and no longer wet yet still shares complex boundaries with open water. By looking back to Table III, it can be concluded low uncertainty values normally indicate accurate classification, while misclassification or unconfident predictions can be inferred by high uncertainty. The relationship between incidence angle values and average uncertainty values across different incidence angles is also investigated and it is found that there is no direct correlation between those two quantities.

To further investigate the spatial correlation between aleatoric and epistemic uncertainty, a bivariate box plot is shown in Fig. 8. Following [45], only the top 5% of data points in terms of the aleatoric uncertainty are selected for analysis since most of the pixels have uncertainty values very close to zero. It can be observed that the median value (indicated by the red line in Fig. 8) of epistemic uncertainty generally increases as the aleatoric uncertainty increases. For data points

TABLE III

TESTING ACCURACY OF LABELED SAMPLES FOR EACH SCENE USING DIFFERENT COMBINATIONS OF MODELS (UNIT: %)

Scene ID	Deterministic CNN	Bayesian CNN	Deterministic CNN +IRGS	Bayesian CNN +IRGS	Bayesian CNN+ IRGS+ uncertainty
20100418	99.6	99.9	100	100	100
20100426	97.8	99.9	99.6	100	100
20100510	99.9	99.9	100	100	100
20100524	99.4	99.0	97.9	97.9	99.3
20100605	98.8	97.5	97.5	97.5	97.5
20100623	99.5	99.9	99.8	99.9	99.9
20100629	99.6	99.7	99.6	99.6	99.7
20100712	98.2	98.1	99.9	99.1	98.1
20100721	99.9	100	100	100	100
20100730	88.8	89.7	99.1	97.5	97.5
20100807	99.7	96.9	100	100	100
20100816	88.7	86.3	93.2	92.4	95.0
20100907	100	99.9	100	100	100
20100909	99.9	99.9	100	100	100
20101003	99.3	99.5	99.6	100	99.6
20101021	77.2	77.8	76.9	77.2	98.3
20101027	99.1	99.4	99.8	99.7	99.7
20101114	92.6	99.6	99.9	99.8	99.8
20101120	99.8	99.5	100	100	99.9
20101206	99.7	99.4	100	100	99.5
20101214	99.7	99.5	100	100	100
Average	97.0	97.2	98.2	98.1	99.2

with aleatoric uncertainty values larger than 0.4, the mean epistemic values stay the same or decrease. Nevertheless, since the number of those data points only consists of a very tiny portion of the top 5% data points, it can still be inferred that the model tends to give more uncertain predictions as the aleatoric uncertainty increases.

In addition, the relationship between uncertainty and SIC is also studied based on the ice charts provided. For faster computation, 1% of the data points in each scene is randomly selected to obtain their corresponding ice concentrations in the ice chart for analysis. In consequence, the number of data points within open water or 90%+ ice concentration regions are much greater than the others. As shown in Fig. 9, pixels within the regions with intermediate SIC levels (i.e., from 20% to 60%) have relatively high aleatoric uncertainty values. This is reasonable as regions with intermediate SIC levels mostly appear in the melting season or ice–water boundaries. As for the epistemic uncertainty, the pattern is similar except that the value is also relatively high when SIC is 10%. Both aleatoric and epistemic uncertainty values decrease significantly under high SIC levels (larger than 70%). Note that the low uncertainty of the 70% bin might be caused by insufficient areas of study. In comparison, the mean uncertainty values for data points in open water are slightly higher, which is mainly caused by the blurry ice–water boundaries and rough water surface under high wind speeds.

D. Uncertainty Map-Incorporated Classification Results

The thresholding parameters presented in Table I are determined after the uncertainty presented in Fig. 6. Specifically, we adopt the minimum and maximum values of each aleatoric and epistemic uncertainty value interval in Fig. 6 as the threshold values. Based on the predefined uncertainty threshold values, different combinations of ice pixel percentage

