The Importance of the Incidence Angle for GLCM Texture Features and Ancillary Data Sources for Automatic Sea Ice Mapping

by

Fernando Jose Pena Cantu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2024

© Fernando Jose Pena Cantu 2024

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Fernando J. Pena Cantu is the sole author of all chapter except chapter 4. All chapters were written under the supervision of Dr. David Clausi and Dr. Andrea Scott.

Components of Chapter 3 were submitted to the International Geoscience and Remote Sensing Symposium (IGARSS) 2023, in Pasadena California. Additionally, Chapter 4 is based on a paper co-written by Xinwei Chen and Fernando J. Pena Cantu. Both articles are shown in the table below.

Chapter	Citation	Author's Contributions
3	Fernando J. Pena Cantu, Xinwei Chen, Linlin Xu, Max Ian A. Manning, K. Andrea Scott, and David A. Clausi. Incidence angle dependence of texture features from dual polarization RADARSAT-2 sea ice imagery. In IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, pages 172–175, 2023. [52]	Primary author of manuscript.
4	Xinwei Chen, Fernando J. Pena Cantu, Muhammed Patel, Linlin Xu, Neil C. Brubacher, K. Andrea Scott, and David A. Clausi. A comparative study of data input selection for deep learning-based automated sea ice mapping. *International Journal of Applied Earth Observation and Geoinformation*, 131:103920, 2024.[8]	Responsible for the conceptualization, methodology, analysis, and visualization of the study. The article was co-written with equal contribution by Xinwei Chen and Fernando J. Pena Cantu.

Abstract

Sea ice is a critical component of Earth's polar regions. Monitoring it is vital for navigation and construction in the Arctic and crucial to understand and mitigate the impacts of climate change. Synthetic aperture radar (SAR) imagery, particularly dual-polarized SAR, is commonly used for this purpose due to its ability to penetrate clouds and provide data in nearly all weather conditions. However, relying solely on HH and HV polarizations for automated sea ice mapping models has limitations, as different ice types and conditions may yield similar backscatter signatures. To enhance the accuracy of these classification models, researchers have explored the integration of additional features, including hand-crafted texture features, learned features, and supplementary data sources.

This thesis makes two main contributions to the field of automated sea ice mapping. The first contribution investigates the dependence of incidence angle (IA) on gray level co-occurrence matrix texture features (GLCM) and its impact on sea ice classification. The methodology involved extracting GLCM features from SAR images in dB units and analyzing their dependence on IA using linear regression and class separability metrics. In addition, a Bayesian classifier was trained to compare the classification performance with and without incorporating the IA dependence. The results indicated that the IA effect had a minor impact on classification performance ($\approx 1\%$), with linear regression results indicating that the IA dependence accounts for approximately less 10% of the variance in most cases.

The second contribution evaluates the importance of various data inputs for automated sea ice mapping using the AI4Arctic dataset. A U-Net based model was trained with SAR imagery, passive microwave data from AMSR2, weather data from ERA5, and ancillary data. Ablation studies and the addition of individual data inputs were conducted to assess their impact on model performance. The results demonstrated that including AMSR2, time, and location data significantly increased model performance, especially for the classification accuracy of major ice types in stage of development (SOD). ERA5 data had mixed effects, as it was found not to increase performance when AMSR2 was already included.

These findings are critical for the development of more accurate and efficient automated sea ice mapping systems. The minimal impact of IA dependence on GLCM features suggests that accounting for IA may not be necessary, simplifying the feature extraction process. Identifying the most valuable data inputs allows for the optimization of model performance, ensuring better resource allocation and enhanced operational capabilities in sea ice monitoring. This research provides a foundation for future studies and developments in automated sea ice mapping, contributing to more effective climate monitoring and maritime navigation safety.

Acknowledgements

I extend my deepest gratitude to my supervisors, Professor David Clausi and Professor Andrea Scott, for their unwavering support and guidance over the past two years. Their mentorship provided me with invaluable opportunities to explore the fields of computer vision and remote sensing, significantly enriching my academic experience.

I am immensely grateful to Max Manning for developing the Python library for the Magic Software, which greatly facilitated my research. Special thanks are also due to Prof. Xinwei Chen and Prof. Linlin Xu for their outstanding work on the AutoIce Challenge and our collaborative efforts on feature selection. Their contributions were instrumental in advancing my research.

I owe a great deal of appreciation to my lab teammates: Neil Curtis Brubacher, Muhammed Patel, Javier Noa, Kiernan McGuigan, Lily de Loe, Chris Czarnecki, and Jinman Park. Their camaraderie and collaborative spirit made our lab an inspiring and enjoyable place for research and discussion.

Finally, I would like to express my heartfelt thanks to my family for their unwavering support and encouragement throughout my graduate studies. Their belief in my abilities has been a constant source of motivation.

Dedication

To anyone looking to learn more about sea ice.

Table of Contents

Author's Declaration ii

Statement of Contributions iii

Abstract iv

Acknowledgements v

Dedication vi

List of Figures x

List of Tables xiii

1 Introduction 1

1.1 Motivation 1

1.2 Thesis Organization 3

2 Background 5

2.1 Sea Ice Mapping 5

2.2 Synthetic Aperture Radar 7

2.2.1 SLAR vs SAR 9

2.2.2 Backscatter Coefficient (σ°) 10

2.2.3 Polarization . . . 11
2.2.4 Incidence Angle Effect . . . 11

3 Incidence Angle Dependence of GLCM Texture Features 14
3.1 Gray-Level Co-Occurrence Matrix (GLCM) . . . 15
3.2 Previous Works . . . 18
3.3 Methodology . . . 19
3.3.1 Dataset . . . 19
3.3.2 Data Pre-processing . . . 22
3.3.3 Experiments . . . 24
3.4 Results and Discussion . . . 28
3.4.1 GIA Results . . . 28
3.4.2 Slope and R^2 . . . 28
3.4.3 Class Separability . . . 31
3.4.4 GLCM Parameter Dependence . . . 35
3.5 Conclusion . . . 41

4 Data Input Selection for Automated Sea Ice Mapping 42
4.1 Deep Learning . . . 43
4.1.1 Training Deep Learning Models . . . 43
4.1.2 Neural Networks . . . 45
4.1.3 Convolutional Neural Networks (CNNs) . . . 47
4.2 AI4Arctic Sea Ice Challenge Dataset . . . 53
4.3 Methodology . . . 56
4.3.1 Data input processing . . . 56
4.3.2 U-Net-based Sea Ice Mapping Model . . . 56
4.3.3 Training Specifications . . . 59
4.3.4 Model Evaluation Metrics . . . 59

4.3.5 Cross-Validation and Hypothesis Testing 60
4.4 Results and Discussion 61
4.4.1 Performance of Full Model 61
4.4.2 Results Removing and Adding Data Inputs 65
4.4.3 Passive Microwave (AMSR2) Feature Analysis 66
4.4.4 Analysis on Spatial-Temporal Features 78
4.4.5 ERA5 Feature Analysis 82
4.4.6 Feature Importance Ranking 85
4.5 Conclusion 86

5 Conclusion **88**
5.1 Future Work 89

References **91**

Letters of Copyright Permission **99**

List of Figures

2.1 Example of daily sea ice chart from Canadian Ice Service. 6

2.2 Egg code schematic. 7

2.3 Typical geometry of SAR system adapted from Remote Sensing of the Environment [34, Chapter 8]. 10

2.4 Illustration of SAR polarizations. 12

2.5 2D histograms illustrating the distribution of backscatter values against the IA for SAR systems. 12

2.6 Density plot and scatter plot comparing HH polarization backscatter in dB versus IA for OW and MYI [39]. 13

3.1 Illustration of GLCM computation. 17

3.2 Findings from Scharien *et al.* [58], which illustrate the dependence between GLCM features and IA. The data were gathered during winter in the Canadian Arctic using Sentinel-1 C-band. 20

3.3 Results from Guo *et al.* [22], which show the class dependence between GLCM features and IA. The data were gathered during winter in the Arctic Ocean using TerraSAR-X. 21

3.4 Visualization of the 29 RADARSAT-2 C-band SAR imagery scenes from the Beaufort Sea. 22

3.5 2D histogram showing the distribution between ENT and IA for FYI in HH polarization 25

3.6 2D histogram showing the relationship between backscatter [dB] of open water and the IA. Data from all the scenes were used to create this histogram. 29

3.7 Analysis of GLCM features in relation to IA across different ice classes and polarizations . . . 31
3.8 2D histograms showing the relationship between IA and GLCM features for different ice classes and polarizations. . . . 32
3.9 2D histograms illustrating the relationship IA and the GLCM feature COR across different ice classes and polarizations. . . . 33
3.10 Illustration of the impact of inter-scan banding on GLCM features . . . 34
3.11 JM distance analysis for GLCM features across different ice classes and polarizations. . . . 36
3.12 Relationship between JM distance and GLCM parameters. . . . 38
3.13 Relationship between R^2 and GLCM parameters. . . . 39
3.14 Slope of the linear regression between COR and IA as a function of different GLCM parameters for HV polarization in OW ice class . . . 40

4.1 Small ResNet architecture. . . . 51
4.2 U-net architecture overview . . . 52
4.3 Class distribution between training and testing dataset for SIC, SOD, and FLOE. . . . 55
4.4 A diagram illustrating the U-Net-based sea ice mapping pipeline implemented in this research. . . . 57
4.5 Confusion matrices for SIC, SOD, and FLOE on the test and validation dataset. . . . 63
4.6 An example SAR scene in the testing dataset (file ID name . . . 67
4.7 Relative sensitivity of AMSR2 for oceanic geophysical parameters as a function of observing frequency [32]. . . . 68
4.8 An example comparing the mapping results from models with two different feature combinations: HH + HV + AMSR2 89.0 GHz and HH + HV + AMSR2 6.9 GHz. . . . 69
4.9 The corresponding AMSR2 feature maps of the SAR scene in Fig.4.6 in all available frequencies. . . . 72
4.10 The confusion matrix results for removing AMSR2 features . . . 73

4.11 The value change in each element of the SIC classification confusion matrix (test dataset after adding AMSR2 features, in comparison with the "HH, HV only" model. 74
4.12 Confusion matrix results of FLOE when adding ASMSR2. 75
4.13 An example SAR scene in the testing dataset under melting conditions. . . 76
4.14 Mapping results, comparing models without time or without location information. 79
4.15 Comparison example between full model and Model without time and location 80
4.16 Confusion matrix SOD results for time and location removal. 81
4.17 Comparison scene between "HH, HV only", "HH, HV and T2M", and "HH, HV and Wind". 83
4.18 The corresponding ERA5 feature maps of the SAR scene in Fig.4.17. . . . 84
4.19 The value change in each element of the SIC classification confusion matrix after adding ERA5 features, in comparison with the "HH, HV only" model. 84

List of Tables

2.1 Classes of SOD used by the Canadian Ice Service [5]. 7
2.2 Classes of FLOE used by Canadian Ice Service [5]. 8

3.1 GLCM features used in this thesis. 18
3.2 Results from GIA Bayesian classifier. 29
3.3 Percentage increase in JM Distance obtained by changing GLCM parameters. 37
3.4 Percentage increase between the maximum and minimum R^2 values obtained by changing GLCM parameters. 40

4.1 The features (data inputs) available in the dataset, which are all utilized for investigation in this research. 54
4.2 Layer specifications of the U-Net-based model using all available data inputs. 58
4.3 Specifications of model training. 59
4.4 Overall performance of the "Full Model". 62
4.5 Monthly performance for the "Full Model". **Bold** highlights lowest three scores. 64
4.6 Summary of numerical results from ablation studies on different feature types. 64
4.7 Summary of numerical results obtained from the "adding features" experiments. 65
4.8 Numerical results obtained from the models (with SAR HH and HV inputs) adding AMSR2 channels individually. 66
4.9 Comparison of the monthly performance on SOD between "Full Model" and "Remove AMSR2". Statistical significant decreases are shown in bold. . . . 74

4.10 Comparison of the monthly performance on SIC and SOD between "HH HV Only" and "HH HV and AMSR2". Statistical significant improvements are shown in bold. 77

4.11 Comparison of the monthly performance on SOD between "HH HV Only" and "HH HV + Time". Statistical significant improvements are shown in bold. 79

4.12 Numerical results obtained from models with SAR HH and HV inputs and individual ERA5 features. 82

4.13 A ranking of importance for different features in the combined score and individual parameters. 86

Chapter 1

Introduction

1.1 Motivation

Monitoring sea ice is vital for safe navigation and construction, particularly in regions such as the Arctic Ocean and its marginal seas, including the Kara, Barents, White, Chukchi, East Siberian, Laptev, Greenland, and Beaufort Seas. Additionally, sea ice monitoring is crucial for understanding and mitigating the impacts of climate change on both the environment and human activities.

Navigation: Icebreakers and other maritime vessels require up-to-date ice reports and forecasts to navigate waters close the polar regions. [5]. Additionally, sea ice data is crucial for vessel construction and strengthening, ensuring that ships are built to withstand ice stress [5].

Construction: In addition to navigation, long-term climatological ice data are essential for marine engineers to make informed decisions regarding dock construction and other infrastructure projects in the Arctic Ocean [5]. The harsh and dynamic conditions of the Arctic environment necessitate robust data to guide safe and sustainable construction practices.

Climate Modeling: Sea ice plays a crucial role in climate modeling, particularly in the phenomenon known as polar amplification [61], where changes in the net radiation balance produce larger temperature changes near the poles than the planetary average. Ice-albedo feedback, where melting ice reduces surface reflectivity, leading to further warming and ice melt, is a significant contributor to polar amplification [29]. The state of sea ice significantly influences high-latitude warming in climate models, with thinner ice correlating with higher

polar amplification. Coupled ocean and sea ice models enhance polar weather forecasts, emphasizing the importance of precise sea ice edge positioning [16].

Increased Importance of Sea Ice Monitoring due to Climate Change: The decline in sea ice due to climate change has wide-ranging impacts. Reductions in sea ice are expected to increase the number of navigable days, especially in the Beaufort Sea, leading to an increase in maritime traffic in the Arctic, making the need for daily ice charts even more imperative [45]. Indigenous populations in the Arctic are particularly vulnerable, as their sanitation, food supply, and infrastructure are closely tied to the natural environment [50]. Wildlife, such as polar bears, also face significant challenges, with habitat loss affecting their survival and conservation efforts [51]. Additionally, coastal erosion increases with less sea ice, as stronger wave action and permafrost degradation exacerbate the loss of land along the Arctic coasts [19]. Precise and frequent sea ice monitoring is indispensable for navigating the challenges posed by climate change.

The Canadian Ice Service (CIS) is the Canadian organization responsible for generating operational daily ice charts for navigation in Canada's navigable waters. These charts are crucial to ensure the safe passage of vessels through ice-infested waters. The charts are predominantly derived from dual-polarized synthetic aperture radar (SAR) imagery, which provides two images: one in the HH polarization and another in the HV polarization [6, Chapter 1].

Highly trained ice analysts at CIS manually interpret SAR images, supplemented by ancillary data, to produce operational ice charts. This manual process, while effective, is time-consuming. With the increasing volumes of data generated by modern SAR systems, the need to produce daily updates, and the expected rise in maritime traffic, there is a need to develop automated sea ice charting processes. Automation would streamline the interpretation of SAR data, enabling the delivery of near-real-time sea ice products with increased spatial coverage.

Despite the potential of SAR imagery, computer vision algorithms have historically struggled to differentiate between various ice types, in part due to the incidence angle effect [39, 15, 42] and wind roughening [17]. The pixel information from dual-polarized SAR imagery lacks sufficient discriminative power to reliably distinguish between ice classes. To enhance algorithm performance, researchers have explored incorporating additional features. These features can be categorized as follows:

- **Hand-Crafted Texture Features:** These include features such as the gray level co-occurrence matrix (GLCM) [65], Gabor filters [12], and Markov random fields [70].

- **Learned Features:** These are features extracted from HH and HV polarizations through the use of convolutional neural networks (CNNs) [68].
- **Other Data Sources:** These encompass passive microwave data [43], weather data [7], temporal and locational information, and multispectral data [55].

While GLCM texture features have been extensively studied to improve sea ice classification, there is limited research on how incidence angle (IA) effects impact their classification performance. In contrast, while sea ice analysts use a variety of data sources to produce ice charts (such as SAR, passive microwave, optical data, weather information, etc.), most deep learning algorithms have traditionally relied solely on SAR data [17, 26, 35, 43, 55, 59]. Consequently, there is still significant potential to explore the impact of these diverse features on sea ice mapping performance. As such, the objectives of this thesis are to address these gaps by:

1. Investigate the dependence between GLCM textural features and incidence angle for sea ice classification by:
    - Measuring the dependence of GLCM textural features on the angle of incidence for HH and HV polarizations; and
    - Evaluating the classification performance of a model with and without incorporating the incidence angle dependence.
2. Evaluating the impact on seasonal performance of the following additional data sources incorporated into a state-of-the-art sea ice classification model:
    - Passive microwave data from AMSR2;
    - Weather data from ERA5; and
    - Spatial (latitude / longitude) and temporal (month) data .

1.2 Thesis Organization

This thesis is organized as follows:

- Chapter 2 provides the background and literature review on sea ice mapping and synthetic aperture radar (SAR). It begins with an introduction to sea ice charts,

including elements such as egg code, sea ice concentration, stage of development, and floe. The chapter also covers the fundamentals of SAR, its advantages, and its specific applications for sea ice mapping. Detailed discussions on SAR geometry and polarization are also included.

- Chapter 3 discusses the importance of investigating the dependence of GLCM features on IA. It starts with a leave-one-out cross-validation study using a Bayesian classifier to identify an optimal set of GLCM parameters and compares results with and without considering the IA. This is followed by a linear regression analysis to determine the slope and coefficient of determination (R^2), representing the dependence and its strength. The Jeffries-Matusita distance is calculated to evaluate the separability of GLCM features with and without incorporating IA dependence. Finally, the sensitivity of multiple GLCM features is assessed by calculating JM distance, slope, and R^2.

- Chapter 4 examines the performance of different data types in sea ice classification using a CNN with a U-Net architecture. The impact of each data type is investigated through two approaches: removing variables to compare results against a model utilizing all features, and adding variables to compare results against a model using only HH and HV from SAR imagery. The effect of data types on seasonal performance is also explored.

- Chapter 5 concludes the thesis, providing comprehensive insights and discussions regarding the direction of future research.

Chapter 2

Background

2.1 Sea Ice Mapping

Sea ice mapping plays a pivotal role in the study of Earth's climatic systems, providing crucial data for projecting climate change impacts [16, 29]. Over recent decades, the Arctic has experienced a marked decrease in both the extent and thickness of sea ice due to significant regional warming. As Arctic sea routes that were once impassable become increasingly navigable, there is a growing interest in exploiting these routes to establish more direct connections between global markets [45].

Canadian Ice Services (CIS) have undertaken the critical task of producing daily sea ice charts. An example of these can be found in Fig. 2.1. These charts are essential for safe maritime navigation, particularly for operations conducted by the Canadian Coast Guard and icebreaker fleets tasked with facilitating ship routing and escort services through icy waters. The primary source of data for these charts is synthetic aperture radar (SAR) imagery, captured by satellites such as RADARSAT-1 and RADARSAT-2 [5].

A key feature of these sea ice charts is the incorporation of the "egg code" system (Fig. 2.2), a standardized symbolic method used to convey detailed information about the ice conditions within an operator-defined polygon (area) [5]. The egg code contains the following three types of information to describe ice conditions:

- Sea Ice Concentration (SIC): The concentration of ice in a given polygon/area;
- Stage of Development (SOD): The age and thickness of the ice; and

- Form of Ice (FLOE): The shape of the ice and the size of floes.

Table 2.1 and 2.2 show the different classes used by CIS for SOD, and FLOE respectively.

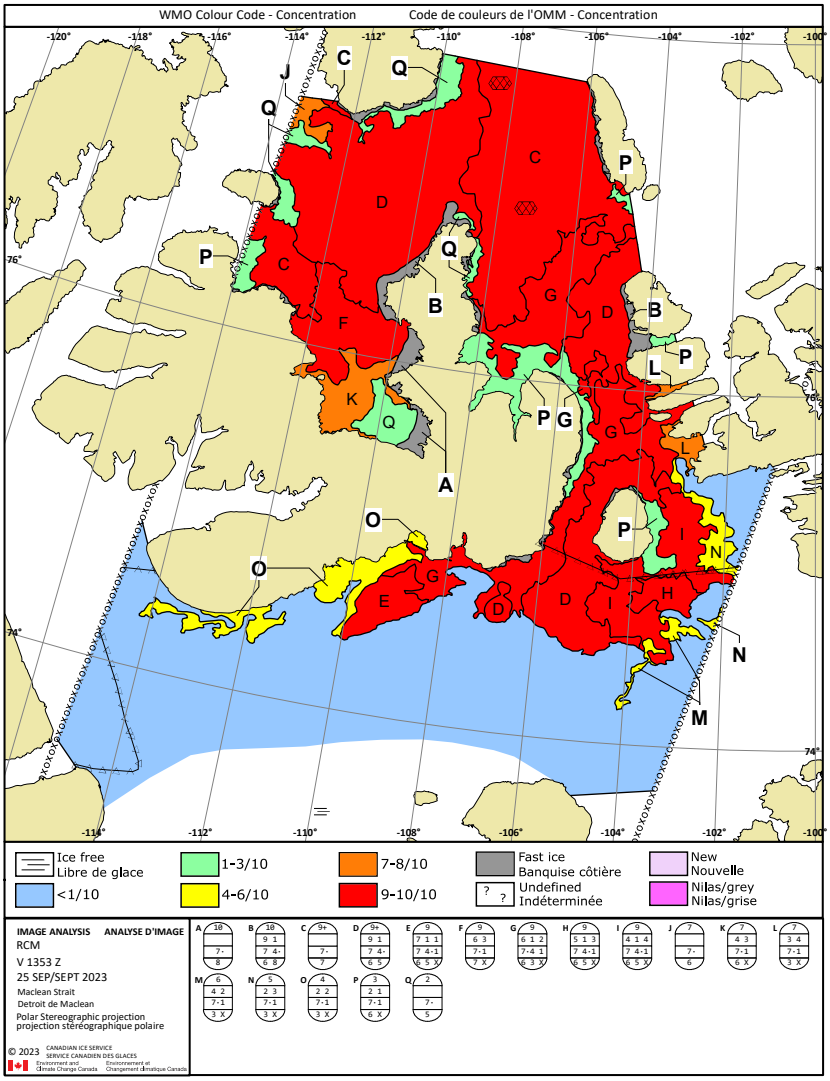


Figure 2.1: Example of daily sea ice chart from Canadian Ice Service [21]. Copy is available at https://ice-glaces.ec.gc.ca/prods/WIS144CT/20230925135323_WIS144CT_0012791450.pdf

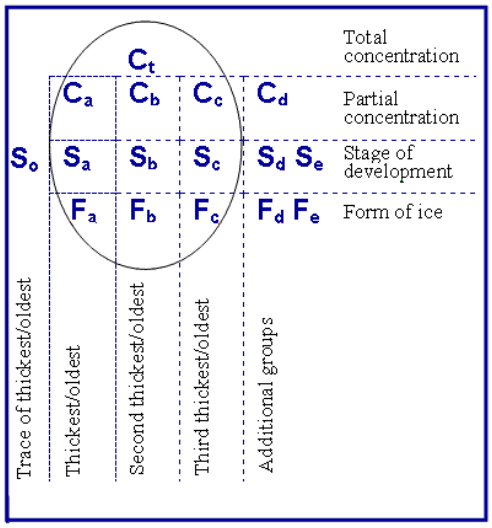


Figure 2.2: Egg code schematic [5]. Copy is available at https://ice-glaces.ec.gc.ca/content_contenu/ice_codes/manice/CH3/CH3_3.4.html

Table 2.1: Classes of SOD used by the Canadian Ice Service [5].