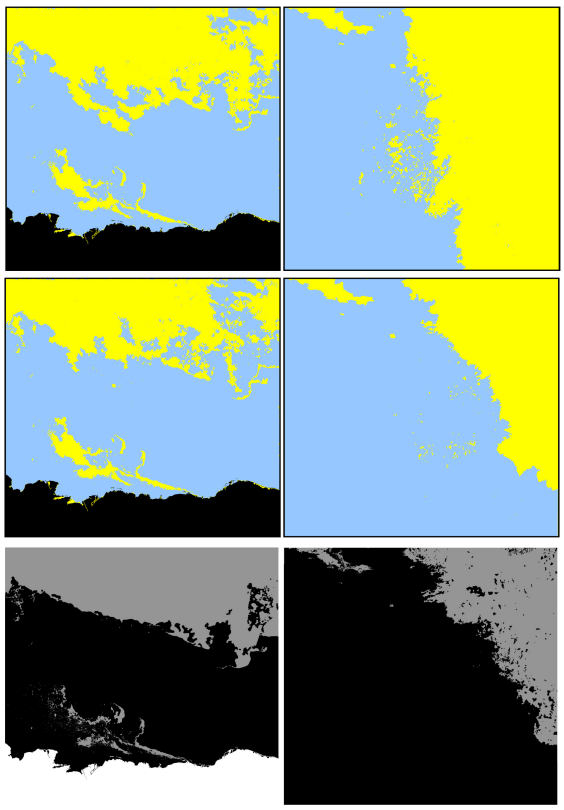

Fig. 11. Comparison between prediction results with and without incorporating the uncertainty value thresholding algorithm for scenes 20100816 and 20101021. (Top) Results without incorporating uncertainty information. (Middle) Results after incorporating uncertainty information. (Bottom) Corresponding manually labeled label maps for visual evaluation.

threshold values are tested and it is found that the given ones in Table I produce the highest classification accuracy. Since all images are used to conduct the uncertainty analysis shown in Fig. 6, the threshold values are determined by all the images. We acknowledge that it would better maintain the independence between the training and testing sets if the leave-one-out scheme is also applied to the determination of the thresholding parameters. Nevertheless, due to the limited amount of data, some of the threshold values might not be well determined, which might negatively affect the classification results. A set of threshold parameters that receive optimal performance among all the scenes facilitates us to apply them to other images in future works.

The numerical results after combining the proposed algorithm have been presented in the last column of Table III and the yellow bars in Fig. 6. The average testing accuracy is further improved by 1.1%, which is mainly attributed to the significant improvement in scene 20101021. Most of the pixels on the rough water surface that are misclassified as ice have been corrected effectively, as shown in Fig. 10(o). More importantly, the yellow bars in Fig. 6 show that the misclassification rates under relatively high uncertainty values have dropped significantly.

The final prediction results of the scenes presented in Fig. 4 are shown in Fig. 10, with corresponding scenes in the HH channel and manually labeled label maps attached.

Although those maps are labeled mainly based on visual observation and might not be 100% accurate, they can be a good reference for visual evaluation of the classification results. By referring to the HH imagery and reference label maps, it can be observed that other than scene 20101021, the large areas of misclassification in scenes 20100807 (in the middle top with low SIC) and 20100816 (on the left side of the ice–water boundaries) have been reduced significantly as well. Examples of prediction results with and without incorporating the uncertainty value thresholding algorithm are also given in Fig. 11 for easier comparison. The prediction maps with uncertainty information incorporated are given in the bottom row of Fig. 10, in which different colors are assigned to regions with different mean aleatoric uncertainty values. It can be observed that the assigned color becomes darker as the uncertainty value increases. This greatly facilitates us to flag regions that might need to be checked manually by ice analysts.

V. Conclusion

By assuming that each weight in a CNN follows a certain Gaussian distribution, a Bayesian CNN is first built to conduct pixelwise classification between sea ice and open water in SAR imagery with both epistemic and aleatoric uncertainty. The IRGS algorithm is then introduced into the workflow to segment each SAR scene into homogeneous polygons, which are then used to further improve the Bayesian CNN classification result by preserving detailed ice–water boundaries and eliminating noise-like misclassifications. By analyzing the uncertainty maps provided by the Bayesian CNN, a thresholding-based algorithm is proposed to correct the miscalssified polygons and significantly reduce misclassification rates under high uncertainty values. The training and testing of the proposed model are implemented on PyTorch installed on a local PC running Ubuntu 22 with a 3.6-GHz AMD Ryzen 7 eight-core CPU, 40-GB memory, and a 64-bit operating system. The training times (a total of around 370 000 patches are input for training) for the Bayesian CNN and a conventional AlexNet are 34 and 20 min, respectively. On the other hand, the average execution times for the two networks mentioned above to produce pixelwise classification results for an SAR scene are 11 and 3 min. Therefore, the Bayesian CNN fits in well with the time requirement for operational sea ice monitoring.

Results generated from multiple dual-polarized SAR scenes collected from RADARSAT-2 in the Beaufort Sea show that while relatively high classification accuracy on labeled samples can be achieved, factors, such as the presence of melting ice, rough water surface under high wind speeds, as well as the blurry ice–water boundaries during the melting season, degrade the accuracy and confidence of our model predictions significantly. Therefore, apparently, it is not sufficient to evaluate the reliability of model predictions only by looking into the classification accuracy of labeled samples and visual inspection. Since it has been demonstrated that the misclassification rate increases with the increase of uncertainty, the uncertainty maps calculated from the Bayesian CNN outputs provide extra information that facilitates us to locate the correct ice–water boundaries and pixels that might be misclassified. Scenes that are obtained during the melting season generally have the highest mean uncertainty values, followed by scenes obtained during freeze-up. The decomposition of predictive uncertainty into aleatoric and epistemic uncertainty manifests that the main source of uncertainty comes from the aleatoric part, which is caused by the large variability in the patterns of both ice and water under different circumstances (e.g., different wind speeds, ice concentrations, and ice types). Although the classification results might be further improved by techniques such as data augmentation and changing the window size, they cannot reduce the uncertainty values. The bivariate plot between aleatoric and epistemic uncertainty indicates that the model tends to be more uncertain about its predictions on regions with higher data uncertainty values. In addition, by referring to the sea ice charts, it is further validated that data points in regions with intermediate ice concentrations tend to have larger uncertainty values than those in regions of open water or high ice concentration levels.

The proposed method that consists of the Bayesian CNN, the IRGS algorithm, and the uncertainty value thresholding-based algorithm is able to conduct accurate polygon-based classification on more than 99% of the labeled pixels. Regions with high uncertainty values are highlighted to give more informative prediction maps. In the future, an uncertainty-guided model can be developed to perform better classification on those uncertain samples. Besides, since in recent years, many advanced CNNs have been proposed for more complex classification tasks, it is worthwhile to convert them into Bayesian CNNs as well and see whether the classification performance can be further improved. In addition, it is also necessary to introduce the uncertainty analysis to sea ice typing, which is a more challenging classification task in SAR-based sea ice monitoring. As ice charts derived from manual analysis of SAR imagery typically ignore small details within ice fields, it would be valuable to utilize classification derived from other data (e.g., optical/IR/NIR) as additional and more independent reference for evaluation in future studies.

Acknowledgment

The authors would like to thank MDA for the 21-scene RADARSAT-2 dataset and D. Isaacs for providing the corresponding image analysis charts. RADARSAT-2 data and products are copyright of MDA Ltd. 2011, all rights reserved. RADARSAT is an official trademark of the Canadian Space Agency.

References

[1] M. Liu, Y. Dai, J. Zhang, X. Zhang, J. Meng, and Q. Xie, "PCA-based sea-ice image fusion of optical data by HIS transform and SAR data by wavelet transform," *Acta Oceanolog. Sinica*, vol. 34, no. 3, pp. 59–67, Mar. 2015.

[2] S. Paul and M. Huntemann, "Improved machine-learning-based open-water–sea-ice–cloud discrimination over wintertime Antarctic sea ice using MODIS thermal-infrared imagery," *Cryosphere*, vol. 15, no. 3, pp. 1551–1565, Mar. 2021.

[3] S. Andersen, R. Tonboe, L. Kaleschke, G. Heygster, and L. T. Pedersen, "Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic sea ice," *J. Geophys. Res.*, vol. 112, no. C8, 2007, Art. no. C08004.

[4] C. Melsheimer, G. Spreen, Y. Ye, and M. Shokr, "Antarctic sea ice types from active and passive microwave remote sensing," *Cryosphere Discuss.*, to be published, doi: 10.5194/tc-2021-381.

[5] X.-M. Li, Y. Sun, and Q. Zhang, "Extraction of sea ice cover by Sentinel-1 SAR based on support vector machine with unsupervised generation of training data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3040–3053, Apr. 2021.