Stage of Development			
Category	Sub-category	Thickness	Notes
New Ice		<10 cm	Recently formed ice.
Young Ice		10-30 cm	
	Grey Ice	10-15 cm	
	Grey-White Ice	15-30 cm	
First Year Ice		>30 cm	Sea ice of not more than one winter's growth.
	Thin -First Year Ice	30-70 cm	
	Medium-First Year Ice	70-120 cm	
	Thick-First Year Ice	>120 cm	
Old Ice			Sea ice which has survived at least one summer's melt.
	Second-Year Ice		
	Multi-Year Ice		

2.2 Synthetic Aperture Radar

SAR is an advanced form of real aperture side-looking radar (SLAR), both integral to the category of imaging radars used for remote sensing. These systems, deployed on aircraft or satellites, employ an active microwave approach by transmitting microwave signals towards a target and then measuring the scattering of these signals returned to the sensor, known

Table 2.2: Classes of FLOE used by Canadian Ice Service [5].

Form of Ice		
Class	Width	Description
Pancake Ice	$30cm - 2m$	Circular pieces of ice up to 10 cm in thickness.
Small Ice cake		Any ice cake less than 2 m across.
Brash ice	$< 2m$	Accumulation of ice fragments no more than 2m across.
Agglomerated brash		Brash Ice compacted due to primarily wind action.
Ice cake	$2 - 20m$	Any relative piece of ice less than 20 m.
Floe	$20m - 10km$	Any relative flat piece of ice across 20m or more.
Small floe	$20 - 100m$	
Medium floe	$100 - 500m$	
Big floe	$500m - 2000m$	
Vast floe	$2 - 10km$	
Giant floe	$> 10km$	
Fast Ice	N/A	Ice which forms and remains fast along the coast.
Icebergs	N/A	A massive piece of ice of greatly varying shape, protruding 5 m or more above sea level, which has broken away from a glacier.
Growlers	N/A	Piece of ice smaller than a bergy bit and floating less than 1 m above the sea surface.
Floebergs	N/A	A massive piece of ice composed of a hummock or a group of hummocks, frozen together and separated from any surrounding ice.

as backscatter. The backscattered signals are captured by the receiving antenna, and processed to generate images of the Earth's surface [34, Chapter 8],[63].

SAR is the preferred technology for sea ice mapping due to the following key advantages it provides other remote sensing sensors. These advantages include:

1. **All-Weather Capability**: SAR systems operate effectively regardless of weather conditions [63], [34, Chapter 8].

2. **Day and Night Operation**: Unlike optical sensors that require sunlight and can only be operated in daylight conditions, SAR can collect data during both day and night [63], [34, Chapter 8].

3. **Sensitivity to Surface Properties**: SAR is highly sensitive to changes in surface roughness, and dielectric properties, which are useful to differentiate different types of sea ice [34, Chapter 8].

4. **High Spatial Resolution**: Modern SAR systems offer high spatial resolution (exceeding 100 m resolution), which is essential for detailed mapping of sea ice features such as leads, ridges, and polynyas [63, 18].

5. **Wide Area Coverage**: SAR satellites can cover vast areas with each pass, allowing for the monitoring of extensive sea ice regions efficiently and regularly [63].

6. **Polarization**: SAR sensors can utilize different polarizations enhancing its ability to analyze the physical and structural properties of the ice [18], [34, Chapter 8].

2.2.1 SLAR vs SAR

An early form of imaging radar systems is SLAR. This system transmits pulses from its antenna at a specified pulse repetition frequency to achieve the desired ground resolution of the imaged area. These pulses illuminate continuous strips of the surface, covering the entire swath to one side of the satellite's orbit direction [63]. Radar imaging systems are characterized by two types of resolution, as depicted in Fig. 2.3:

The range resolution, which is perpendicular to the azimuth (direction of flight), is determined by the pulse length (τ), the speed of light (c), and the depression angle (γ) as depicted by in (2.1) [34, Chapter 8].

$$R_r = \frac{\tau * c}{2 \cos \gamma} \tag{2.1}$$

Azimuth resolution, in contrast, aligns with the direction of flight and depends on the beam width. The beam width is proportional to the wavelength (λ) of the signal and inversely proportional to the length of the antenna (L). The azimuth resolution is expressed in (2.2), Where H is the altitude of the satellite [34, Chapter 8].

$$R_a = \frac{H}{\sin \gamma} * \frac{\lambda}{L} \tag{2.2}$$

One of the challenges in radar imaging is achieving high azimuth resolution, which can be limited by the necessity to either reduce the wavelength (increasing atmospheric interaction) or enlarge the antenna size (which has practical limits). SAR overcomes these limitations by utilizing the motion of the satellite. By coherently combining signals received at different times due to the platform's movement, SAR systems construct a virtual aperture significantly longer than the physical antenna. This process enhances the azimuth resolution beyond what traditional real-aperture radar (RAR) can achieve [44]. The azimuth resolution in SAR is defined in (2.3) [34, Chapter 8].

$$SAR_r = \frac{L}{2} \tag{2.3}$$

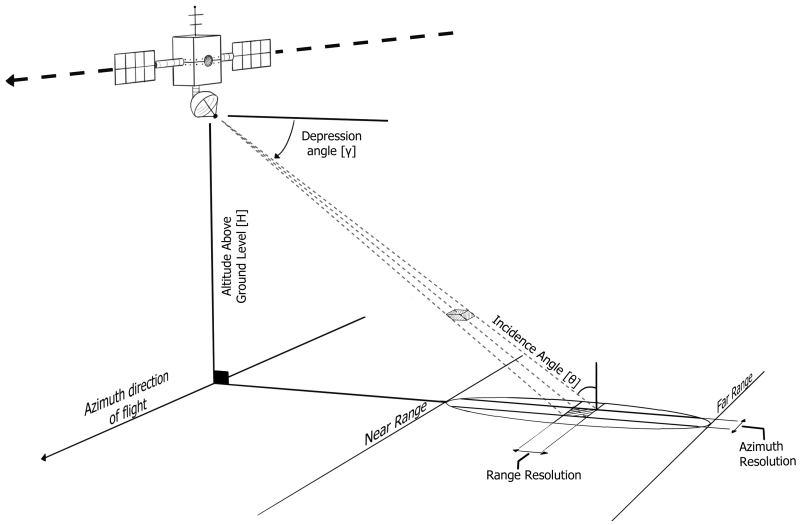


Figure 2.3: Typical geometry of SAR system adapted from Remote Sensing of the Environment [34, Chapter 8].

2.2.2 Backscatter Coefficient (σ°)

The foundational radar (2.4) describes the power received by the satellite [34, Chapter 8].

$$P_r = \frac{P_t * G_t}{4\pi R^2} * \sigma * \frac{A_r}{4\pi R^2} \tag{2.4}$$

where P_r is the received power, P_t is the transmitted power, G_t is the antenna gain, σ is the radar cross-section, A_r is the receiving antenna area, and R is the distance to the target [34, Chapter 8].

One of the problems of using the concept of radar cross section to measure physical properties of the target (e.g, sea ice) is its dependence on the surface area of the imaged footprint (i.e., the ground resolution of the sensor). Therefore, when radar sensors of different resolutions view the same surface, then σ may be different. A more useful quantity is obtained by dividing σ by the area of the observed terrain. This parameter is called backscatter coefficient or "sigma naught" σ°. The backscatter coefficient can be expressed in different forms. It can be represented in linear units, where the backscatter coefficient is given directly with values typically ranging between 0 and 1, or in decibels (dB), which is a logarithmic scale. The formula to convert the linear sigma naught to decibels is:

$$\sigma^{\circ}_{dB} = 10 * log_{10}(\sigma^{\circ}) \tag{2.5}$$

The backscatter coefficient is determined by the surface roughness, the volume of the medium, physical and electrical properties, as well as the radar parameters; namely the wavelength, polarization, and the local incidence angle of the incident signal [63]. The backscatter coefficient is ultimately what is being imaged in SAR, and it is what is used to differentiate between ice types.

2.2.3 Polarization

Radar antennas transmit and receive polarized electromagnetic energy, where the electric fields of the radar waves oscillate within specific, predetermined planes. These planes can either be horizontal (H) or vertical (V), giving rise to four main combinations of transmit-receive polarizations: HH, HV, VH, and VV. Fig. 2.4 illustrates an example where the transmitted signal has a horizontal polarization (H). The transmitted signal scatters off the ground, resulting in both horizontally (H) and vertically (V) polarized signals being emitted. Thus, it shows an example of HH and HV. Each of the four polarization pairs yields distinct insights about the observed surfaces for example:

1. **Open Water Observation:** Water with high wind speeds tends to appear brighter in like-polarization (HH/VV) compared to calm water, which displays reduced brightness. In contrast, cross-polarization (HV/VH) brightness is generally independent of wind speed [18].

2. **Ice Characterization:** HV and VH polarization is particularly effective for distinguishing between smooth and deformed ice. Deformed ice exhibits strong depolarization effects, making it more distinguishable in HV polarization than in others [18].

2.2.4 Incidence Angle Effect

Due to the sideways-looking nature of the radar, as illustrated in Fig. 2.3, the IA decreases with increasing range. This results in variations in backscatter, even when imaging the same material. Generally, as the IA increases, the backscatter coefficient (σ°) decreases,

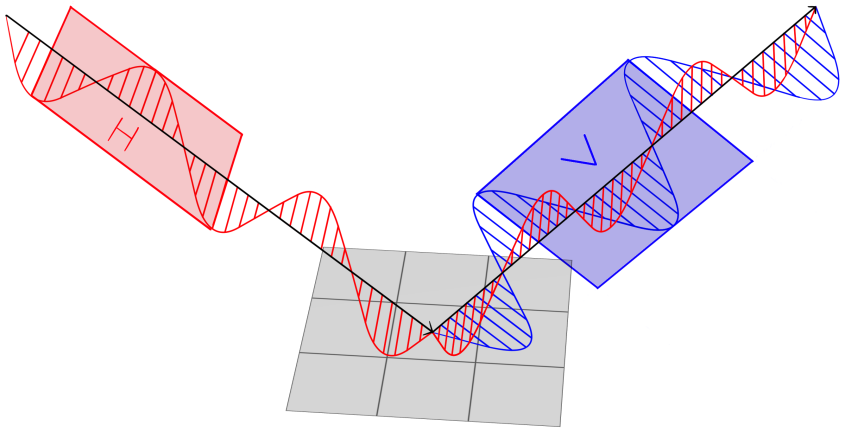


Figure 2.4: Illustration of SAR polarizations. The horizontally polarized (H) transmitted signal is shown in red. Upon scattering from the ground, both horizontally (H) and vertically (V) polarized signals are emitted, with the vertically polarized signal shown in blue.

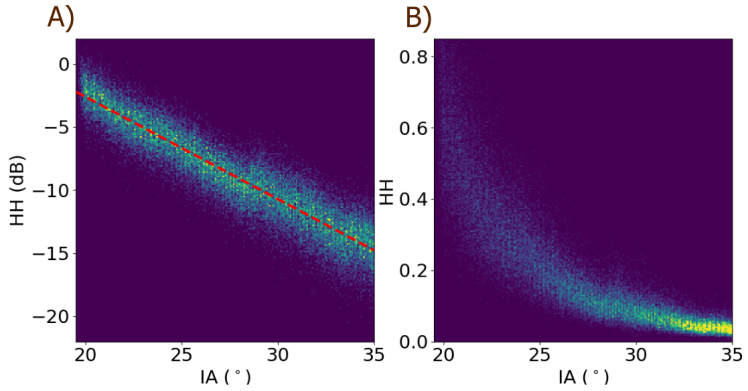


Figure 2.5: 2D histograms illustrating the distribution of backscatter values against the IA for SAR systems. The left histogram (A) shows backscatter in decibels (dB) with a red dashed trendline indicating a negative linear correlation, while the right histogram (B) presents the same data in linear units, revealing the backscatter pattern following an exponential decay [40].

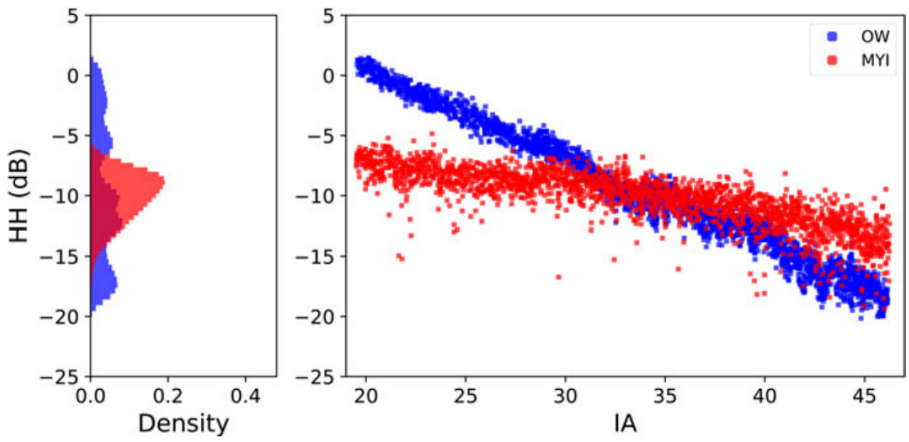


Figure 2.6: Density plot and scatter plot comparing HH polarization backscatter in dB versus IA for OW and MYI [39].

exhibiting an approximately exponential decay. When represented in a logarithmic domain, this decay appears linear [39, 40], as shown in Fig. 2.5.

Different materials demonstrate distinct backscatter responses as a function of incidence angle. For example, as shown in Fig. 2.6 OW exhibits a faster decrease in σ^0 with increasing incidence angle compared to MYI. Ultimately, this leads to the two classes (OW and MYI) being indistinguishable in HH at certain IA's.

One of the primary reasons for the different material responses to the IA effect is the surface roughness. Smooth surfaces, such as calm open water, act as specular reflectors, directing the majority of the radar signal away from the satellite. This results in very low backscatter coefficients at higher IA. Conversely, rough surfaces scatter the radar signals in multiple directions, including back towards the satellite. In this context, a surface can be considered smooth if it follows (2.6) [34, Chapter 8].

$$h < \frac{\lambda}{25 * sin\gamma} \tag{2.6}$$

where h is the roughness of the surface, λ is the wavelength of the radar signal, and γ is the depression angle. In contrast, a surface is considered rough if it follows (2.7) [34, Chapter 8].

$$h > \frac{\lambda}{4.4 * sin\gamma} \tag{2.7}$$

Chapter 3

Incidence Angle Dependence of GLCM Texture Features

Sea ice classification using SAR images presents significant challenges, as the SOD classes may not be easily distinguishable with SAR data alone. One major factor contributing to this difficulty is the IA effect. Variations in IA can cause changes in the backscatter signal, making it challenging to differentiate between different SOD classes at certain angles.

To enhance class separability, researchers have investigated additional features beyond the traditional HH and HV channels. Gray-level co-occurrence matrix (GLCM) texture features have proven to be powerful tools for improving classification accuracy by providing richer and more detailed textural information. These features have been effectively utilized in several studies. For instance, Soh and Tsatsoulis [65] demonstrated the robustness of GLCM for analyzing sea ice textures, achieving over 90% classification accuracy using a Bayesian classifier . Clausi [12] compared various extraction methods for SOD classification and found that co-occurrence probabilities offered the highest classification accuracy. Zakhvatkina *et al.* [71] employed GLCM features in a neural network-based algorithm to classify different sea ice types, achieving high accuracy rates. Leigh *et al.* [38] developed an operational ice/water classification system that utilizes GLCM texture features and backscatter data to achieve a 96% classification accuracy.

Most studies utilizing GLCM texture features for sea ice classification do not consider the relationship between GLCM texture features and IA. However, recent research suggests a potential correlation between GLCM features and IA [58, 22]. Therefore, the aim of this chapter is to evaluate the dependence of GLCM features on IA for sea ice classification. Our contributions are as follows:

- We measured the dependence of GLCM features on IA in both HH and HV channels using RADARSAT-2 data.
- We evaluated the importance of IA dependence for each texture feature using the coefficient of determination (R^2).
- We compared the class separability, obtained using the Jeffries-Matusita distance, for each texture feature with and without accounting for IA.
- We assessed the classification performance of a Bayesian classifier with and without incorporating IA dependence.
- We investigated the sensitivity of dependence, coefficient of determination, and Jeffries-Matusita Distance to GLCM parameters.

The remainder of the chapter is structured as follows: The next section introduces GLCM texture features, detailing their meanings, calculations, and the specific features used in this study. Following this, we provide a comprehensive review of previous work in the field, highlighting key findings and limitations. The subsequent section outlines our methodology, including the dataset, pre-processing steps, algorithms, and experimental design. Finally, we present our results and discuss the implications for sea ice classification.

3.1 Gray-Level Co-Occurrence Matrix (GLCM)

GLCM features represent a pivotal advancement in the field of image texture analysis. Developed by Robert M. Haralick in the early 1970s [27], GLCM features provide a second-order statistical method that goes beyond the simple, direct pixel value analyses offered by first-order texture measures such as variance [24]. GLCM uniquely evaluates the relationships between pairs of pixels based on their spatial arrangement with each other, offering deeper insights into the textural structure of an image.

GLCM operates on a specified window within an image, analyzing pixel pairs according to a defined set of parameters:

- **Interpixel Distance** (d): Specifies the spatial distance between pixel pairs.
- **Displacement Angle** (θ): The orientation for analyzing pixel pairs, common angles are 0, 45, 90, and 135 degrees.

- **Window Size** (w): Defines the window size in n-pixels x n-pixels used for GLCM calculations.
- **Number of Gray Levels** (k): The granularity of gray levels used, often reduced from 256 to a smaller number, such as 16.

The matrix $P_{w,d,\theta,k}$ represents the frequency of each pixel pair combination within the defined parameters, and is normalized to produce a probability matrix $Pr_{w,d,\theta,k}(i,j)$:

$$Pr_{w,d,\theta,k}(i,j) = \frac{P_{w,d,\theta,k}(i,j)}{\sum_{i=1}^{k}\sum_{i=j}^{k} P_{w,d,\theta,k}(i,j)} \tag{3.1}$$

This probability matrix captures how likely any two pixel levels are to appear together in a specific spatial relationship within an image window centered at pixel i,j, thus quantifying the textural patterns that may not be discernible through first-order metrics.

Fig. 3.1 illustrates the calculation of the matrix $P_{w,d,\theta,k}$ using parameters $w = 5$, $d = 1$, $\theta = 0$, and $k = 4$. Initially, a cropped section of the image, matching the specified window size w, is extracted around the pixel of interest. This cropped image is then quantized to the selected number of gray levels, in this case, 4. Subsequently, by employing the specified interpixel distance d and displacement angle θ, the frequency of all possible gray level combinations between pixel pairs is computed, resulting in the matrix $P_{w,d,\theta,k}$.

To accommodate variations such as rotation or shifts in the observation angle, which could skew the texture analysis, GLCM is typically computed across multiple principal directions—0, 45, 90, and 135 degrees—and the results are averaged. This process dilutes the influence of any particular directional texture, providing a more consistent and generalizable representation of texture across the image:

$$Pr_{w,d,k}(i,j) = \frac{1}{4}\sum_{\theta} Pr_{w,d,\theta,k}(i,j) \quad for\ \theta\ = 0, 45, 90, 135 \tag{3.2}$$

While many statically texture measures can be calculated from the GLCM (Haralick suggested 14), in this thesis we will be looking at eight texture features shown in Table 3.1. Mean will be used to calculate other features not as a feature itself. To make the features easier to explain, we divided them into three groups listed below [13].

- **Contrast Group**: This group evaluates the tonal variations between pairs of pixels within a specified window, helping to assess the contrast or intensity differences.

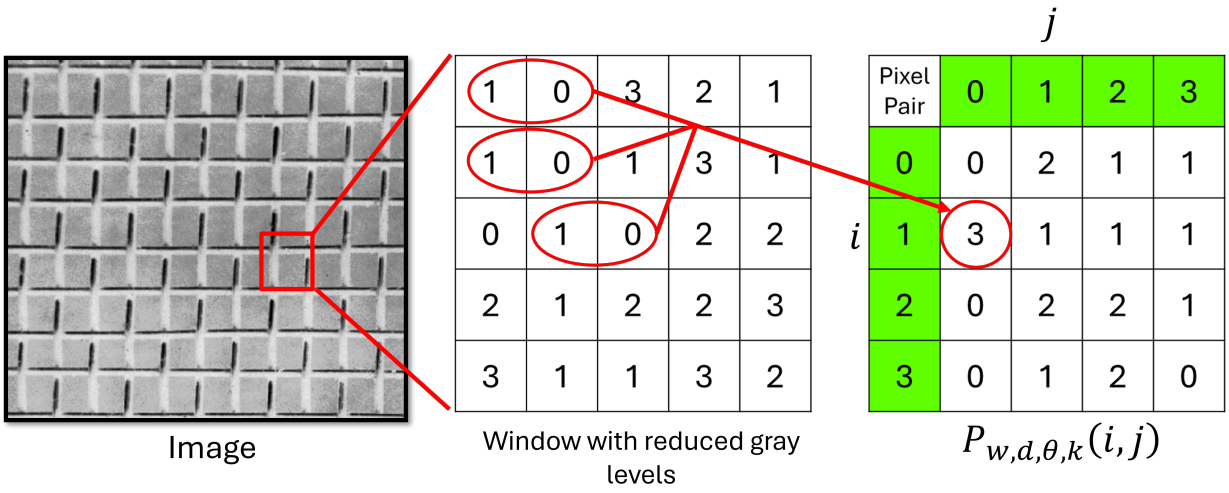


Figure 3.1: Illustration of GLCM computation.