[6] H. Lyu, W. Huang, and M. Mahdianpari, "A meta-analysis of sea ice monitoring using spaceborne polarimetric SAR: Advances in the last decade," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6158–6179, 2022.

[7] M. S. Tamber, K. A. Scott, and L. T. Pedersen, "Accounting for label errors when training a convolutional neural network to estimate sea ice concentration using operational ice charts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1502–1513, 2022.

[8] T. Xie, W. Perrie, C. Wei, and L. Zhao, "Discrimination of open water from sea ice in the Labrador Sea using quad-polarized synthetic aperture radar," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111948.

[9] J. A. Karvonen, "Baltic Sea ice SAR segmentation and classification using modified pulse-coupled neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1566–1574, Jul. 2004.

[10] S. Ochilov and D. A. Clausi, "Operational SAR sea-ice image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4397–4408, Nov. 2012.

[11] S. Leigh, Z. Wang, and D. A. Clausi, "Automated ice–water classification using dual polarization SAR satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5529–5539, Sep. 2014.

[12] N. Zakhvatkina, A. Korosov, S. Muckenhuber, S. Sandven, and M. Babiker, "Operational algorithm for ice–water classification on dual-polarized RADARSAT-2 images," *Cryosphere*, vol. 11, pp. 33–46, Jan. 2017.

[13] W. Tan, J. Li, L. Xu, and M. A. Chapman, "Semiautomated segmentation of Sentinel-1 SAR imagery for mapping sea ice in Labrador coast," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Ses.*, vol. 11, no. 5, pp. 1419–1432, May 2018.

[14] 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," *Cryosphere*, vol. 14, no. 8, pp. 2629–2645, Aug. 2020.

[15] M. Jiang, D. A. Clausi, and L. Xu, "Sea ice mapping of RADARSAT-2 imagery by integrating spatial contexture with textural features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7964–7977, 2022.

[16] Y. Zhang et al., "Sea ice and water classification on dual-polarized Sentinel-1 imagery during melting season," *Cryosphere Discuss.*, to be published, doi: 10.5194/tc-2021-85.

[17] 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, p. 2165, Jul. 2020.

[18] Y. Ren, X. Li, X. Yang, and H. Xu, "Development of a dual-attention U-Net model for sea ice and open water classification on SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4010205.

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

[20] H. Lyu, W. Huang, and M. Mahdianpari, "Eastern Arctic sea ice sensing: First results from the RADARSAT constellation mission data," *Remote Sens.*, vol. 14, no. 5, p. 1165, Feb. 2022.

[21] M. Jiang, L. Xu, and D. A. Clausi, "Sea ice–water classification of RADARSAT-2 imagery based on residual neural networks (ResNet) with regional pooling," *Remote Sens.*, vol. 14, no. 13, p. 3025, Jun. 2022.

[22] M. Jiang, X. Chen, L. Xu, and D. A. Clausi, "Semi-supervised sea ice classification of SAR imagery based on graph convolutional network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 1031–1034.

[23] Q. Yan and W. Huang, "Sea ice sensing from GNSS-R data using convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1510–1514, Oct. 2018.

[24] A. Stokholm, T. Wulf, A. Kucik, R. Saldo, J. Buus-Hinkler, and S. M. Hvidegaard, "AI4SeaIce: Toward solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4304013.

[25] D. Malmgren-Hansen et al., "A convolutional neural network architecture for Sentinel-1 and $AMSR_2$ data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1890–1902, Mar. 2021.

[26] J. Karvonen, "Baltic sea ice concentration estimation from C-band dual-polarized SAR imagery by image segmentation and convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4301411.

[27] N. Saberi, K. A. Scott, and C. Duguay, "Incorporating aleatoric uncertainties in lake ice mapping using RADARSAT–2 SAR images and CNNs," *Remote Sens.*, vol. 14, no. 3, p. 644, Jan. 2022.

[28] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[29] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021.

[30] P. Ortiz, M. Orescanin, V. Petkovic, S. W. Powell, and B. Marsh, "Decomposing satellite-based classification uncertainties in large Earth science datasets," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4106211.

[31] L. Alagialoglou, I. Manakos, M. Heurich, J. Cervenka, and A. Delopoulos, "A learnable model with calibrated uncertainty quantification for estimating canopy height from spaceborne sequential imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4410913.