  - **Contrast (CON)**: This feature calculates average of $(i-j)^2$, which increases as the disparity between these values increases.
  - **Dissimilarity (DIS)**: This feature calculates average of $|i-j|$, with higher values indicating greater contrast within the window.
  - **Homogeneity (HOM)**: This feature measures how similar pixels i and j are with decreasing value as $(i-j)^2$ increases.
  - **Inverse Difference (INV)**: This feature how similar pixels i and j are with decreasing value as $|i-j|$ increases.

- **Orderliness Group**: This group assesses the regularity and uniformity of pixel value differences within the window, providing insights into the structured nature of the texture.
  - **Entropy (ENT)**: This measures the randomness or complexity of the texture, increasing as the texture becomes more disordered.
  - **Angular Second Moment (ASM)**: It reflects the orderliness of the pixels inside the window, decreasing as the image texture becomes more disordered.

- **Statistical Group**: This group focuses on statistical values derived from the GLCM values, not the original pixel values

– **Standard Deviation (STD)**: Standard deviation of GLCM values.
– **Correlation (COR)**: It quantifies the linear dependency between pixels i and j.

Table 3.1: GLCM features used in this thesis.

Parameter (Abbrev.)	Derivation
Contrast (CON)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2 Pr(i,j)$
Dissimilarity (DIS)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} Pr(i,j)\|i-j\|$
Homogeneity (HOM)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{Pr(i,j)}{1+(i-j)^2}$
Inverse Difference (INV)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{Pr(i,j)}{1+\|i-j\|}$
Angular Second Moment (ASM)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} Pr(i,j)^2$
Entropy (ENT)	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} -Pr(i,j)ln[Pr(i,j)]$
Mean (μ_i)	$\sum_{i=0}^{N-1} iPr(i,j)$
Standard Deviation (STD & σ_i)	$\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu_i)^2 Pr(i,j)}$
Correlation (COR)	$\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu_i)(j-\mu_i)Pr(i,j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}$

3.2 Previous Works

Lohse *et al.* [40] stands out as the first refereed to explore the dependency between GLCM feature and IA. Their investigation focused on Sentinel-1 C-band data, analyzing only the HH polarization due to its higher signal-to-noise ratio. In contrast, HV polarization was not examined due to its lower signal strength and the presence of scalloping effects that could compromise the reliability of texture analysis. The authors found no correlation between GLCM texture features and IA when using dB units for the backscatter, which linearizes the decay in backscatter (2.2.4). This result suggests that GLCM texture features are resilient to changes in IA.

Building on the background laid by Lohse *et al.* [40], Scharien *et al.* [58] also investigated the dependence between GLCM features using HH polarization Sentinel-1 data. One notable difference in their methodology was the use of linear units for backscatter instead

of dB units as used by Lohse *et al.* [40]. The authors found that linear units enhanced the separability between FYI and MYI. Their findings, illustrated in Fig. 3.2, reveal a pronounced dependency of GLCM features on IA, especially significant for MYI, with a correlation of determination (R^2) of approximately 0.45. This demonstrates a stronger link between IA and texture characteristics in MYI compared to FYI, suggesting variable dependency across different ice types. Lohse *et al.* [40] did note that GLCM features extracted from linear units had a heavy non-linear relation with IA. This suggests that heavy dependence with IA, may only be present when linear units are used to extract GLCM features.

Finally, Guo *et al.* [22] investigated the relationship between GLCM features and IA for sea ice classification using TerraSAR-X's HH-polarization imagery. The study identifies a linear correlation between GLCM and IA features when using dB units. However, the relationship found was quite weak. Fig. 3.3 illustrates the correlations that they found.

Despite the significant contributions that these studies bring to the understanding of the relationship between GLCM features and IA, they share several limitations. None of these previous studies examined the relationship between GLCM features and IA for HV polarization, focusing solely on HH. Additionally, they do not provide information about the importance of the dependence between GLCM features and IA for sea ice classification tasks. Lastly, they do not investigate whether the dependence between the GLCM features and IA is affected by the choice of GLCM parameters.

This chapter addresses these three issues. We evaluate the dependence on both HH and HV polarizations, build a Bayesian classifier to compare results with and without considering IA dependence, and investigate the impact of GLCM parameters on the dependence between GLCM features and IA.

3.3 Methodology

3.3.1 Dataset

Our research is based on a dataset initially introduced by Leigh *et al.* [38] and subsequently augmented by Chen *et al.* [10]. Comprising 29 SAR scenes, the data was captured over the Beaufort Sea from 2010 to 2011 (all seasons) by the RADARSAT-2 C-band satellite in ScanSAR mode. The scenes feature dual-polarization (HH and HV) and span an IA range from 20° to 49°. Each image, with a pixel spacing of 50 meters covers approximately 10000 $\times$ 10000 pixels. Coarse polygon SOD labels were obtained using sea ice charts prepared by

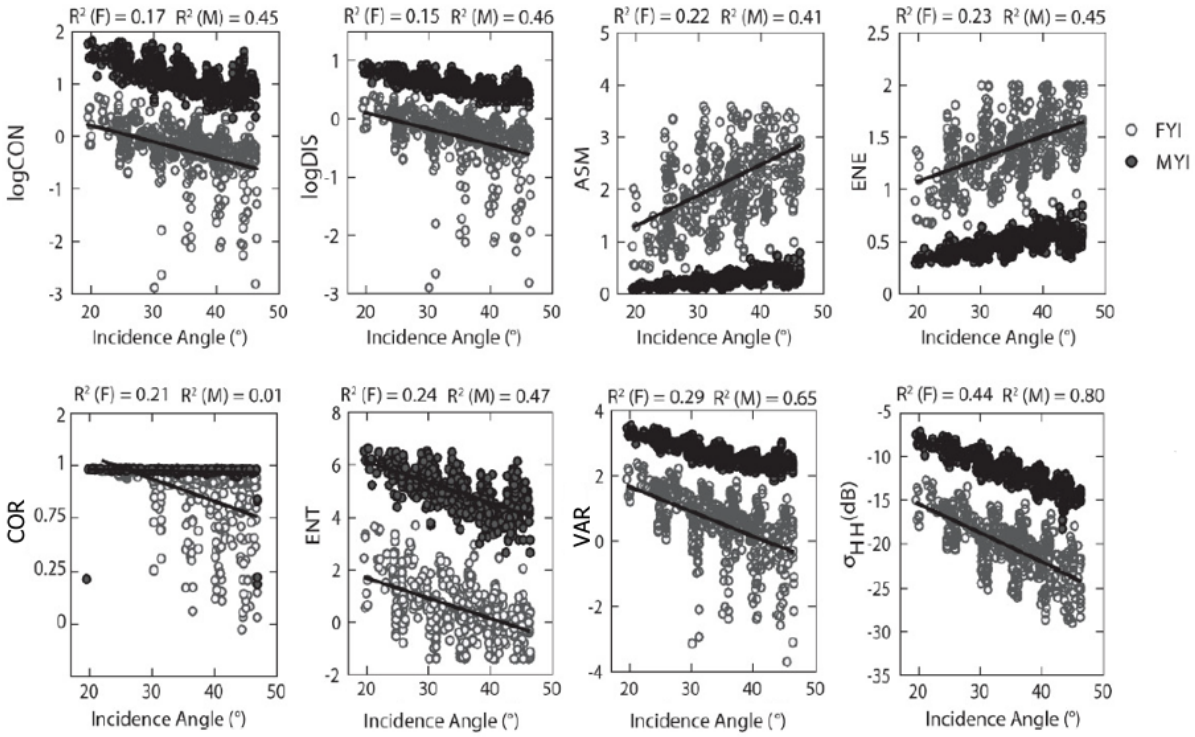


Figure 3.2: Findings from Scharien *et al.* [58], which illustrate the dependence between GLCM features and IA. The data were gathered during winter in the Canadian Arctic using Sentinel-1 C-band.

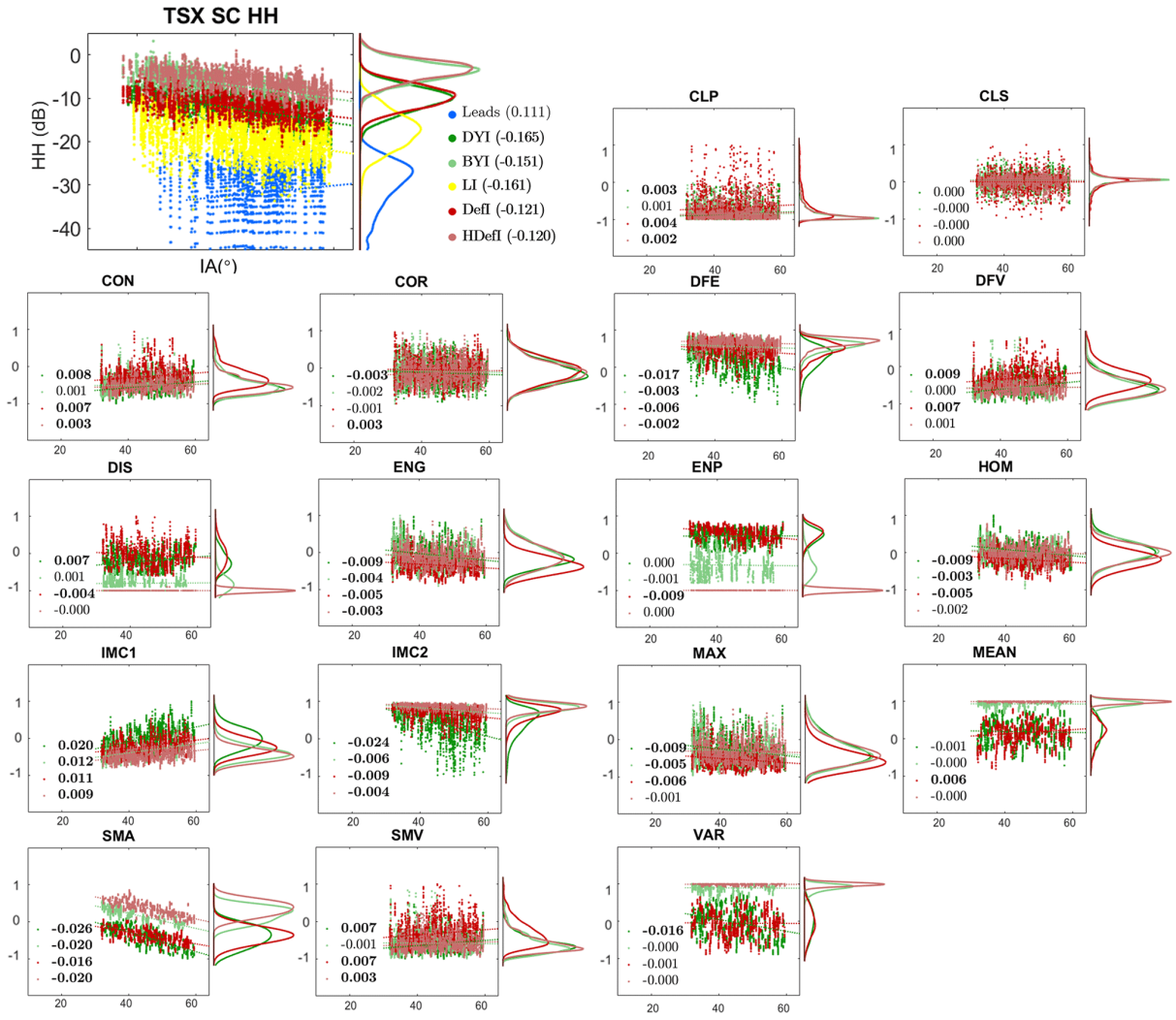


Figure 3.3: Results from Guo *et al.* [22], which show the class dependence between GLCM features and IA. The data were gathered during winter in the Arctic Ocean using TerraSAR-X. The ice classes utilized in the study are leads, dark young ice (DYI), bright young ice (BYI), level ice (LI), deformed ice (DefI), and heavily deform ice (HDefI).

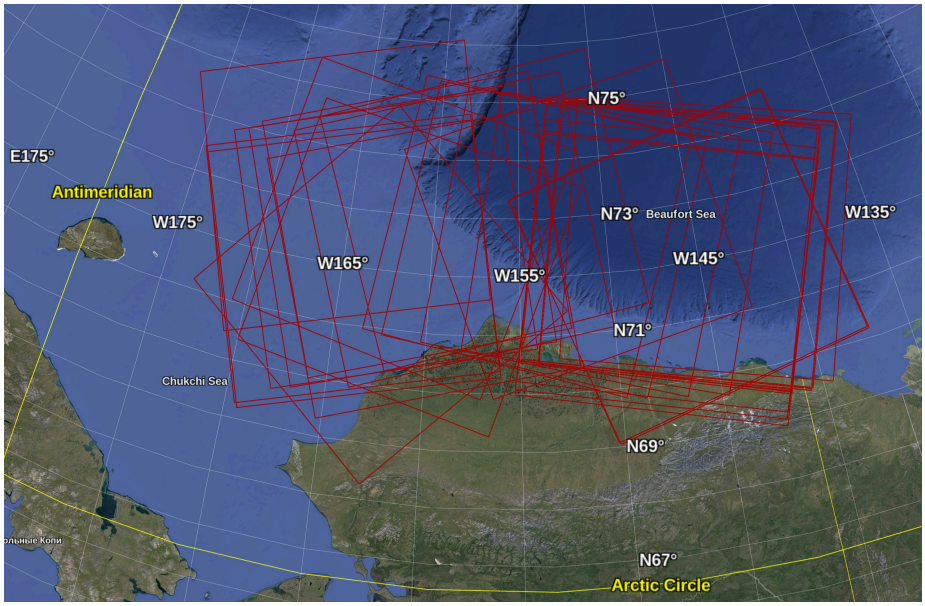


Figure 3.4: Visualization of the 29 RADARSAT-2 C-band SAR imagery scenes from the Beaufort Sea.

a trained analyst from the Canadian Ice Service (CIS). Pixel labels were then derived from these polygons through a performance evaluation framework (PEF) [48]. The execution of the PEF system involved conducting a highly detailed version of iterative region growing using semantics (IRGS) [69] classification, followed by manual labeling of the thousands of identified regions [38].

The SOD classes found in this dataset are: open water (OW), first-year ice (FYI), and multi-year ice (MYI). The geographic coverage of the dataset is depicted in Fig. 3.4, offering a visual overview of the study area.

3.3.2 Data Pre-processing

The first step in the data pre-processing is radiometric calibration to convert the raw pixel values to the normalized radar cross-section (σ°) The RADARSAT-2 radiometric calibration uses a lookup table containing beta, gamma, and sigma gain values (g) and a shared offset (b), to transform the raw pixel intensity (P) using the formula defined in Equation 3.3 [41].

$$\sigma^o = \frac{P^2 + b}{g_\sigma} \tag{3.3}$$

Subsequently, the IA (θ) of the SAR signal is derived by the ratio of σ^0 to β^0, as detailed in Equation 3.4 [64].

$$\theta = sin^{-1}(\frac{\sigma^0}{\beta^0}) = sin^{-1}(\frac{g_\beta}{g_\sigma}) \tag{3.4}$$

To linearize the relationship between backscatter intensity and IA, we convert the backscatter measurements into the logarithmic scale (dB), as shown in Equations 3.5 and 3.6. Additionally, speckle noise reduction and image downsampling are achieved through a 4x4 block averaging.

$$\sigma^\circ_{HH\,dB} = 10 * log_{10}(\sigma^0_{HH}) \tag{3.5}$$
$$\sigma^\circ_{HV\,dB} = 10 * log_{10}(\sigma^0_{HV}) \tag{3.6}$$

To extract the GLCM features, we first normalize the data, using the following equation:

$$\sigma^\circ_{norm} = \frac{\sigma^\circ_{dB} - min(\sigma^\circ_{dB})}{max(\sigma^\circ_{dB}) - min(\sigma^\circ_{dB})} \tag{3.7}$$

For the HH and HV polarizations, a different normalization range was applied due to their distinct intensity ranges. Specifically, HH is normalized using a maximum of 0 dB and a minimum of -40 dB, while HV is normalized using a maximum of -15 dB and a minimum of -40 dB.

The process of extracting GLCM features using a window, as shown in Fig. 3.1, may result in features that include data from multiple classes at their intersection. To prevent class mixing, we employ both a land mask and an edge mask. The land mask is set to true for pixels representing land in the SAR image and false elsewhere. The edge mask is created by drawing a white square (a square of true values) of the same size as the window used for GLCM feature extraction at pixels located at the edge between different ice classes. These class edges are detected using horizontal and vertical Sobel filters applied on the labels obtained from sea ice charts. GLCM features from pixels where either the land mask or edge mask is true are discarded.

3.3.3 Experiments

To evaluate the dependence of IA on GLCM features for SOD classification, we measured the following for each GLCM feature:

1. The dependence with respect to IA
2. The strength of the dependence
3. The separability between SOD classes with and without taking into account the IA

First, we separated the GLCM data by class using the labels provided. Following this, we aggregated all pixel GLCM features from all the images in the dataset. We then proceeded to construct a linear regression for each GLCM feature against the IA for each polarization and ice class as shown in the 2D histogram in Fig. 3.5. The dependence of each GLCM feature on IA is represented by the slope calculated from the linear regression. Using the slope as the measure of dependence assumes a linear relationship between GLCM features and IA. We adopted this assumption for the following reasons:

- It is the simplest form of dependence.
- Previous works (Section 3.2) suggest an approximately linear relationship.
- Our results also show that the relationship is approximately linear (Fig. 3.8).

To calculate the strength of the dependence, we used the coefficient of determination (R^2), as shown in Equation 3.8.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})} \qquad \text{Coefficient of determination} \qquad (3.8)$$

$$\sigma^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2 \qquad \text{Variance} \qquad (3.9)$$

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2 / n}{\sum_i (y_i - \bar{y}) / n} = 1 - \frac{\sigma^2_{reg}}{\sigma^2} = \frac{\sigma^2 - \sigma^2_{reg}}{\sigma^2} \qquad (3.10)$$

where y_i is the value of the GLCM feature in pixel i, f_i is the predicted GLCM value using linear regression with respect to IA, and $\bar{y}$ is the mean value of the GLCM feature. As

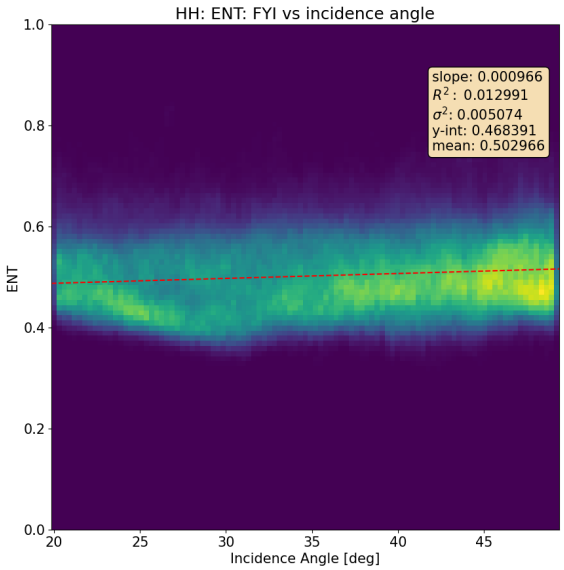


Figure 3.5: 2D histogram showing the distribution between ENT and IA for FYI in HH polarization. The plot illustrates the relationship between ENT and IA, with the color intensity representing the density of data points. The red dashed line indicates the linear regression fit, highlighting a slight positive correlation.

shown in Equations 3.9 and 3.10, R^2 is a statistical measure that determines the proportion of the variance of the dependent variable that can be explained by the independent variable. For example, an R^2 value of 0.2 indicates that the IA accounts for 20% of the variance of the GLCM feature. Therefore, R^2 indicates the importance of the dependence between GLCM features and IA, with higher values indicating higher importance.

To assess class separability based on GLCM features, we used the Jeffries-Matusita (JM) distance, which ranges from 0 (indicating non-separability) to 2 (indicating complete separability). The JM distance is calculated using the Bhattacharya coefficient and Bhattacharya distance, with $P(X)$ and $Q(X)$ representing the probability mass functions of GLCM features for different classes (e.g., OW and FYI) for a given texture feature, as shown in Equation 3.13.

$$BC(P,Q) = \sum_{x \in X} \sqrt{P(X)Q(X)} \qquad \text{Bhattacharya coefficient} \tag{3.11}$$

$$D_B(P,Q) = -\ln(BC(P,Q)) \qquad \text{Bhattacharya Distance} \tag{3.12}$$

$$JM_{(P,Q)} = \sqrt{2(1-\exp(-D_B(P,Q)))} \qquad \text{Jeffries-Matusita Distance} \tag{3.13}$$

As mentioned before, the goal is to assess how the IA influences GLCM features by measuring the JM distance, slope and R^2 values. However, these measurements may be affected by GLCM parameters as well. Instead of investigating how each combination of GLCM parameters affects JM distance, slope and R^2 values, we decided to study the set of GLCM parameters that maximizes performance of a SOD classifier.

Since the GLCM parameters (number of gray levels, interpixel distance, and window size) can influence JM distance, slope, and R^2 values, instead of studying every possible combination of GLCM parameters, we focused on studying the effect of the incidence angle (IA) on a parameter set that maximizes classification performance. To do so, we conducted a comprehensive GLCM parameter optimization study where we utilized a Bayesian classifier as our classification model. The Bayesian classifier that was chosen is the Gaussian Incidence Angle (GIA) classifier developed by Lohse *et al.* [39]. This classifier was selected due to it's simplicity and transparency, in contrast to the "black box" nature of neural networks, and its capability to linearly model the relationship between GLCM features and IA, as detailed in Equation 3.15. Furthermore, this is the same classifier used by Lohse *et al.* [40] and Guo *et al.* [22] in their studies, facilitating easier comparison of results.

$$p_i(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_i)^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-\mu_i)\right) \tag{3.14}$$

Multivariate Gaussian Distribution

$$p_i(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-(\alpha_\mathbf{i}-\beta_\mathbf{i}\theta))^T\mathbf{\Sigma}_i^{-1}(\mathbf{x}-(\alpha_\mathbf{i}-\beta_\mathbf{i}\theta))\right) \tag{3.15}$$

Multivariate Gaussian Distribution with IA dependence (GIA)

The GIA classifier functions similarly to a Bayesian classifier, where the likelihood $p(\mathbf{x}|\omega_i)$ follows a multivariate Gaussian distribution. The key difference is that the mean (μ) is not constant but is described as a linear function ($\alpha - \beta_\mathbf{i}\theta$) with respect to IA ($\theta$). To fit our model, we needed to calculate α, β, and Σ, the covariance matrix of each class. The data point $\mathbf{x} \in \mathbb{R}^{1\times d}$ (with d representing the dimensionality of the data).

The data point $\mathbf{x} \in \mathbb{R}^{1\times d}$, where d denotes the dimensionality of the data, represents a vector containing the backscatter values (σ°_{HH}, σ°_{HV} in dB) along with the GLCM texture features of a specific pixel. Consequently, the parameters $\alpha \in \mathbb{R}^{1\times d}$ and $\beta \in \mathbb{R}^{1\times d}$ are determined by performing a linear regression on the training dataset $\mathbf{X} \in \mathbb{R}^{n\times d}$, where n represents the total number of pixels. The covariance matrix Σ is also computed from the training data.

The optimization study employed leave-one-out cross-validation (LOOCV) to assess model performance. In LOOCV, the model is trained on all images except one, which is reserved for testing. This process is repeated for each image in the dataset. Given that the dataset contains 29 images, this results in the training of 29 models and the generation of 29 results, which are then averaged to produce a single performance metric. This iterative process ensured comprehensive model validation. The F1 score was chosen as the primary evaluation metric. This metric, defined in Equation 3.16, considers true positives (TP), false negatives (FN), and false positives (FP), averaging the scores across all classes to provide an overall performance measure.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN} \qquad F_1 = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.16}$$

For this study, we explored a range of parameters, including interpixel distance (d), the number of gray levels (k), and various window sizes (w), to identify the combination that maximizes classification performance. The combinations of parameters are shown below.

$$d = [1, 3, 5], \qquad k = [16, 32, 64, 128], \qquad w = [11, 15, 25]$$

These parameter combinations were chosen based on specific considerations. Clausi [13] demonstrated that classification performance often decreases as the number of gray levels increases, so we capped k at 128 to avoid unnecessary experiments with higher values. For window size, we set a maximum of 25 to preserve a larger number of MYI pixels, as larger window sizes increase the size of the edge mask, which would significantly reduce the number of MYI pixels in the dataset. Lastly, we selected a maximum interpixel distance of 5 to ensure it does not exceed half of the window size.

3.4 Results and Discussion

3.4.1 GIA Results

The results of the GIA Bayesian classifier are summarized in Table 3.2. The "Features" column indicates which features were used as input. For instance, HH+HV+$GLCM_{HH}$ denotes that HH, HV, and all texture features extracted from HH were used to train the model. The "No IA" or "IA" labels indicate whether the GIA classifier utilized the IA information by setting $\mu = \alpha + \beta * \theta$. Notably, when the GLCM features are included, the HH and HV backscatter features always utilize the IA information. This approach was adopted to isolate and observe the influence of IA specifically on the GLCM features. Table 3.2 presents four sets of experiments, each comparing the performance of the GIA classifier with a specific set of features when using IA versus not using IA.

When using GLCM features from either HH or HV, the performance increase when accounting for IA is relatively small, with a +0.91% and +0.85% increase in performance for $GLCM_{HH}$ and $GLCM_{HV}$, respectively. When using all available features (HH+HV+$GLCM_{HH+HV}$), the performance increase due to IA is very small and assumed to be insignificant (+0.06%). Notably, these performance gains are primarily driven by increases in the F1 scores for FYI and MYI. On the other hand, accounting for the IA effect often resulted in a reduction of the F1 score for OW, although the overall performance still showed a net increase.

These results are not entirely unexpected. When only utilizing HH+HV, the performance increase with IA is just +1.33%, despite the strong effect that IA has on HH polarization as shown in Fig. 3.6. The results also indicate that as more variables are added that can separate the SOD classes, the need to account for the IA effect diminishes. Overall, these results suggest that when using HH, HV, and GLCM features, the IA effect plays a minor role in classification performance.

3.4.2 Slope and R^2

In this subsection, we examine the individual characteristics of each GLCM feature. Evaluating how each individual feature changes with respect to all GLCM parameters is not realistic. As such, we utilized the GLCM parameters that produced the best-performing GIA model with all features (d=1, k=32, w=25).

Fig. 3.7 presents the slope and R^2 values for the GLCM features. Generally, the R^2 values are quite low, suggesting that the IA effect explains, at most, 25% of the variance in

Table 3.2: Results from GIA Bayesian classifier. The "Features" column indicates the set of features the model takes as input. The "No IA" or "IA" indicates the use of IA information by setting $\mu = \alpha + \beta * \theta$. The numbers in () indicate the performance change with respect to the model that does not utilize IA.

Features	**d**	**k**	**w**	**Overall**	**OW**	**FYI**	**MYI**
HH+HV: No IA	3	64	25	44.24%	63.50%	53.17%	8.53%
HH+HV: IA	3	64	25	45.57% (+1.33%)	64.94% (+1.43%)	54.21% (+1.05%)	9.15% (+0.62)
HH+HV+$GLCM_{HH}$: No IA	1	32	11	55.27%	64.13%	56.91%	13.50%
HH+HV+$GLCM_{HH}$: IA	5	64	11	56.18% (+0.91%)	60.60% (-3.53%)	60.44% (+3.53%)	12.70% (-0.80%)
HH+HV+$GLCM_{HV}$: No IA	5	64	11	51.49%	59.17%	64.56%	10.44%
HH+HV+$GLCM_{HV}$: IA	5	64	11	52.35% (+0.85)	59.14% (-0.02%)	64.72% (+0.17%)	11.02% (+0.58%)
HH+HV+$GLCM_{HH+HV}$: No IA	1	32	25	59.25%	61.44%	68.01%	14.35%
HH+HV+$GLCM_{HH+HV}$: IA	1	32	25	59.31% (+0.06%)	60.97% (-0.47%)	68.48% (+0.47%)	14.62% (+0.28%)

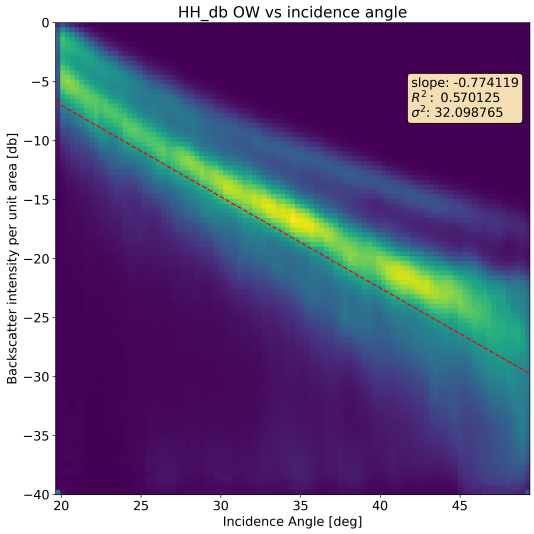


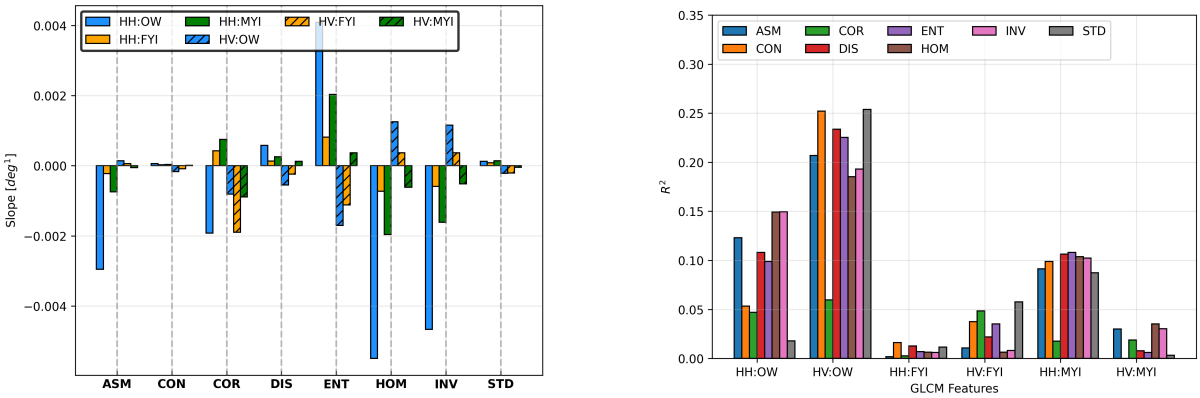
Figure 3.6: 2D histogram showing the relationship between backscatter [dB] of open water and the IA. Data from all the scenes were used to create this histogram.

the GLCM features. This indicates that the influence of IA on GLCM features is generally weak, meaning that changes in IA have only a minor effect on the values of these features. The IA effect is the strongest for GLCM features extracted in OW, particularly for the HV polarization ($R^2 \approx 0.22$). In contrast, GLCM features derived from FYI exhibit a very weak relationship with IA ($R^2 \approx 0.025$), while those from multi-year ice MYI show a slightly stronger relationship, especially in HH polarization ($R^2 \approx 0.09$).These observations corroborate the GIA classifier's outcomes, where the minimal impact of IA on GLCM features results in only a slight performance gain when accounting for IA.

These findings are consistent with previous studies by Lohse *et al.* [40] and Guo *et al.* [22], who also reported a weak relationship between GLCM features extracted from σ°_{HH} in dB and IA. This consistency is notable despite several key differences between the studies. Firstly, the study areas differ: Guo focused on the High Arctic Ocean, Lohse examined regions across the entire Arctic Circle, while our study is centered on the Beaufort Sea. Additionally, while both Lohse and Guo used data exclusively from the winter season, our study incorporates data collected throughout the year. Another difference lies in the satellite platforms used; Lohse utilized data from Sentinel-1, and Guo used TerraSAR-X, which operates in the X-band. In contrast, our analysis is based on data from RADARSAT-2, which operates in the C-band. Despite these differences in study areas, seasonal coverage, and satellite systems, the observed weak relationship between GLCM features and IA remains consistent across the studies. Notably, these results are only valid, for GLCM features extracted from backscatter in dB, as Lohse *et al.* [40], Scharien *et al.* (2021) [58] and Guo *et al.* [22] noted that a nonlinear relationship between GLCM features and IA is present when the features are extracted σ° in linear units.

Figs. 3.8 and 3.9 show the 2D histograms for DIS, ENT, and COR. For DIS and ENT, the slope with respect to IA is clearly visible, especially for OW in HV. This aligns with the findings in Fig. 3.7, where both ENT and DIS have relatively high R^2 values for HV:OW. In contrast, COR tends to have a very low R^2, as shown in Fig. 3.9, with the histogram appearing quite flat and having a relatively large variance.

Additionally, some distributions do not follow a Gaussian distribution. For example, DIS in HH: FYI tends to have a positively skewed distribution, with the tail not being significantly affected by IA. In contrast, the DIS feature in HH for OW is heavily dependent on IA, with the tail being largest at high IA. The IA effect on the tail might be the reason for the drop in performance of the GIA classifier for HH+HV+$GLCM_{HH}$: OW when the IA effect was considered, as the DIS: OW distribution in HH does not maintain a linear relationship at higher IA levels. For DIS in HV, the distributions do not have such long tails, leading to a smaller variance. However, they do suffer from inter-scan banding, which gives the signal a sinusoidal pattern.



(a) Slope for each GLCM feature.

(b) $|R^2|$ for each GLCM feature.

Figure 3.7: Analysis of GLCM features in relation to IA across different ice classes and polarizations. (a) Slope of the linear regression between GLCM features and IA for each feature and ice class, indicating the dependence of GLCM features on IA. (b) Coefficient of determination (R^2) for each feature and ice class, showing the strength of the dependence between GLCM features and IA. Each plot was obtained using GLCM texture features calculated with the following parameters: interpixel distance (d) = 1, number of gray levels (k) = 32, and window size (w) = 25.

An example of the interaction between inter-scan banding and GLCM features can be seen in Fig. 3.10, which shows the inter-scan banding in the HV image and it's corresponding homogeneity feature map. The bands present in the HV image are also present in HOM feature map but they are shifted since HOM does not follow the intensity of HV image but rather changes in intensity also know as texture.

3.4.3 Class Separability

Fig. 3.11 (a) shows the JM separation distance for GLCM feature and class combinations when accounting for IA. Overall, OW - FYI and FYI - MYI tend to have low separability (≈ 0.5), while OW-MYI tends to have high separability (> 1). This indicates that FYI is difficult to separate from OW and MYI, but MYI is easier to separate from OW. GLCM features focused on contrast (COR, DIS, HOM, INV) tend to have lower separability in HV than in HH. In contrast, COR, ENT, and STD tend to have the same or higher separability in HV compared to HH. This is particularly true for COR, which has much lower separability in HH than other GLCM features but much higher separability in HV.

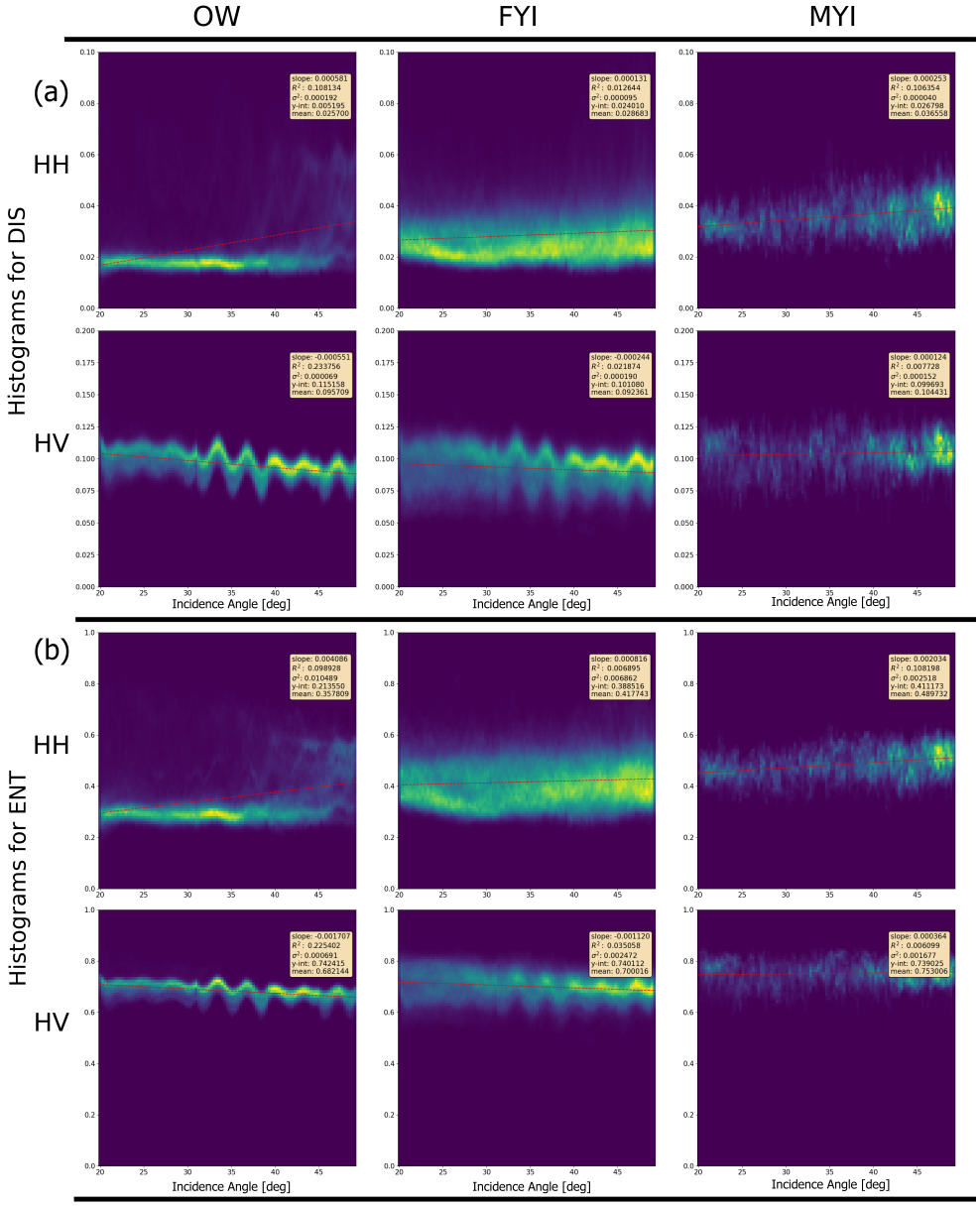


Figure 3.8: 2D histograms showing the relationship between IA and GLCM features for different ice classes and polarizations. (a) Histograms for DIS across OW, FYI, and MYI classes for HH and HV polarizations. (b) Histograms for ENT across OW, FYI, and MYI classes for HH and HV polarizations. Each histogram was obtained using GLCM texture features calculated with the following parameters: interpixel distance (d) = 1, number of gray levels (k) = 32, and window size (w) = 25.

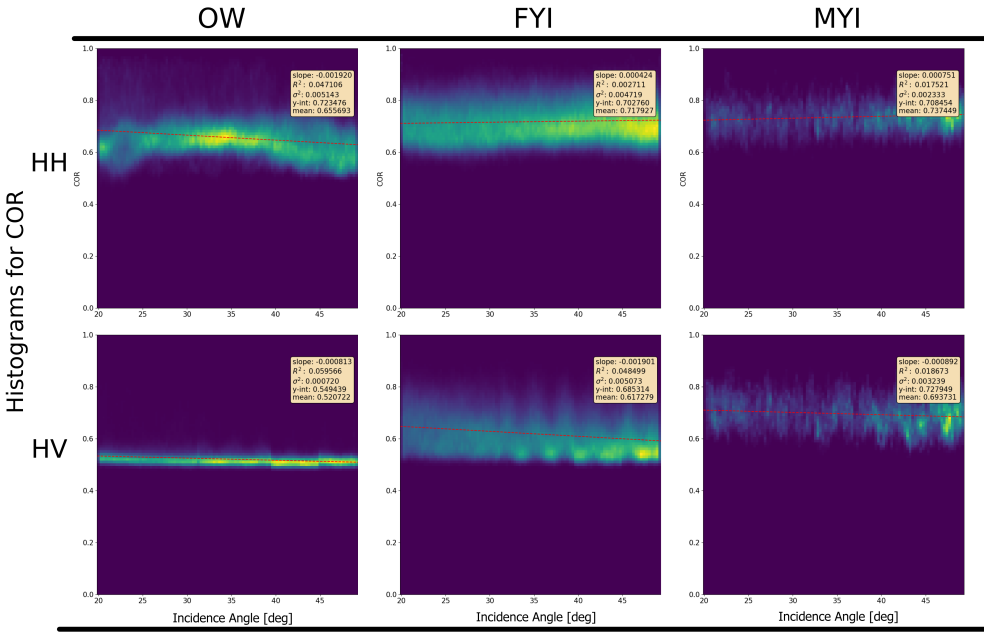


Figure 3.9: 2D histograms illustrating the relationship IA and the GLCM feature COR across different ice classes and polarizations. Each histogram was obtained using GLCM texture features calculated with the following parameters: interpixel distance (d) = 1, number of gray levels (k) = 32, and window size (w) = 25.

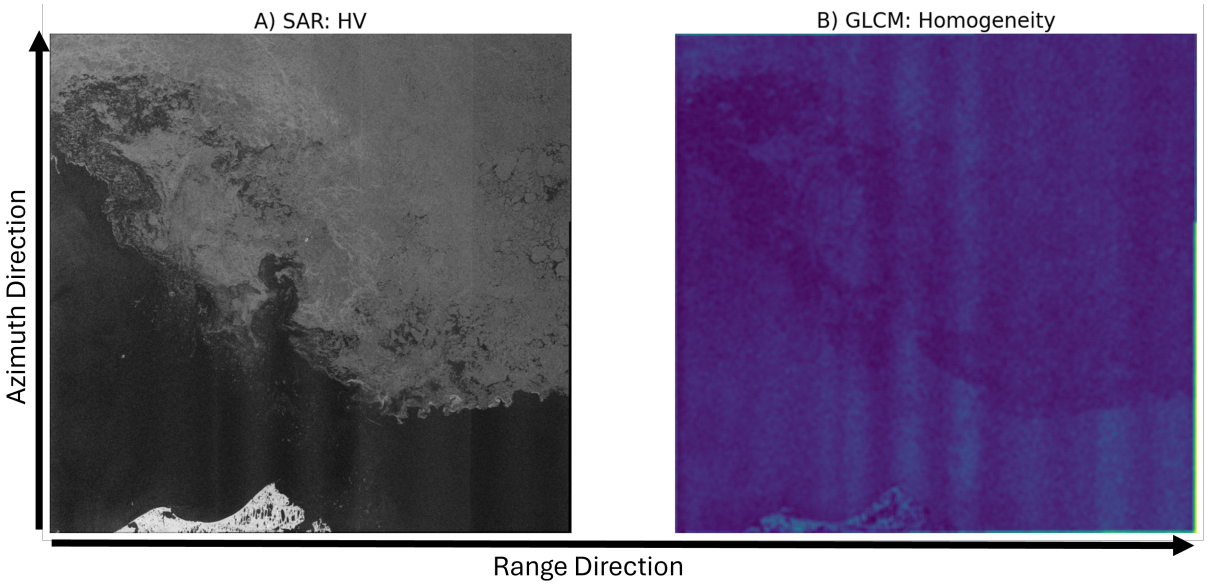


Figure 3.10: Illustration of the impact of inter-scan banding on GLCM features. A) displays the HV channel from the SAR image, where inter-scan banding is evident. B) shows the corresponding GLCM homogeneity feature map, where brighter areas indicate regions of higher homogeneity. The banding effects in the SAR image are reflected in the GLCM feature map, demonstrating how these artifacts can propagate into derived texture features. File Name: 20110720, captured on 2011, July 20^{th}.