[32] 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.

[33] 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 (CVPRW)*, Jun. 2015, pp. 28–35.

[34] 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.

[35] Y. Gal et al, "Uncertainty in deep learning," Ph.D. thesis, Univ. Cambridge, Cambridge, U.K., 2016.

[36] A. Graves, "Practical variational inference for neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 2348–2356.

[37] D. P. Kingma, T. Salimans, and M. Welling, "Variational dropout and the local reparameterization trick," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 2575–2583.

[38] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," 2015, *arXiv:1506.02158*.

[39] K. Shridhar, F. Laumann, and M. Liwicki, "A comprehensive guide to Bayesian convolutional neural network with variational inference," 2019, *arXiv:1901.02731*.

[40] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.

[41] K. Shridhar, F. Laumann, and M. Liwicki, "Uncertainty estimations by softplus normalization in Bayesian convolutional neural networks with variational inference," 2018, *arXiv:1806.05978*.

[42] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 84–90.

[44] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5574–5584.

[45] Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, "Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation," *Comput. Statist. Data Anal.*, vol. 142, Feb. 2020, Art. no. 106816.

[46] Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, "Uncertainty quantification using Bayesian neural networks in classification: Application to ischemic stroke lesion segmentation," in *Medical Imaging With Deep Learning*, 2018. [Online]. Available: https://openreview.net/forum?id=SkP2Q9sG

[47] 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.

[48] Q. Yu and D. A. Clausi, "IRGS: Image segmentation using edge penalties and region growing," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 30, no. 12, pp. 2126–2139, Dec. 2008.

[49] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1302–1317, Apr. 2012.

[50] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
[51] D. A. Clausi, A. K. Qin, M. S. Chowdhury, P. Yu, and P. Maillard, "MAGIC: MAp-guided ice classification system," *Can. J. Remote Sens.*, vol. 36, pp. S13–S25, Jan. 2010.
[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

Xinwei Chen received the B.Eng. degree in information engineering from the South China University of Technology, Guangzhou, China, in 2017, and the Ph.D. degree in electrical engineering from the Memorial University of Newfoundland, St. John's, NL, Canada, in 2021.

He is currently a Post-Doctoral Fellow with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include machine learning and image processing with applications in remote sensing.

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

She was a Post-Doctoral Researcher with the Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Toronto, ON, Canada, where she was part of a team involved in the development of a sea ice data assimilation system. In 2012, she joined the Department of Systems Design Engineering, University of Waterloo, as a Faculty Member with a specialization in sea ice remote sensing and data assimilation.

Linlin Xu (Member, IEEE) received the B.Eng. and M.Sc. degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in remote sensing from the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada, in 2014.

He has authored or coauthored various papers in high-impact remote sensing journals and conferences. His research interests include machine learning, remote sensing, and environmental monitoring.

Mingzhe Jiang (Graduate Student Member, IEEE) received the B.Eng. degree in electronic information engineering and the M.Eng. degree in signal and information processing from the Hefei University of Technology, Hefei, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in systems design engineering with the University of Waterloo, Waterloo, ON, Canada.

His research interests include machine learning and image processing with applications in remote sensing.

Yuan Fang received the B.Eng. and M.Sc. degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2016 and 2018, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2022.

She is currently a Post-Doctoral Fellow with the Department of Systems Design Engineering, University of Waterloo. Her research interests include remote sensing image processing, hyperspectral imaging, machine learning, and environmental monitoring.

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

He then worked in medical imaging with Mitra Imaging, Waterloo. He started his academic career in 1997 as an Assistant Professor in geomatics engineering with the University of Calgary, Calgary, AB, Canada. In 1999, he returned to the University of Waterloo, where he is currently a Professor and specializes in the field of intelligent systems and is the Associate Dean of Research and External Partnerships with the Faculty of Engineering. He has many contributions, conducting research primarily in remote sensing, computer vision, image processing, and algorithm design. He has published extensively and received many scholarships, paper awards, research awards, and teaching excellence awards; his efforts have led to successful commercial implementations, culminating in the creation, and sale of a tech company.

Dr. Clausi is a fellow of the Canadian Academy of Engineering. He was the Co-Chair of the International Association for Pattern Recognition Technical Committee 7—Remote Sensing from 2004 to 2006.