These results are similar to those found by Lohse *et al.* [40], who reported that OW - MYI is quite separable, whereas FYI - MYI has low separability across all GLCM features. Furthermore, similar to our results, Lohse *et al.* found that CON has low separability in all class combinations for HH. However, Lohse *et al.* [40] found OW - FYI to be quite separable, in contrast to our results, which showed the complete opposite from our results. Several factors may explain this discrepancy. First, Lohse's study utilized a much larger window size (51), which likely contributed to higher separability in the data. Additionally, all samples in Lohse's study were obtained during the winter season, whereas this dataset includes samples from different seasons.

Fig. 3.11 (c) shows the percentage increase in separability when accounting for the IA effect. Adding the IA effect increases the separability of all features. However, the increase is mainly seen in features with low separability (< 0.25), with the best example being HOM in HV:OW vs. MYI, which shows an increase in separability by approximately 500%. It is also notable that most of the increases occur in HV, possibly because HV GLCM features tend to be less separable than GLCM features from HH. These results may also explain why the GIA classifier did not show significant improvement in classification performance when incorporating IA dependence, as improving features that are not very separable does not significantly enhance the overall separability of the SOD classes.

3.4.4 GLCM Parameter Dependence

This subsection explores the importance of GLCM parameters for IA dependence. In the first part, we investigate the effect of GLCM hyperparameters on class separability. Fig. 3.12 shows the effect of each parameter on the JM distance for ENT. Each row in Fig. 3.12 has three plots, each showing the change in JM distance with respect to the number of gray levels, distance, and window size. Each line in the plots represents a set of samples where only one parameter changes while the others remain constant. For example, in a plot comparing JM distance and window size, each line contains points with the same number of gray levels and distance, with only the window size varying. Each row shows the JM distance for specific polarizations and class comparisons (HH: OW & FYI).

Figure 3.12 reveals several key patterns. First, larger window sizes tend to increase the separation distance between classes. Second, separability peaks at a low number of gray levels (≈ 25). Finally, the interpixel distance appears to have little to no effect on class separability. These findings align with existing literature; Lohse *et al.* [40] also observed that GLCM separability improves with larger window sizes, and Clausi [13] found that for many GLCM features, classification accuracy peaks at low numbers of gray levels.

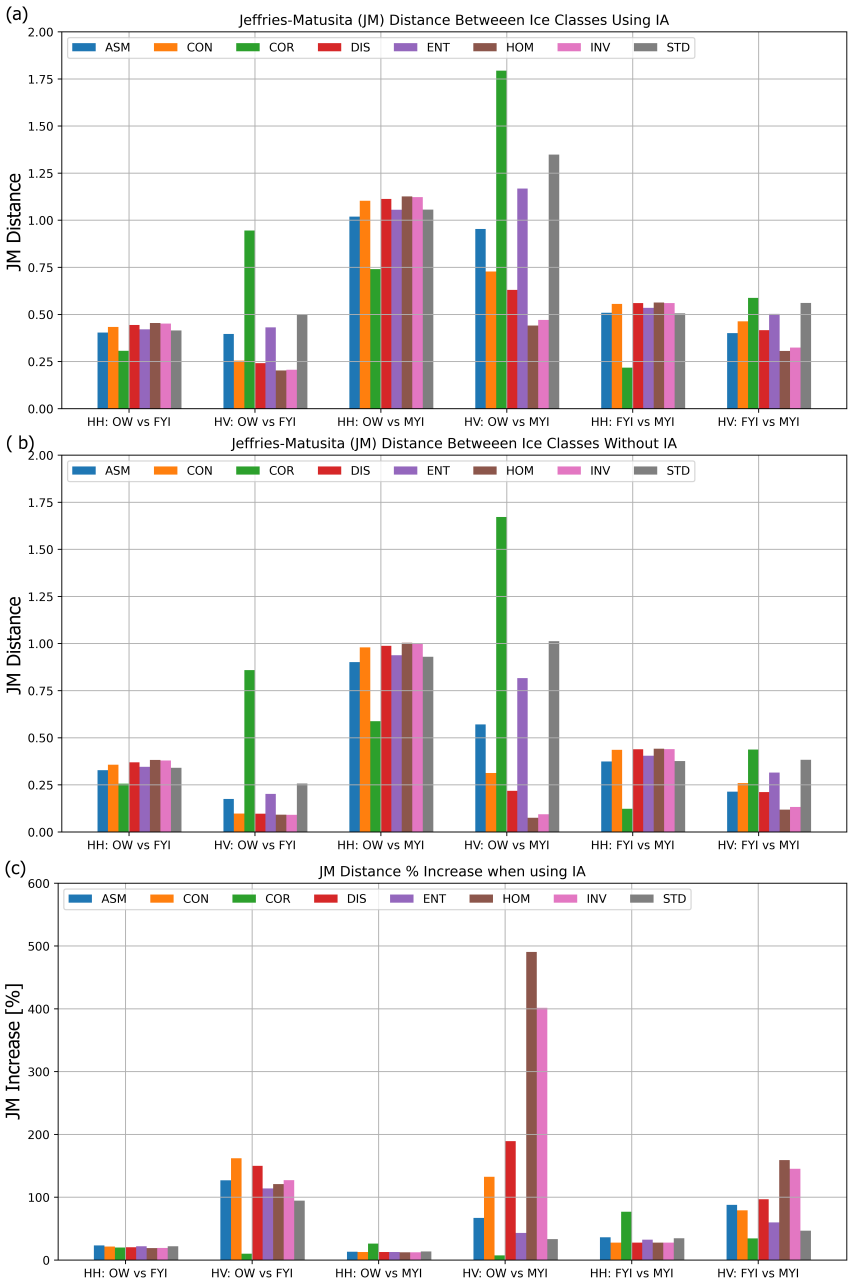


Figure 3.11: JM distance Analysis for GLCM features across different ice classes and polarizations. (a) JM distance between ice classes incorporating the IA dependence. (b) JM distance between ice classes without incorporating the IA dependence. (c) Percentage increase in JM distance when incorporating the IA dependence, illustrating the impact of IA on the separability of ice classes.

Table 3.3: Percentage increase in JM Distance obtained by changing GLCM parameters.

GLCM Features	HH			HV		
	OW & FYI	OW & MYI	FYI & MYI	OW & FYI	OW & MYI	FYI & MYI
ASM	179.66%	104.47%	108.12%	379.79%	201.60%	134.68%
ENT	105.64%	65.24%	89.75%	347.02%	219.72%	158.92%
CON	155.93%	92.98%	106.10%	221.31%	224.15%	149.90%
DIS	184.31%	108.02%	117.83%	223.60%	241.44%	151.75%
HOM	197.32%	116.37%	129.76%	416.80%	608.27%	327.41%
INV	210.73%	125.91%	133.68%	260.63%	418.40%	248.50%
COR	1525.26%	1242.50%	435.62%	4060.84%	1817.93%	1196.57%
STD	122.00%	72.35%	85.32%	96.48%	77.28%	89.07%

Additionally, the overall pattern between parameters and GLCM changes very little with different class combinations (OW & FYI, FYI & MYI). However, changes in polarization lead to significant changes in the relationship between GLCM parameters and separability. While not shown in Fig. 3.12, similar GLCM features like DIS and CON or ASM and ENT show similar patterns with respect to GLCM parameters and separability. Table 3.3 shows the percentage increase in JM distance obtained by changing GLCM parameters. These increases are often larger than 100%, indicating that parameter optimization is crucial for class separability, with window size and number of gray levels being the most important parameters.

Figure 3.13 illustrates how R^2 varies with different GLCM parameters. Unlike the JM distance, there is no consistent pattern across polarizations or ice types; instead, patterns are observed primarily among similar GLCM features, such as CON and DIS. Table 3.4 highlights the variation between the maximum and minimum R^2 values. Most of these values exceed 100%, indicating that the influence of IA on GLCM features is highly dependent on the specific GLCM parameters used. This effect is particularly pronounced for COR, where the large percentage increases are due to the minimum R^2 being close to zero, with the slope occasionally crossing zero as the GLCM parameters change, as shown in Figure 3.14. Notably, correlation is the only feature associated with a reversal in slope direction when altering the GLCM parameters. This may not too surprising as correlation has a very low R^2 value, indicating that IA has a very weak or not effect on it. These findings suggest two key conclusions: the direction of the relationship between GLCM features and IA is generally independent of the GLCM parameters, whereas the strength of this relationship is highly dependent on the specific GLCM parameters used.

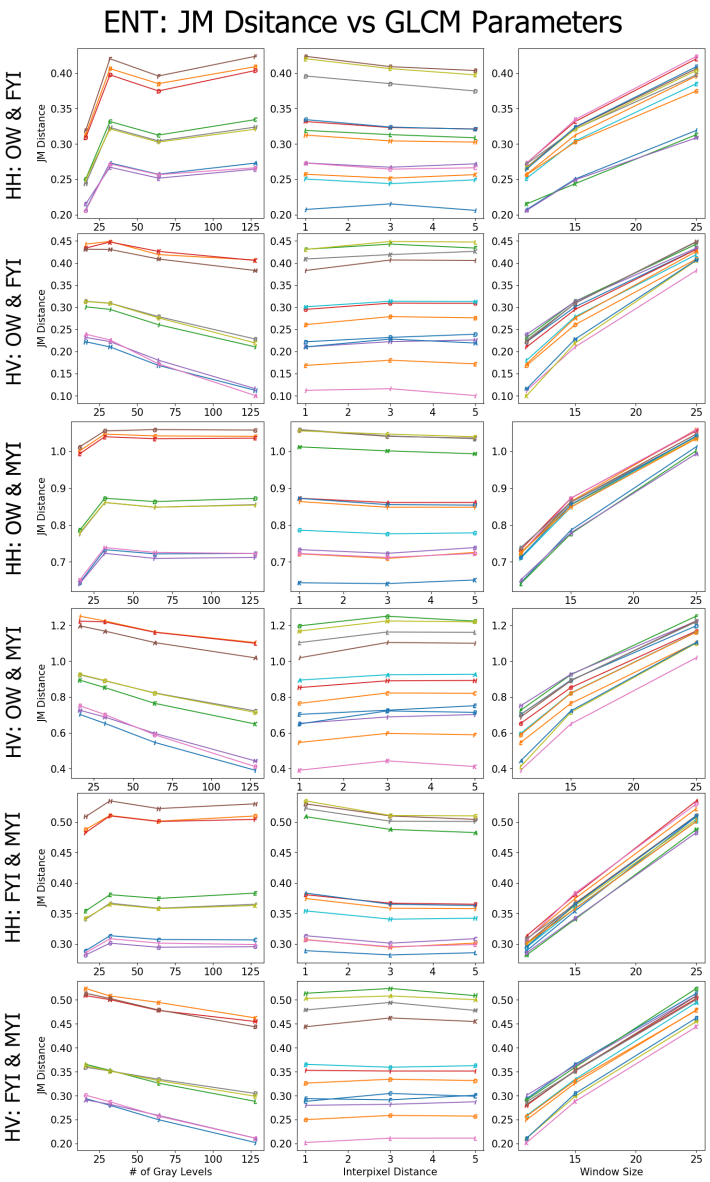


Figure 3.12: Each plot shows the change in JM distance with respect to one parameter (gray levels, interpixel distance, or window size) while holding the other two parameters constant. Each line in the plot represents a unique set of samples. Colors are utilized to differentiate the lines, with no significance attached to lines sharing the same color.

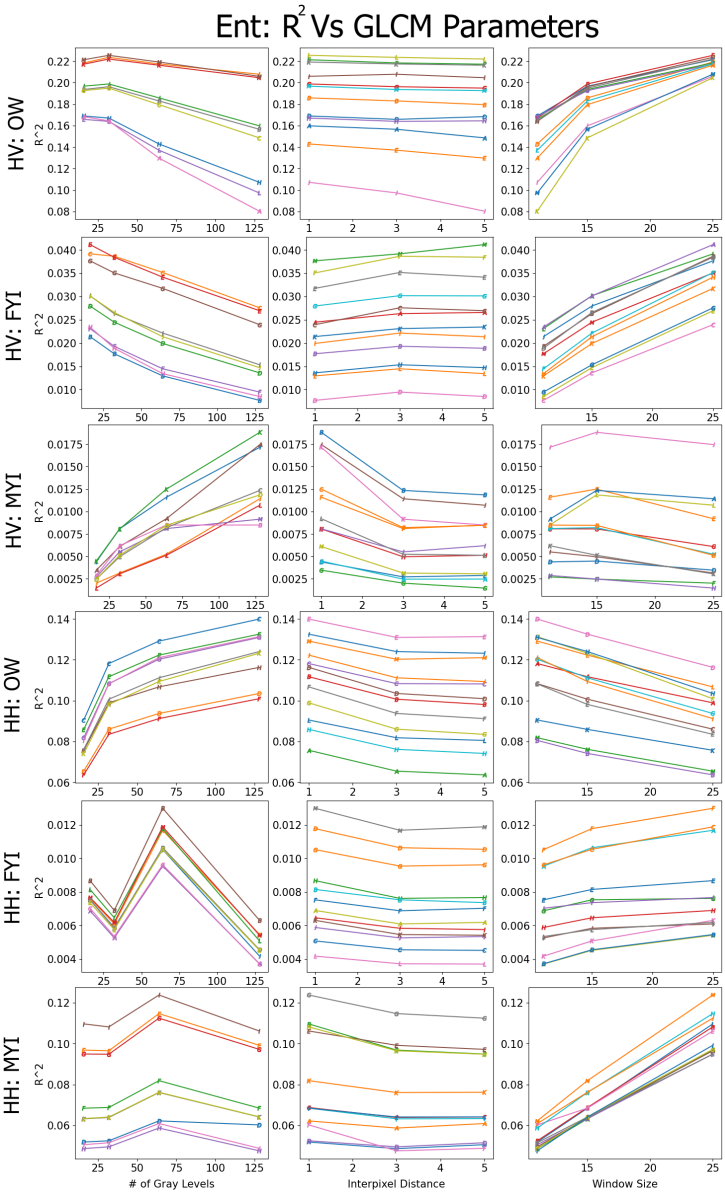


Figure 3.13: Each plot shows the change in R^2 with respect to one parameter (gray levels, interpixel distance, or window size) while holding the other two parameters constant. Each line in the plot represents a unique set of samples. Colors are utilized to differentiate the lines, with no significance attached to lines sharing the same color.

Table 3.4: Percentage increase between the maximum and minimum R^2 values obtained by changing GLCM parameters.

	HH			HV		
	OW	**FYI**	**MYI**	**OW**	**FYI**	**MYI**
ASM	105.17%	661.30%	165.48%	198.90%	695.00%	232.16%
ENT	119.64%	251.63%	160.22%	180.39%	434.54%	1179.82%
CON	1318.31%	293.66%	249.05%	72.64%	181.55%	6667.44%
DIS	199.62%	173.41%	206.73%	72.83%	252.93%	1272380.44%
HOM	153.46%	310.35%	177.79%	276.97%	1262.18%	850.20%
INV	142.57%	228.63%	171.73%	133.41%	741.03%	1027.89%
COR	15411691.87%	2528007.80%	90317.28%	6464717.98%	38432310.26%	460909.54%
STD	238.32%	83.38%	158.39%	39.04%	66.26%	230413.87%

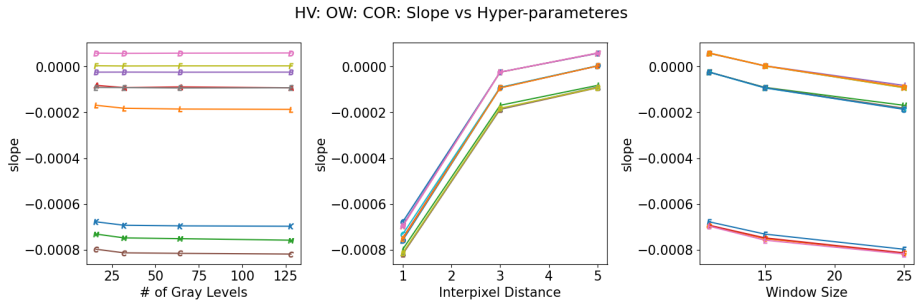


Figure 3.14: Slope of the linear regression between COR and IA as a function of different GLCM parameters for HV polarization in OW ice class. Each line represents a set of samples where only one parameter changes while the others remain constant.

3.5 Conclusion

To conclude, our study involved extracting GLCM features from SAR imagery in dB units, focusing on data from the Beaufort Sea that included the SOD classes of OW, FYI, and MYI. We aggregated data from all images and constructed a linear regression for each GLCM feature against the IA for each polarization and ice class. The dependence measure was determined by the slope from the linear regression, and the strength of dependence was quantified using the coefficient of determination (R^2). Class separability was assessed using the Jeffries-Matusita distance for each class combination, GLCM feature, and polarization. Recognizing the influence of GLCM parameters on results, we optimized these parameters for classification performance by selecting those that maximized the performance of the Gaussian Incidence Angle Bayesian classifier [39], ensuring accuracy through leave-one-out cross-validation.

The results indicate that incorporating IA information into the classification process with the GIA Bayesian classifier provides only marginal performance improvements. There were small increases in classification accuracy, such as +1.33% for HH+HV and less than 1% for other combinations involving GLCM features. This suggests that the IA effect is relatively minor when using comprehensive feature sets like HH, HV, and GLCM features. Our analysis also revealed that the relationship between IA and GLCM features is generally weak, with $\mathbf{R^2}$ values accounting for at most 25% of the variance. The IA effect was more pronounced in OW, especially for HV polarization, but remained weak for FYI and MYI. Class separability increased with the inclusion of IA, particularly for features with initially low separability, but this increase was not substantial enough to make features with low separability (< 0.25) highly separable (> 1). Our findings align with previous works [40, 22], despite differences in study areas, seasonal coverage, satellite systems, and signal frequency, suggesting that the IA effect may not be a critical factor in sea ice classification for GLCM features extracted from SAR images in dB units.

Chapter 4

Data Input Selection for Automated Sea Ice Mapping

In recent years, the application of machine learning (ML) and deep learning (DL) techniques, particularly convolutional neural networks (CNNs), has significantly advanced automated sea ice parameter estimation from dual-polarized SAR imagery [68, 2, 46, 17, 67, 53]. However, as mentioned in Section 2.2.4 the similar backscatter signatures produced by distinct incidence angle, ice types, varying concentrations, different wind and snow depth and wetness conditions pose a challenge for accurate mapping when relying solely on SAR imagery [43].

To address these challenges, researchers have increasingly turned to multi-sensor approaches that integrate various data sources to resolve ambiguities and improve accuracy [20]. Notable examples include the fusion of Sentinel-1 SAR imagery with AMSR2 passive microwave data for sea ice concentration (SIC) estimation [43, 35], the use of SAR and optical images to predict the stage of development (SOD) [26], and the classification of ice and open water using HH polarization combined with multi-spectral data from MODIS [55]. In addition to these multi-sensor approaches, incorporating spatiotemporal information has proven beneficial for enhancing model performance. For instance, separating SAR data based on acquisition month or season to train specific sea ice classifiers has shown to be effective [49].

The AI4Arctic Sea Ice Challenge dataset [4], released in 2022, represents a significant resource for advancing DL-based sea ice mapping. This dataset includes Sentinel-1 SAR images, passive microwave data from the Advanced Microwave Scanning Radiometer 2 (AMSR2), numerical weather prediction data from the ECMWF Reanalysis v5 (ERA5),

and other ancillary data. With the abundance of satellite and ancillary data, to understand the efficacy of various data inputs in enhancing model performance is crucial. Despite the importance of this topic, a comprehensive investigation into the significance of different data inputs across multiple sea ice mapping tasks remains lacking.

This study builds upon our previous preliminary research publications [9, 7] by conducting a thorough analysis of the various data inputs and their contributions to sea ice mapping. Using ablation studies and statistical significance testing, we identify which features are essential for improving sea ice mapping performance, which features have minimal impact, and which are redundant. Unlike our earlier research efforts, this analysis focuses on understanding the impact of different features on the performance of key sea ice parameters: SOD, SIC, and FLOE, along with their respective subclasses. We also explore the reasons behind the varying influences of these features on specific classes. Based on our findings, we provide well-informed recommendations for selecting the most effective features when developing DL models for sea ice mapping.

The chapter is organized as follows: Section 4.1 provides a brief introduction to deep learning models and their applications in sea ice mapping. Section 4.2 introduces the AI4Arctic Challenge dataset and describes the various types of data it contains. The methodology, including the CNN architecture and experimental setup, is detailed in Section 4.3. Finally, Section 4.4 presents the experimental results, along with comparisons and discussions.

4.1 Deep Learning

This section provides an introduction to deep learning (DL), offering a foundational understanding necessary to grasp the concepts presented in this chapter.

4.1.1 Training Deep Learning Models

Training deep learning models involves several key steps: feeding the model with data, calculating the loss (or error) between the predictions and the actual data, and adjusting the network's weights to minimize this loss over time [3]. This process requires careful handling of the following components:

Data

For supervised training (which is the focus of this thesis), the data comprises training samples ($\mathbf{x}$) and their corresponding ground truth labels ($\mathbf{y}$). For example, when training a classifier to distinguish between dogs and cats, the images of dogs or cats serve as $\mathbf{x}$, and $\mathbf{y}$ represents the labels (either "dog" or "cat"), however the labels are represented with numbers such as 0 and 1 for cat and dog respectively.

Model

The deep learning model, denoted as $f(\cdot)$, is parameterized by its learnable parameters or weights ($\boldsymbol{\theta}$). When data is fed into the model, it produces a prediction: $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{\theta})$.

Loss Functions

The choice of loss function depends on the specific prediction task. For instance, regression tasks may use mean squared error (MSE), while classification tasks often employ cross-entropy loss:

$$\text{MSE} = \mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{4.1}$$

$$\text{Cross-Entropy } = \mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{4.2}$$

where N is the number of samples, y_i is the true label represented by a real number, and $\hat{y}_i$ is the predicted probability of the positive class for classification or a real number for regression tasks.

Backpropagation and Optimization

Training deep learning models involves minimizing the loss function $\mathcal{L}$. This requires tuning each weight (θ_i) in the model. The most efficient method for this is backpropagation, which computes the gradient of the loss function with respect to each weight using the chain rule, moving backwards from the output to the input layer. This gradient information is then

used to update the weights using an optimization algorithm such as stochastic gradient descent (SGD) [54]:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L} \tag{4.3}$$

where θ represents the weights, η is the learning rate, and $\nabla_\theta \mathcal{L}$ is the gradient of the loss function. Advanced optimizers such as Adam [36], which adjust the learning rate during training, are often used to improve convergence.

Agorithm 1 shows how a deep neural network is trained. In this context, an epoch refers to a complete pass through the entire training dataset, while a batch represents a smaller subset of the training data. During each batch, the model parameters θ are updated iteratively until all epochs are completed.

Algorithm 1 Training a Deep Neural Network

Input: Training data $(\mathbf{x}, \mathbf{y})$, number of epochs E, learning rate η
Output: Trained model parameters θ

Initialize parameters θ

for *each epoch* $e \in \{1, \ldots, E\}$ **do**
  for *each batch* $(\mathbf{x}, \mathbf{y})$ *in training data* **do**
    $\hat{\mathbf{y}} = \text{model}(\mathbf{x}; \theta)$ `// Forward pass`
    $\mathcal{L}$= loss_function$(\mathbf{y}, \hat{\mathbf{y}})$ `// Compute loss`
    $\nabla_\theta \mathcal{L} = \text{backward}(\mathcal{L})$ `// Backward pass`
    $\theta = \theta - \eta \nabla_\theta \mathcal{L}$ `// Update weights`
  end
end

4.1.2 Neural Networks

The concept of perceptrons, introduced by Frank Rosenblatt in 1958 [57], serves as the foundation for many neural network architectures. Perceptrons are simple binary classifiers that can linearly separate data by adjusting weights based on input features. Mathematically, a perceptron computes the weighted sum of the input features and applies a step function to determine the output:

$$\hat{y} = \sigma(\mathbf{w}^T\mathbf{x} + b) \tag{4.4}$$

$\mathbf{x} \in \mathbb{R}^{1\times d}$ is the input vector, $\mathbf{w} \in \mathbb{R}^{1\times d}$ is the weight vector, d represents the dimensionality of both the input and weight vectors, b is the bias, and σ is the step function, defined as:

$$\sigma(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

This linear separability is achieved by finding an optimal hyperplane that can distinguish between different classes in a dataset.

Building on the idea of perceptrons, multilayer perceptrons (MLPs) incorporate multiple layers of neurons, including hidden layers, which enable them to model more complex interactions. An MLP consists of an input layer, one or more hidden layers, and an output layer. Each neuron in a hidden layer applies a non-linear activation function, such as the sigmoid or the rectified linear unit (ReLU), to the weighted sum of its inputs. This allows MLPs to capture intricate patterns in data that single-layer perceptrons cannot.

The operation of an MLP can be represented mathematically as follows:

$$\mathbf{h}^{(1)} = \sigma((\mathbf{W}^{(1)})^T\mathbf{x} + \mathbf{b}^{(1)}) \tag{4.6}$$

$$\mathbf{h}^{(2)} = \sigma((\mathbf{W}^{(2)})^T\mathbf{h}^{(1)} + \mathbf{b}^{(2)}) \tag{4.7}$$

$$\vdots$$

$$\hat{\mathbf{y}} = (\mathbf{W}^{(L)})^T\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)} \tag{4.8}$$

where $\mathbf{x} \in \mathbb{R}^{1\times d^{(0)}}$ is the input, $\mathbf{h}^{(l)} \in \mathbb{R}^{1\times d^{(l)}}$ represents the activations at layer l, $\mathbf{W}^{(l)} \in \mathbb{R}^{d^{(l)}\times d^{(l-1)}}$ and $\mathbf{b}^{(l)}$ are the weights and biases at layer l, and σ is a non-linear activation

function. Traditionally, the sigmoid function was commonly used as an activation function in MLPs:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{4.9}$$

However, ReLU has become the preferred choice due to simplicity and effectiveness when training deep networks. ReLU activation is defined as:

$$\sigma(z) = \max(0, z) \tag{4.10}$$

A significant theoretical foundation for MLPs is the universal approximation theorem, which states that an MLP with at least one hidden layer containing a sufficient number of neurons can approximate any continuous function to any desired degree of accuracy [31].

4.1.3 Convolutional Neural Networks (CNNs)

One limitation of MLPs or fully connected networks (FCNs) is that they treat input data as a flat vector. For even small images, this approach requires a very high number of parameters to extract features, leading to inefficiencies and a high susceptibility to overfitting when dealing with raw pixel data. While FCNs theoretically should be capable of handling image data, their practical performance is often suboptimal.

To address these limitations, we can leverage the inherent structure in image data by making several key assumptions that significantly reduce the number of learnable parameters:

- **Local Connectivity**: Nearby pixels in an image are more likely to share relevant features than distant pixels. Thus, only pixels close to each other need to be connected.
- **Parameter Sharing and Location Invariance**: Features such as edges, textures, and patterns are likely to appear in various locations in an image. Applying the same set of weights to different parts of the image drastically reduces the number of parameters.

Following these assumptions, we replace the matrix multiplication used in MLPs with convolution operations. The convolution operation for multiple channels can be defined as:

$$S(l,i,j) = \sum_{k=1}^{K} (\mathbf{I}(k,i,j) * \mathbf{K}(k,l,i,j)) = \sum_{k=1}^{K} \sum_{m=-M}^{M} \sum_{n=-N}^{N} \mathbf{I}(k,i+m,j+n)\mathbf{K}(k,l,m,n) + b \tag{4.11}$$

where:

- $\mathbf{I} \in [0,1]^{3\times H\times W}$ represents the input image.
- $\mathbf{K} \in \mathbb{R}^{L\times K\times 2M\times 2N}$ is a learnable kernel filter applied across the entire image.
- b is the bias term.
- H and W are the height and width of the image, respectively.
- K represents the number of input channels, and L represents the number of output channels.

This operation is similar to applying an edge filter to an image, but here the filter is learnable. The kernel $\mathbf{K}$ can be thought of as a set of 3D filters that are applied to the entire image, with each filter creating a new image channel. After the convolution operation, a nonlinear activation function (typically ReLU) is applied to the output:

$$F(l,i,j) = \sigma\left(\sum_{k=1}^{K} (\mathbf{I}(k,i,j) * \mathbf{K}(k,l,i,j))\right) \tag{4.12}$$

To illustrate why convolutional layers are much more efficient than MLPs for image processing, consider an input image of height H, width W, and 3 RGB channels. If we use an MLP to create a hidden layer of the same size as the input, we would need $H^2 \times W^2 \times 3$ learnable parameters. On the other hand, a convolutional layer would need $K^2 \times 3^2$ learnable parameters, where K is the width of a square filter, and the factor of 3 accounts for the input and output channels.

For an image with $H = W = 21$ and using a kernel size of $K = 3$, the MLP would require 1,750,329 learnable parameters, while the convolutional layer would only need 81. This significant reduction in parameters highlights the efficiency of convolutional layers in handling image data.

Architecture of CNN Models

Modern CNNs frequently follow the ResNet architecture as seen in Fig. 4.1, which is widely recognized for its effectiveness in various image recognition tasks [28]. A typical CNN architecture comprises the following structures:

- **Backbone**: The backbone of a CNN consists of the primary network blocks responsible for feature extraction. It includes convolutional blocks and downsampling layers. The backbone can be described as follows:
  - **CNN Block**: Each CNN block consists of convolutional layers followed by batch normalization and ReLU activation functions. The structure typically follows this pattern:

    input $\rightarrow$ [Conv $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Conv $\rightarrow$ BatchNorm] $\rightarrow$ input+ $\rightarrow$ ReLU

    - **BatchNorm**: Batch normalization normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation, typically bringing the values close to the range of -1 to 1. This stabilizes the learning process and acts as a regularization layer [33].
    - **Residual Connection**: The input is added to the output at the end of the CNN block, known as a residual connection. This facilitates gradient flow and information retention, addressing the vanishing gradient problem [28].
  - **Downsampling Layer**: After a certain number of convolutional blocks, a downsampling layer is added to reduce the height (H) and width (W) of the feature maps. Common downsampling methods include:
    - **Max Pooling**: Reduces the spatial dimensions by selecting the maximum value within each rectangular region of the feature map, highlighting prominent features and providing spatial invariance.
    - **Average Pooling**: Downscales the input by computing the average value within each rectangular region, resulting in smoother and more generalized feature maps. In classification models, the last layer often uses average pooling to reduce H and W to 1.
    - **Strided Convolution**: Performs the convolution operation with a stride greater than one, reducing spatial dimensions by skipping positions, thereby combining feature extraction and downsampling in a single step.

- **Head**: The head of the network is responsible for performing the downstream task. For classification tasks, this typically involves a fully connected network that processes the extracted features to produce the final classification output.

CNN Models for Sea Ice Mapping

In the early stages of using CNNs for sea ice mapping, the architectures employed were quite similar to the ResNet architecture shown in Fig. 4.1. This architecture is primarily designed for image classification or regression rather than segmentation. To adapt it for sea ice mapping, SAR images (HH and HV) were divided into patches of approximately 50 x 50 pixels, which were then processed by the network for classification or regression tasks. For example, Wang *et al.* [68] used a CNN to estimate SIC from SAR images, while Boulze *et al.* [2] used a CNN to predict SOD.

As the field progressed, architectures specifically designed for segmentation began to gain prominence, with one of the most notable being the U-Net architecture [56], originally developed for biomedical segmentation. This architecture, shown in Fig. 4.2, became popular for sea ice mapping applications. For example, [17] utilized a U-Net inspired architecture to predict SIC, incorporating wind data from the European Centre for Medium-Range Weather Forecasts (ECMWF) as a third channel. Similarly, Nagi *et al.* [46] used a U-Net network for floe segmentation, Stokholm *et al.* [67] applied a U-Net network for SIC prediction, and Ren *et al.* Ren *et al.* [53] implemented a dual attention U-Net for classifying open water and sea ice. Furthermore, Chen *et al.* [7] employed a U-Net architecture to win the AutoIce Challenge in 2023.

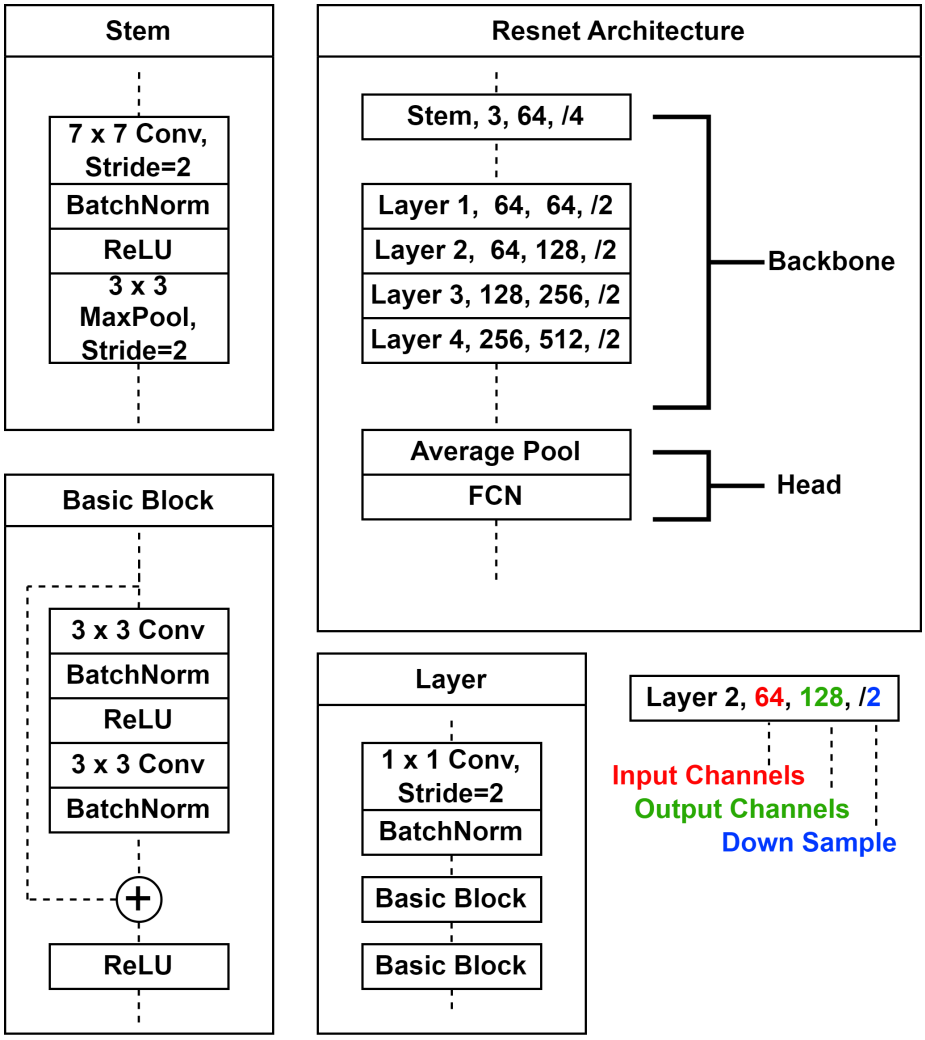


Figure 4.1: Small ResNet architecture.

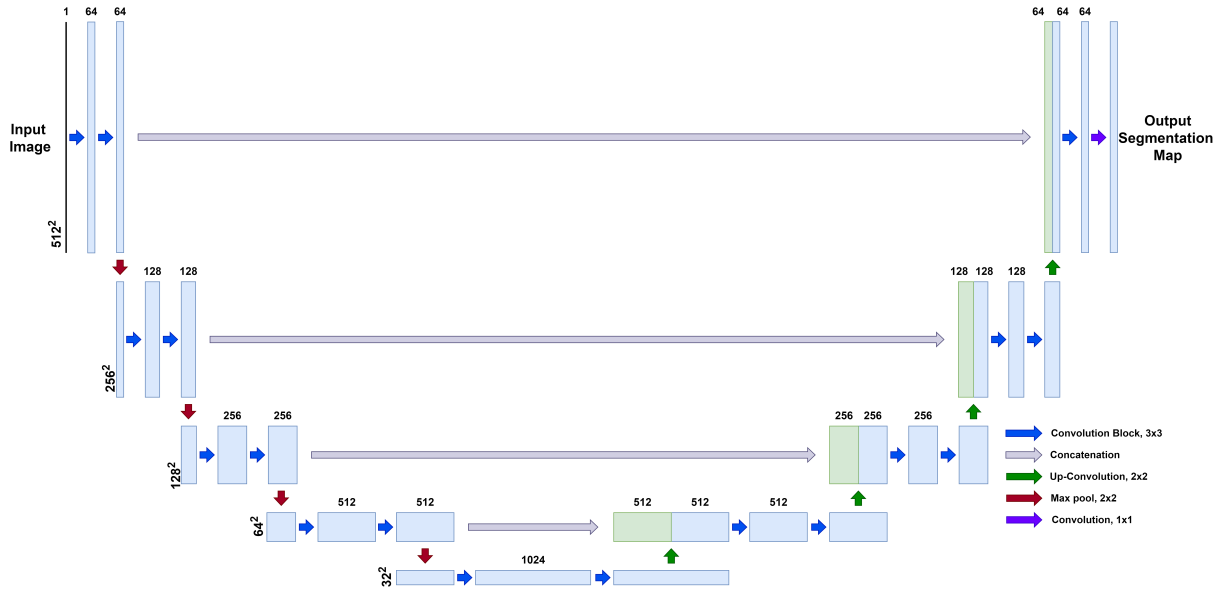


Figure 4.2: U-net architecture overview: Blue boxes signify multi-channel feature maps, with channel numbers labeled above and dimensions noted at the lower left. Green boxes reflect copied feature maps from the blue. Arrows illustrate the operations connecting to the next feature map.

4.2 AI4Arctic Sea Ice Challenge Dataset

The AI4Arctic Sea Ice Challenge Dataset, unveiled in 2022, is a comprehensive dataset comprising 533 observations, distributed between the training dataset (513) and the held-out testing dataset (20). The dataset encompasses two distinct versions: a raw version and a ready-to-train version. For the purposes of this study, we utilized the ready-to-train version, which is pre-processed and optimized for DL model implementations, as detailed in the data manual [4]. Each file in the dataset encapsulates a diverse array of multi-source data, as outlined in Table 4.1.

The dataset incorporates Sentinel-1 Extra Wide Swath (EW) SAR scenes, offering data in both HH and HV polarizations. Notably, these SAR scenes have undergone noise correction using the NERSC algorithm [37]. To balance computational efficiency and high-resolution mapping, a 2×2 kernel has been applied to downsample the pixel spacing to 80 m (∼5000×5000 pixels) in the ready-to-train version.

In tandem with SAR data, the dataset includes corresponding AMSR2 brightness temperatures, resampled to align with the Sentinel-1 geometry with a 2 km grid. Simultaneously, ERA5 weather parameters are retrieved and resampled to align with the Sentinel-1 geometry. Finally, supplementary ancillary data, detailed at the bottom of Table 4.1, is added to further enrich the dataset. Information about the spatial resolutions of those data is given in Table 4.1.

The data is then normalized to fall within the 1 to -1 range, achieved by scaling to the minimum and maximum values for each variable, in order to facilitate the model learning. Data collection spans the period from January 2018 to December 2021, encompassing the Arctic regions surrounding Canada and Greenland. While ERA5, and AMSR2 products are available daily, the 533 observations provided are not daily.

Pixel-based labels are vital for model training and evaluation. In the ready-to-train version, these labels are derived from polygon-based egg codes found in ice charts. Specifically, the SIC attributed to the egg code of each polygon is assigned to all pixels within the polygon as ground truths. Beyond total SIC, each polygon contributes partial SICs, each associated with a specific SOD and FLOE. The SOD and FLOE labels for pixels within a polygon are determined by the dominant ice type and its corresponding FLOE, defined as the ice type with at least 65% in the partial concentration. Polygons lacking a dominant ice type are excluded from both training and evaluation processes.

The distribution of classes for SIC, SOD, and FLOE in training and testing are shown in Fig. 4.3. The distribution is heavily unbalance with most of the pixels belonging to open water, and some of the classes having less than 1% of total pixels.

Table 4.1: The features (data inputs) available in the dataset, which are all utilized for investigation in this research.

Data Input type	**Data input (feature) Name**	**Spatial Resolution**	**Total number of Channels**
SAR variables	HH	80 x 80 m	3
	HV		
	Incidence angle (IA)		
Passive microwave (AMSR2 brightness temperature) data in horizontal and vertical polarizations	AMSR2 6.9 GHz	2 x 2 km	14
	AMSR2 7.3 GHz		
	AMSR2 10.7 GHz		
	AMSR2 18.7 GHz		
	AMSR2 23.8 GHz		
	AMSR2 36.5 GHz		
	AMSR2 89.0 GHz		
ERA5 weather data	Eastward component of 10 m wind speed (u10)		6
	Northward component of 10m wind speed (v10)		
	Temperature 2m above the surface (T2M)		
	Skin temperature (SKT)		
	Total water column water vapor (TCWV)		
	Total column cloud liquid water (TCLW)		
Ancillary data	Distance-to-land map	80 x 80 m	4
	Time information: acquisition month	N/A	
	Location information: SAR latitude grid	~ 19 x 19 km	
	Location information: SAR longitude grid		

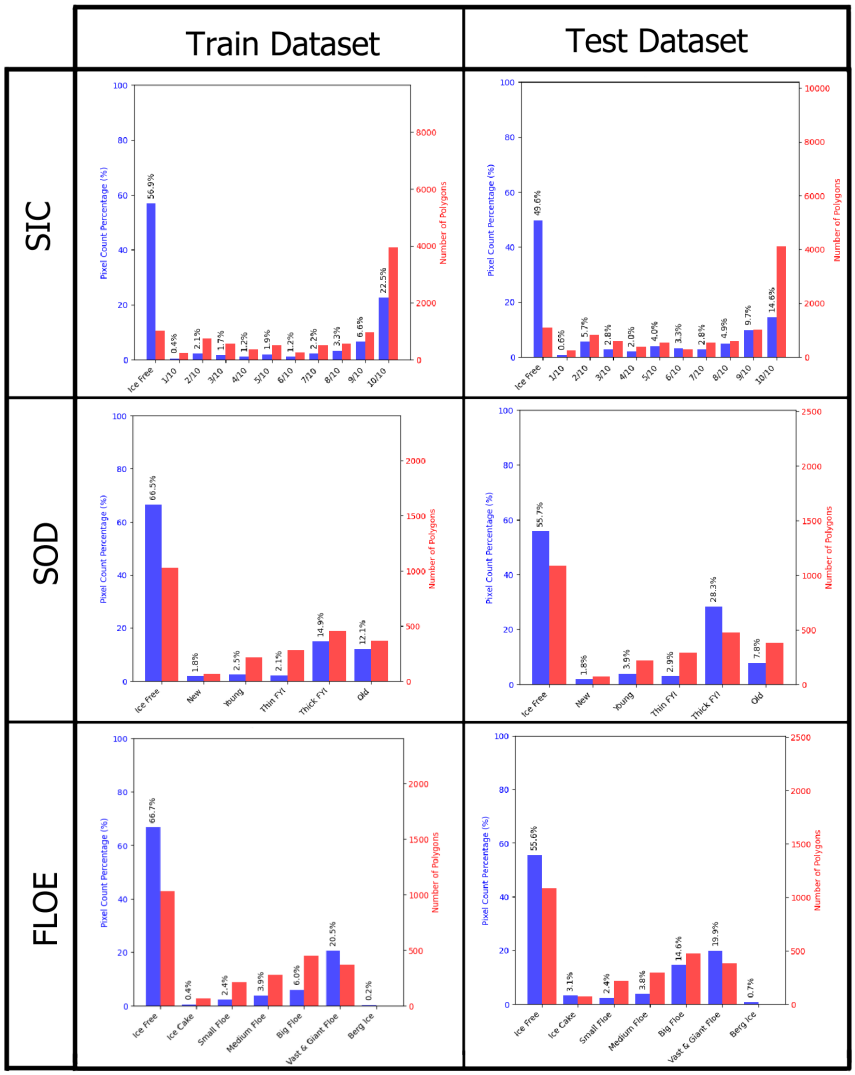


Figure 4.3: The class distribution between the training and testing datasets for SIC, SOD, and FLOE is analyzed. The pixel count percentage indicates the proportion of pixels belonging to a specific class, while the number of polygons represents the count of polygons assigned to that class. For SOD and FLOE, the specified class must be the dominant one.

4.3 Methodology

4.3.1 Data input processing

Utilizing the diverse features as inputs for the CNN-based sea ice mapping model involves a structured three-step process, as illustrated in Fig. 4.4. For our network we utilized a similar U-Net architecture as the winning architecture in the AutoIce Challenge in 2023 [7]. Building on insights from this prior work, we emphasize the pivotal role of downsampling SAR data to significantly enhance mapping accuracy by facilitating the model in capturing contextual information. In the first step, the extensive $\sim 5,000 \times 5,000$ pixel SAR images, alongside the distance-to-land map and corresponding pixel-based label maps, undergo a downsampling by a factor of 10. This results in a more manageable size of approximately $\sim 500 \times 500$ pixels with a pixel spacing of around 800m. Subsequently, the data in subgridded variables (comprising all other variables specified in Table 4.1) are then upsampled to match the resolution of the SAR/label map. This strategic alignment enables the concatenation of all features into a cohesive set serving as data inputs for the model. Finally, to ensure uniformity in data input size during model training, random cropping is applied to the data inputs using a patch size of 256×256. Meaning, when training from each file a random crop of 256 x 256 is extracted. A meticulous grid search, encompassing the selection of downsampling rate and patch size, has been conducted. The values outlined above have been identified as optimal, yielding the most robust and effective model performance.

4.3.2 U-Net-based Sea Ice Mapping Model

Building on our successful participation in the AutoIce sea ice mapping challenge [66, 7], we continue to leverage the effectiveness of the U-Net architecture [56] for our sea ice mapping model. As depicted in Fig. 4.4 and detailed in Table 4.2, the adopted U-Net architecture comprises four encoder and decoder blocks seamlessly connected by a bridge layer. Within each convolution block, two convolutional layers are employed, with the number of filters specified by the digits in Fig. 4.4. The final up-sample layer's output is intricately linked to three distinct 1×1 convolution layers. This connection serves the crucial purpose of generating mapping results for the three sea ice parameters. As mentioned in our previous publication [7], we explored alternative configurations, such as increasing the number of blocks or filters. However, none of these approaches resulted in improved performance.

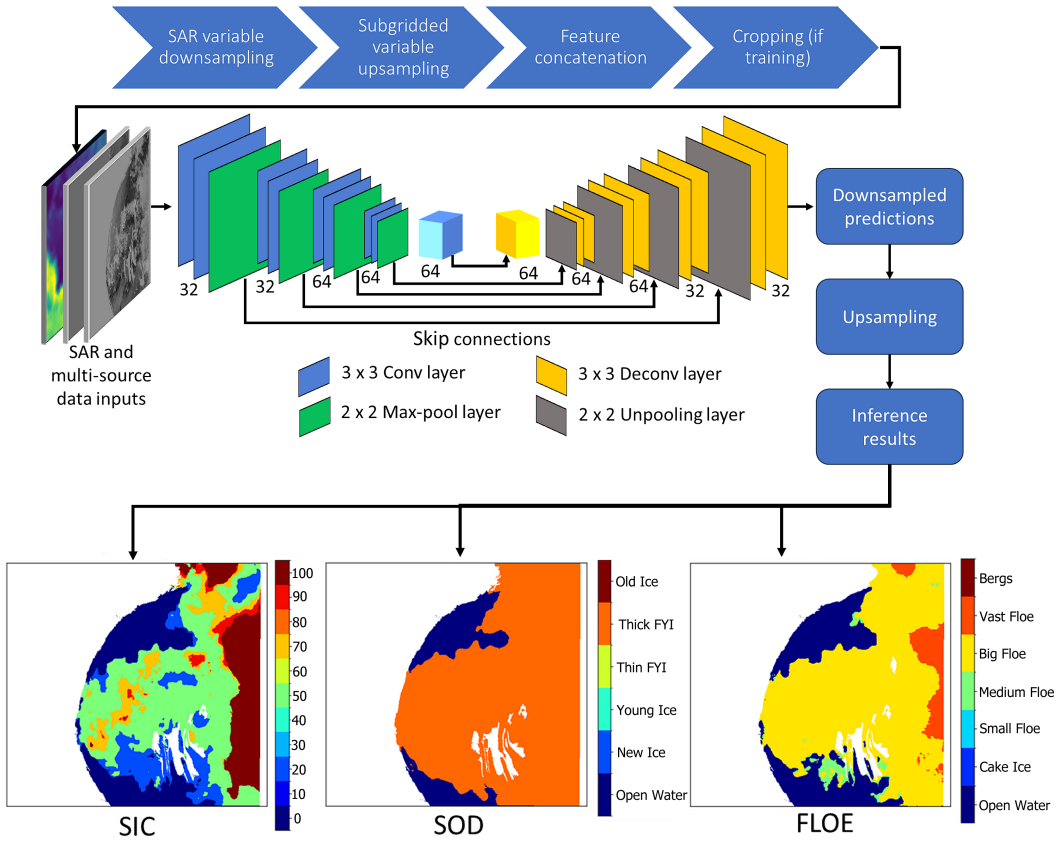


Figure 4.4: A diagram illustrating the U-Net-based sea ice mapping pipeline implemented in this research.

Table 4.2: Layer specifications of the U-Net-based model using all available data inputs.

	Name	Input feature maps	Output feature maps
Encoder	Convolution block 1	27×256×256	32×256×256
	Max-pooling 1	32×256×256	32×128×128
	Convolution block 2	32×128×128	32×128×128
	Max-pooling 2	32×128×128	32×64×64
	Convolution block 3	32×64×64	64×64×64
	Max-pooling 3	64×64×64	64×32×32
	Convolution block 4	64×32×32	64×32×32
	Max-pooling 4	64×32×32	64×16×16
	Bridge	64×16×16	64×16×16
Decoder	Up-sample 1	64×16×16	64×32×32
	Convolution block 5	128×32×32	64×32×32
	Up-sample 2	64×32×32	64×64×64
	Convolution block 6	128×64×64	64×64×64
	Up-sample 3	64×64×64	64×128×128
	Convolution block 7	96×128×128	32×128×128
	Up-sample 4	32×128×128	32×256×256
	Convolution block 8	64×256×256	32×256×256
Outputs	1×1 convolution+ Softmax (SIC)	32×256×256	12×256×256
	1×1 convolution+ Softmax (SOD)	32×256×256	7×256×256
	1×1 convolution+ Softmax (FLOE)	32×256×256	8×256×256

Table 4.3: Specifications of model training.

Learning rate	0.0001
Weight decay	0.001
momentum	0.001
Optimizer	SGD
Batch size	16
Number of iterations per epoch	500
Total epoch	60
Patch size	256
Down Sample	10x
Loss functions	Cross entropy

4.3.3 Training Specifications

Model training specifications are elucidated in Table 4.3. Optimal hyperparameters, crucial for achieving peak performance, are diligently determined through an exhaustive grid search. To guarantee ample exposure per data sample during the training phase, each epoch encompasses 500 iterations. Within each iteration, a batch of patches is randomly extracted from the training scenes. During these iterations, the cross-entropy losses computed for the three parameters are aggregated, collectively constituting the total loss. This strategic amalgamation of losses ensures a comprehensive optimization approach, considering all three sea ice parameters.

4.3.4 Model Evaluation Metrics

To assess the model's performance, we employ the metrics established in the AutoIce Challenge [66]. For SIC, accuracy is quantified using the R2 score. This score captures the inter-class relationships among different SIC levels, acknowledging that, for instance, 10% SIC is closer to 20% than to 30%. Mathematically, the R2 score is defined as:

$$\text{R2 Score} = 1 - \frac{\sum_{i=1}^{N_{\text{pixel}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_{\text{pixel}}} (y_i - \bar{y}_i)^2}, \tag{4.13}$$

where y_i and $\hat{y}_i$ represent the ground truth and estimated SIC of the i^{th} pixel, respectively. $\bar{y}$ denotes the mean value of all y_i among all valid pixels in all SAR scenes of the validation/testing set, and N_{pixel} is the total number of such pixels.

For SOD and FLOE, the F1 score is chosen as the evaluation metric:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{4.14}$$

with Precision and Recall defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}, \tag{4.15}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}, \tag{4.16}$$

where pixel true positives, false positives, and false negatives are aggregated across all valid pixels in the validation/testing set.

To offer a consolidated assessment, a combined score is introduced in [66], defined as:

$$\text{Combined Score} = \frac{2}{5} \times \text{SIC} + \frac{2}{5} \times \text{SOD} \ + \frac{1}{5} \times \text{FLOE}. \tag{4.17}$$

The assigned weights to the three parameters ensure that the combined score does not exceed 100%. Given the operational significance of SIC and SOD over FLOE, a relatively smaller weight is assigned to the latter.

4.3.5 Cross-Validation and Hypothesis Testing

To scrutinize the influence of distinct data types on model efficacy, we embarked on a dual-faceted investigative approach, encompassing both removal and addition studies. In the removal study, we systematically eliminated specific features, such as AMSR2 channels, from the comprehensive "Full Model" (which incorporates all data types) to assess the consequential impact on model performance. Conversely, the addition study is initiated with a foundational model, exclusively employing HH and HV SAR polarization features, to which we added other data types to evaluate performance enhancements.

Ensuring the veracity of our findings, each experiment was subjected to Monte Carlo cross-validation. This entailed the random selection of 20 scenes for validation, with the remaining 493 scenes allocated for model training. Post-training, model performance was

then evaluated. This procedure was replicated over 30 times (30 runs), with results subsequently averaged to mitigate variance. Note that the test scenes do not change during this procedure.

To ascertain the statistical significance of removing/adding specific feature types on model scores, z-tests [23] were conducted between base models (models in the first rows of Tables 4.6 and 4.7) and other models. Assuming that the given sample mean score (including combined scores and individual parameter scores) is approximately normally distributed (due to the central limit theorem) and the t-test approximately equal to the z-test when sample size is larger than 30, the following hypotheses were formulated:

- **Null Hypothesis (H_0):** There is no significant difference between the two scores.
- **Alternative Hypothesis (H_1):** A certain score in the base model is significantly larger than that score in another model (or vice versa). A one-tailed test was chosen for this analysis.

With a significance level α of 0.05, Z-statistics were calculated using the formula:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{4.18}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample mean scores of the cross-validation runs, s_1 and s_2 are the sample standard deviations, and n_1 and n_2 are the number of cross-validation runs or samples for each model. For a one-tailed test, the critical value (1.645) was obtained from the standard normal distribution table. If the calculated Z-statistic falls in the critical region, the null hypothesis is rejected.

4.4 Results and Discussion

4.4.1 Performance of Full Model

This section elaborates on the performance of the "Full Model", which integrates all available data types for sea ice mapping. According to Table 4.4, the model shows robust and anticipated performance across both the cross-validation and test datasets, with the test

Table 4.4: Overall performance of the "Full Model".

Model	Cross Validation Dataset				Test Dataset				Number of runs
	Combined Score	SIC	SOD	FLOE	Combined Score	SIC	SOD	FLOE	
Full Model	91.13	92.21	92.01	87.24	83.93	87.03	86.26	73.09	34

dataset proving more challenging as indicated by its lower average scores. The model exhibits commendable accuracy in estimating SIC and SOD, albeit encountering difficulties in accurately predicting floe sizes, which tend to score lower.

Fig. 4.5 shows the confusion matrices for validation and testing and each metric (SOD, FLOE and SIC). Overall, the confusion matrices show that the model particularly has a lower performance on intermediate classes. In terms of SIC, it often misclassifies categories that are neither 0% nor 100%, yet tends to predict values close to the actual class, reflecting a generally satisfactory performance. This is likely because predicting SIC is a regression problem rather than a classification problem. For SOD, the model encountered significant difficulties with new ice, young ice, and thin FYI. In the domain of floe size, it struggles with most categories, with ice cake and icebergs presenting the most substantial challenges. The observed challenges are likely linked to the limited availability of data for certain ice types. As shown in Fig. 4.3 new ice, young ice, thin FYI ice, ice cake, and iceberg are relatively rare occurrences in the dataset, which may contribute to the model's underperformance in these categories. Furthermore, 10X downsampling likely reduces the ability of the model to see smaller ice floes like cake ice. As pixel spacing changes from 80 m to 800 m.

Finally, Table 4.5 presents the model's monthly performance. The results were derived from over 30 Monte Carlo cross-validation runs. On each run, the validation scores per scene were calculated. Afterward, the scores from each scene across all runs were averaged per month, leading to a score per month. No testing scores could be used since for testing the same 20 scenes are used in every single run. The analysis reveals that the model performs higher during transitional periods of ice melting and freezing, particularly in March and September when ice extents are at their maximum and minimum, respectively. Conversely, the model's performance dips in July, January, and December, coinciding with middle melting and freezing periods. A closer look at specific metrics indicates a significant drop in SIC accuracy only during the months of May, June, and July, while SOD and FLOE size predictions are notably less accurate in December, January, and July, indicating that SIC prediction is only notably worse during melting.

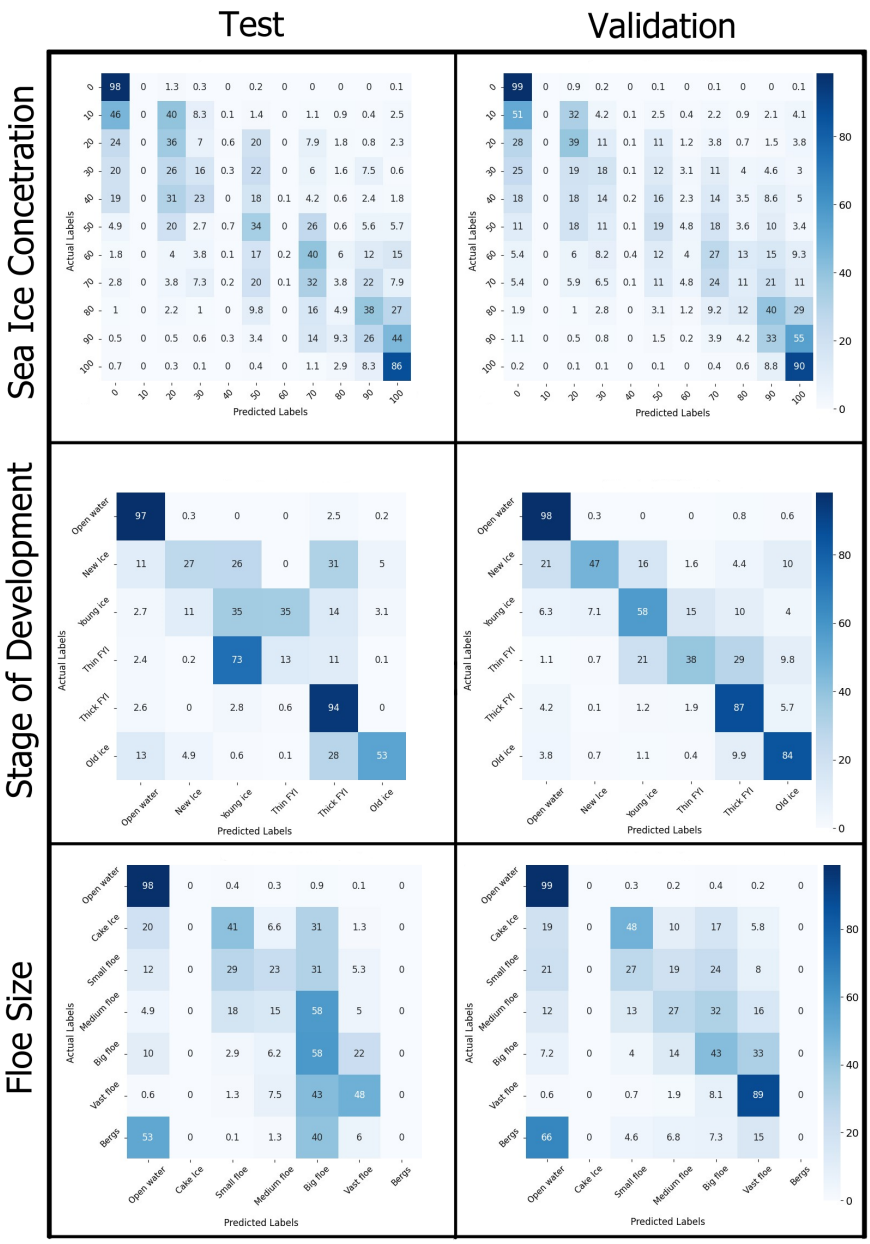


Figure 4.5: Confusion matrices for SIC, SOD, and FLOE on the test and validation dataset.

Table 4.5: Monthly performance for the "Full Model". **Bold** highlights lowest three scores.

Months	**Combine Score**	**Combine Score Std**	**SIC**	**SIC Std**	**SOD**	**SOD Std**	**FLOE**	**FLOE Std**	**Number of runs**
January	**86.95**	10.32	94.58	4.68	**82.98**	18.34	**79.60**	23.69	45
February	90.51	10.28	96.30	3.52	86.82	17.46	86.33	16.40	44
March	91.63	6.12	96.26	2.43	88.53	11.92	88.56	9.22	32
April	91.17	6.15	94.24	4.35	91.60	9.33	84.19	13.64	47
May	90.75	6.68	**91.84**	6.05	94.65	7.92	80.79	23.01	44
June	88.90	7.35	**88.42**	9.58	93.84	6.21	80.00	19.11	37
July	**85.12**	17.65	**86.89**	27.84	**86.95**	13.36	**77.95**	22.29	61
August	90.66	9.34	92.26	8.13	91.25	11.27	86.25	14.70	97
September	93.19	7.81	95.01	6.17	93.44	8.62	89.04	17.84	82
October	91.54	10.29	93.64	8.03	89.76	14.94	90.89	12.03	67
November	88.68	12.34	93.69	8.74	88.30	15.22	**79.44**	26.25	45
December	**85.98**	11.29	94.45	4.46	**79.11**	22.76	82.79	15.79	53

Table 4.6: Summary of numerical results from ablation studies on different feature types. The scores in SIC, SOD, and FLOE correspond to the R2, F1, and F1 scores, respectively. Data inputs that have a statistical significant impact on model performance are in bold.

Model description	Cross validation (%)				Testing (%)			
	Combine Score	SIC	SOD	FLOE	Combined Score	SIC	SOD	FLOE
Full model (No removal)	**91.13**	**92.21**	**92.01**	**87.24**	**83.93**	**87.03**	**86.26**	**73.09**
Remove IA	91.81	92.41	92.71	88.80	84.06	87.33	86.52	72.60
Remove distance map	92.13	93.07	93.02	88.49	83.46	86.89	85.52	72.50
Remove AMSR2	**90.12 (-1.01)**	92.01	**89.46 (-2.54)**	87.65	**80.41 (-3.52)**	87.12	**77.23 (-9.03)**	73.33
Remove ERA 5	92.15	92.64	93.31	88.82	84.10	87.44	86.02	73.59
Remove IA, distance map, ERA 5	91.60	92.29	92.70	88.02	83.55	86.66	85.84	72.76
Remove Time	91.76	93.17	91.92	88.61	**82.49 (-1.45)**	87.10	**82.81 (-3.46)**	72.62
Remove Location	90.78	92.62	90.98	86.71	**82.21 (-1.72)**	87.08	**81.64 (-4.63)**	73.61
Remove Time and location	**90.21 (-0.92)**	91.95	**89.93 (-2.07)**	87.32	**80.45 (-3.48)**	87.23	**77.22 (-9.04)**	73.37
Remove SAR variables (HH, HV, IA)	**89.91 (-1.22)**	**89.21 (-2.99)**	91.60	87.90	**78.51 (-5.42)**	**78.45 (-8.57)**	**82.76 (-3.50)**	**70.13 (-2.96)**

Table 4.7: Summary of numerical results obtained from the "adding features" experiments. Data inputs that have a statistical significant impact on model performance are in bold.

Model description	Cross validation (%)				Testing (%)			
	Combined Score	SIC	SOD	FLOE	Combined Score	SIC	SOD	FLOE
HH, HV only	88.01	90.75	86.10	86.37	75.63	77.58	76.19	70.60
HH, HV, and ERA 5	88.17	91.05	86.20	86.37	**79.30 (+3.67)**	**83.89 (+6.31)**	**77.58 (+1.39)**	**73.54 (+2.95)**
HH, HV, and distance map	**89.36 (+1.34)**	**91.49 (+0.75)**	**87.75 (+1.65)**	**88.29 (+1.92)**	75.20	76.69	75.90	**70.86 (0.29)**
HH, HV, and IA	**88.79 (+0.78)**	**91.44 (+0.70)**	86.96	87.15	75.81	77.79	76.56	70.37
HH, HV, and AMSR2	**90.32 (+2.31)**	**92.30 (+1.55)**	**90.01 (+3.91)**	87.00	**81.15 (+5.52)**	**87.28 (+9.70)**	**78.43 (+2.24)**	**74.33 (+3.73)**
HH, HV, and Time	**89.78 (+1.76)**	91.00	**90.05 (+3.95)**	86.81	**77.93 (+2.31)**	76.62	**82.81 (+6.63)**	70.80
HH, HV, and Location	**89.03 (+1.02)**	90.49	**88.47 (+2.37)**	87.22	**79.78 (+4.16)**	**82.52 (+4.94)**	**81.26 (+5.07)**	**71.36 (+0.76)**
HH, HV, and Time and Location	**91.54 (+3.54)**	**(+88.95) (+2.58)**	**91.65 (+0.91)**	**92.72 (+6.61)**	**78.76 (+3.13)**	75.85	**84.88 (+8.70)**	**72.31 (+1.71)**

4.4.2 Results Removing and Adding Data Inputs

The results of removing and adding data inputs are shown in Tables 4.6 and 4.7. Based on the z-test results, scores significantly lower than the "Full Model" in Table 4.6 and scores significantly higher than the "HH, HV only" model in Table 4.7 are bolded. The ablation studies show that removing passive microwave (AMSR2) channels or time and location channels significantly impacts model scores in both cross-validation and testing sets. For instance, removing all AMSR2 channels reduces combined scores by 1.01% and 3.52% in cross-validation and testing sets, respectively, primarily due to drops in SOD scores (e.g., a 9.03% drop in the testing set). Similarly, removing time and location information channels reduces model accuracy, especially for SOD (9.04% drop in the testing set). Conversely, removing other features (e.g., ERA5, incidence angle, and distance-to-land map) from the Full model does not show statistical significance in affecting model accuracy. These findings are further validated in "adding" experiments detailed in Table 4.7. For instance, adding AMSR2 channels improves scores for all three parameters (2.31% and 5.52% increase in combined scores for cross-validation and testing sets, respectively). Incorporating time and location features also contributes significantly to model performance, with a 3.54% and 3.13% increase in combined scores for cross-validation and testing sets, respectively. Although adding other types of features might improve model scores, these improvements are not consistent between cross-validation and testing sets. For instance, adding ERA5 features improves the combined testing score by 3.67% but shows limited enhancement in cross-validation sets. Looking at the removing and adding results, they suggest that the

Table 4.8: Numerical results obtained from the models (with SAR HH and HV inputs) adding AMSR2 channels individually. The blue and red colors correspond to positive and negative contributions in accuracy, respectively. All results show statistical significance compare to HH, HV baseline.

Model description	Cross validation (%)				Testing (%)			
	Combined Score	SIC	SOD	FLOE	Combined Score	SIC	SOD	FLOE
HH, HV only	88.01	90.75	86.10	86.37	75.63	77.58	76.19	70.60
HH, HV, and AMSR2 6.9 GHz	**91.15 (+3.14)**	**92.08 (+1.34)**	**90.13 (+4.03)**	**91.35 (+4.98)**	**81.98 (+6.35)**	**87.68 (+10.10)**	**80.53 (+4.34)**	**73.49 (+2.89)**
HH, HV, and AMSR2 7.3 GHz	**90.59 (+2.57)**	**92.84 (+2.10)**	**89.65 (+3.55)**	**87.95 (+1.58)**	**80.89 (+5.27)**	**87.06 (+9.48)**	**79.27 (+3.09)**	**71.8 (+1.20)**
HH, HV, and AMSR2 10.7 GHz	**90.25 (+2.24)**	**93.02 (+2.27)**	**88.81 (+2.71)**	**87.61 (+1.24)**	**81.07 (+5.44)**	**87.12 (+9.54)**	**79.43 (+3.24)**	**72.24 (+1.65)**
HH, HV, and AMSR2 18.7 GHz	**90.15 (+2.13)**	**93.2 (+2.46)**	**88.25 (+2.14)**	**87.84 (+1.47)**	**80.88 (+5.25)**	**87.15 (+9.57)**	**78.61 (+2.42)**	**72.89 (+2.29)**
HH, HV, and AMSR2 23.8 GHz	**90.18 (+2.17)**	**92.8 (+2.05)**	**88.53 (+2.43)**	**88.26 (+1.89)**	**80.44 (+4.81)**	**86.92 (+9.34)**	**77.48 (+1.29)**	**73.39 (+2.79)**
HH, HV, and AMSR2 36.5 GHz	**90.3 (+2.29)**	**92.56 (+1.81)**	**89.06 (+2.96)**	**88.27 (+1.90)**	**81.01 (+5.38)**	**87.84 (+10.27)**	**77.68 (+1.49)**	**73.99 (+3.39)**
HH, HV, and AMSR2 89.0 GHz	**84.51 (-3.50)**	**88.79 (-1.96)**	**81.65 (-4.45)**	**81.66 (-4.71)**	**80.15 (+4.52)**	**85.83 (+8.26)**	**77.57 (+1.39)**	**73.94 (+3.34)**

inclusion of ERA5 data is not necessary if AMSR2 is already included. A detailed analysis for each feature type is provided below.

4.4.3 Passive Microwave (AMSR2) Feature Analysis

An example SAR scene in the test data set with the corresponding labels and mapping results is presented in Fig. 4.6 to demonstrate the effectiveness of AMSR2 inputs in improving the classification accuracy of SOD. Without AMSR2 inputs, the model misclassifies the region with young ice as thick first-year ice (FYI). In contrast, this issue is significantly alleviated with the incorporation of AMSR2 (i.e., the full model). The improvement in SIC and FLOE is less significant in this scene.

To analyze the difference between different AMSR2 frequencies on model performance, models with each individual AMSR2 frequency channel (concatenated with SAR HH and HV channels) are trained and compared with each other, as shown in Table 4.8. The improvements in model performance are close for all frequencies except 89.0 GHz. Particularly, the model with AMSR2 6.9 GHz channels obtains the highest combined and SOD scores. Although the 6.9 GHz channels have relatively low resolution, they are less sensitive to noise and only sensitive to sea surface temperature as shown in Fig. 4.7. This could explain why they achieved the highest improvement in mapping accuracy between differ-

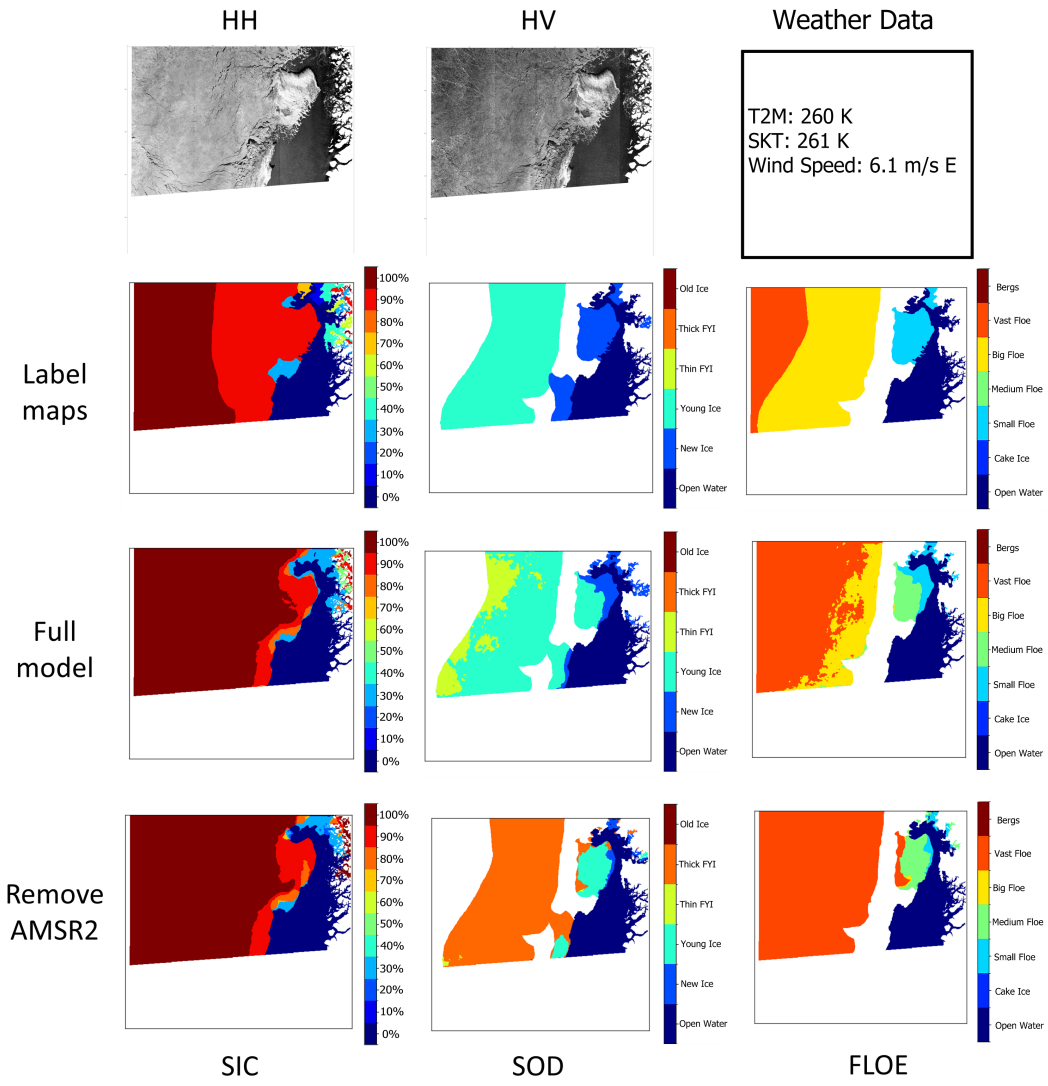


Figure 4.6: An example SAR scene in the testing dataset (file ID name: 20211212T211242_dmi, collected on December 12^{th}, 2021 in Northwest Greenland, center latitude/longitude: $71.64°N, 60.53°W$) with corresponding labels (second row) and mapping results (third and fourth rows). The weather parameters such as T2M shown in this scene corresponds to the average value across the entire scene. The third row corresponds to results from the full model, while the fourth row corresponds to results from the model without AMSR2 inputs "Remove AMSR2 in Table 4.6".

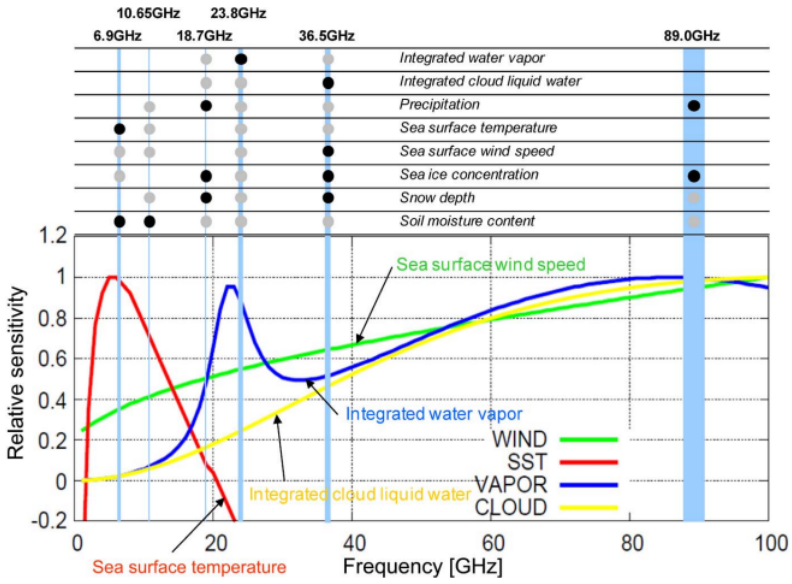


Figure 4.7: Relative sensitivity of AMSR2 for oceanic geophysical parameters as a function of observing frequency [32].

ent frequencies. Conversely, the 89.0 GHz channels are more sensitive to weather-induced noise, causing ambiguities between ice and water.

One may hypothesize that the lower resolution of 6.9 GHz may hinder the model performance by causing poorly delineated edges. However qualitatively, this does not seem to be the case. Fig. 4.8 compares the output of HH + HV + AMSR2 6.9 GHz against HH + HV + AMSR2 89.0 GHz which has a much higher resolution. The scene scores for 6.9 GHz and 89.0 GHz, respectively are SIC: 93.88%, SOD: 91.52%, FLOE: 69.48% and SIC: 93.55%, SOD: 74.11%, FLOE: 69.77%. Both models tend to have boundaries between water and ice that are quite similar to each other. Furthermore, the SIC and FLOE scores are quite similar on both models. However, in this scene for SOD, 6.9 GHz significantly outperforms 89.0 GHz. The main reason being that 89.0 GHz misclassifies entire regions as old ice. It also misclassifies the ice in the fjord as new ice instead of young ice. This indicates that the main difference in performance does not come from edges but rather from misclassification of entire regions.

In addition, in Fig. 4.9, we can see that the channel 89.0 GHz (V) struggles to differentiate between water and ice. These two factors could explain why adding AMSR2 89.0 GHz channels might degrade model performance as seen in Table 4.8.

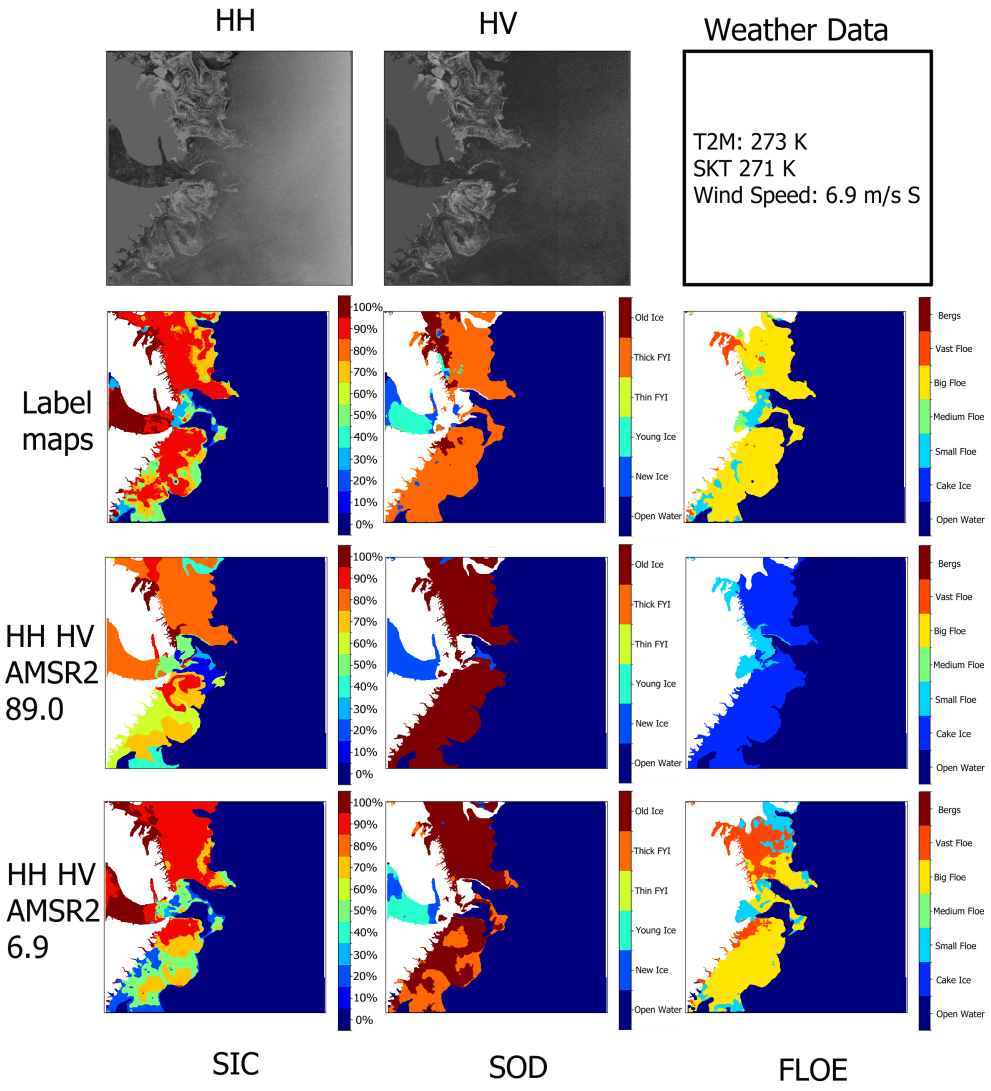


Figure 4.8: An example comparing the mapping results from models with two different feature combinations: HH + HV + AMSR2 89.0 GHz and HH + HV + AMSR2 6.9 GHz. File ID name of the SAR scene: 20201013T080448_dmi, collected on October 10^{th}, 2020 in the east coast of Greenland, center latitude/longitude: $70.10°N, 19.74°W$. The weather parameters like T2M shown in this scene corresponds to the average value across the entire scene.

Compared to the model with only HH and HV as inputs, including all AMSR2 channels improves the SIC, SOD, and FLOE scores by 9.70%, 2.24%, and 3.73% in the testing set as shown in Table 4.7. Adding a single frequency can achieve similar performance. This is due to the high correlation in brightness-temperature patterns across frequencies, which introduces redundancy in feature space.

Furthermore, changes in class-wise accuracy after adding or removing AMSR2 channels are investigated for all three parameters. Fig. 4.10 shows the SOD confusion matrices (in percentage) obtained from (a) the full model, (b) the model without AMSR2 channels, and (c) their percentage difference in each element. Fig. 4.10 (c) demonstrates that removing AMSR2 mostly impacts the classification accuracy of old ice, followed by thick FYI and young ice. Since the other two ice types (new ice and thin FYI) have very few data samples and high ambiguities with neighboring classes, it can be concluded that AMSR2 channels contribute significantly to improving mapping accuracy of major ice types. For SIC, the significant improvement after adding AMSR2 channels is mostly caused by correcting the misclassification of regions with 100% SIC, as shown in Fig. 4.11. This demonstrates that adding AMSR2 inputs can overcome the ambiguity issue in the electromagnetic signature between open water and sea ice caused by certain ice conditions. An example of this can be seen in Fig. 4.13. Where HH-HV model misclassifies a large region of 100% ice as OW. This scene is in June 23^{rd}, well into the melting season. Regarding FLOE, the most significant improvement from adding AMSR2 inputs is seen in the vast floe class, as illustrated in Fig. 4.12. This is also mainly attributed to the reduction of misclassification of ice as OW. Thus, it can be summarized that adding a suitable AMSR2 frequency channel as data inputs improves SOD classification accuracy among most ice types significantly while correcting the misclassification between ice and open water in some regions.

Finally, we aimed to assess how the inclusion or exclusion of passive microwave data influences the model's seasonal performance. Tables 4.9 and 4.10 present the model's monthly performance metrics, evaluated on the validation dataset, with AMSR2 data either omitted from the full model or incorporated alongside HH and HV polarizations. Table 4.9 exclusively reports SOD metrics, as significant variations were observed only in SOD, as elaborated in Table 4.6. A statistical decline in model performance for SOD was recorded during March, April, May, and June, when AMSR2 was removed. Highlighting that the majority of performance degradation occurs during the melting season. However, no changes in performance were found in the month of July, which is the worst performing month during the melt season for the Full model.

In contrast, the addition of AMSR2 to HH and HV polarizations resulted in an improved SOD performance, particularly in the months of December, January, May, June, and July, as highlighted in Table 4.10. This improvement is indicative of a more accurate SOD esti-

mation during both the melting and freezing seasons. However, during transitional periods such as October and April, little to no enhancement in SOD performance was observed, pointing to the minimal impact of AMSR2 during these times. For SIC metrics, notable performance improvements were recorded in April, May, and August, emphasizing that the significant benefits of incorporating AMSR2 predominantly occur during the melting season for SIC. This indicates that AMSR2 significantly enhances the model's ability to more accurately differentiate between ice thicknesses during both melting and freezing periods. Additionally, incorporating AMSR2 greatly improves SIC predictions during the melting periods

Our research findings closely aligns with the existing body of knowledge. Malmgren-Hansen et al. [43] showed significant improvements in their CNN model for SIC estimation by incorporating AMSR2 data alongside SAR imagery from Sentinel-1. In particular, their enhanced model exhibited a notable increase in accuracy between open water and sea ice by 3.48%, and a 3.62% boost in the accuracy for SIC predictions with the integration of AMSR2 data. The positive impact on SIC estimation is anticipated, considering the prevalent use of AMSR2 in various algorithms to estimating SIC [14].

Moreover, as delineated in Table 4.8, the inclusion of AMSR2 data enhances the model's performance in estimating the SOD. The literature indicates the potential of passive microwave data in the estimation of the thickness of sea ice. For example, research by Scott et al. [60] demonstrated that data from AMSR2-E spanning frequencies from 6.9 GHz to 36.5 GHz can predict sea ice thickness up to 50cm. Moreover, AMSR2 is currently utilized in identifying thin sea ice ($\leq$30cm) [11]. This understanding elucidates the observed decline in model performance for classifying new ice ($\leq$10cm) and young ice (10-30cm) when AMSR2 data is omitted, as illustrated in Fig. 4.10.

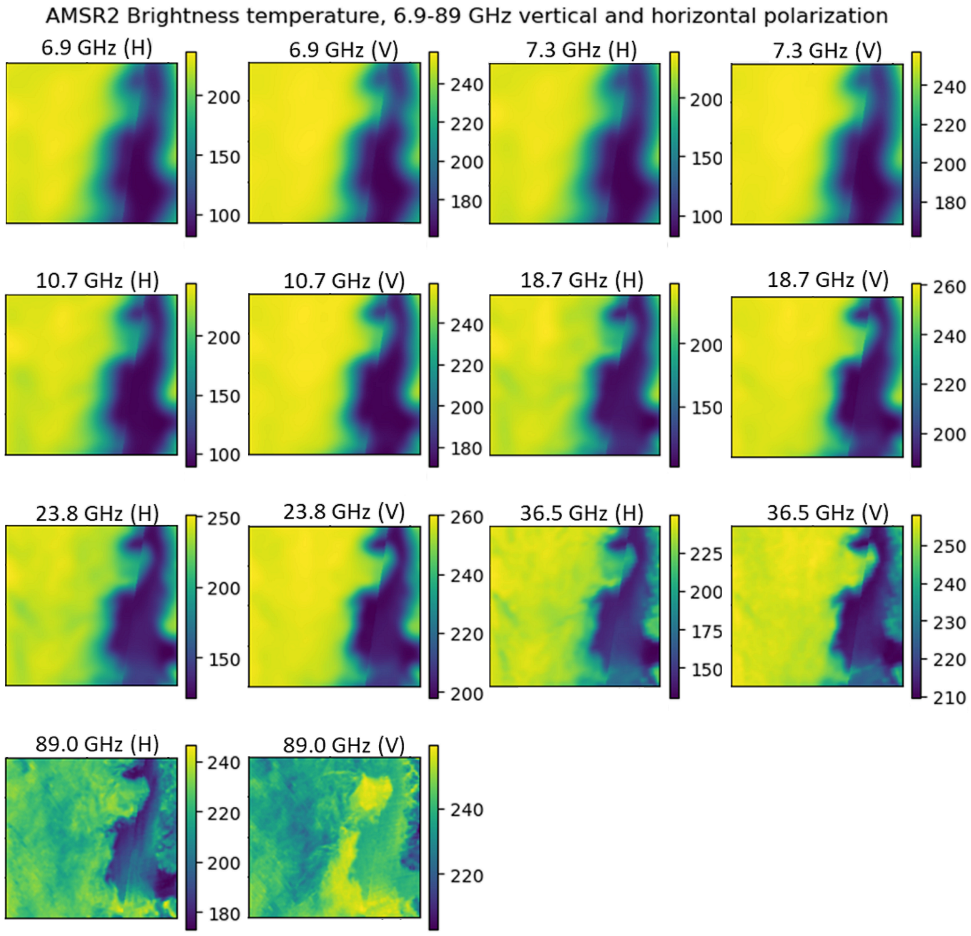


Figure 4.9: The corresponding AMSR2 feature maps of the SAR scene in Fig.4.6 in all available frequencies.

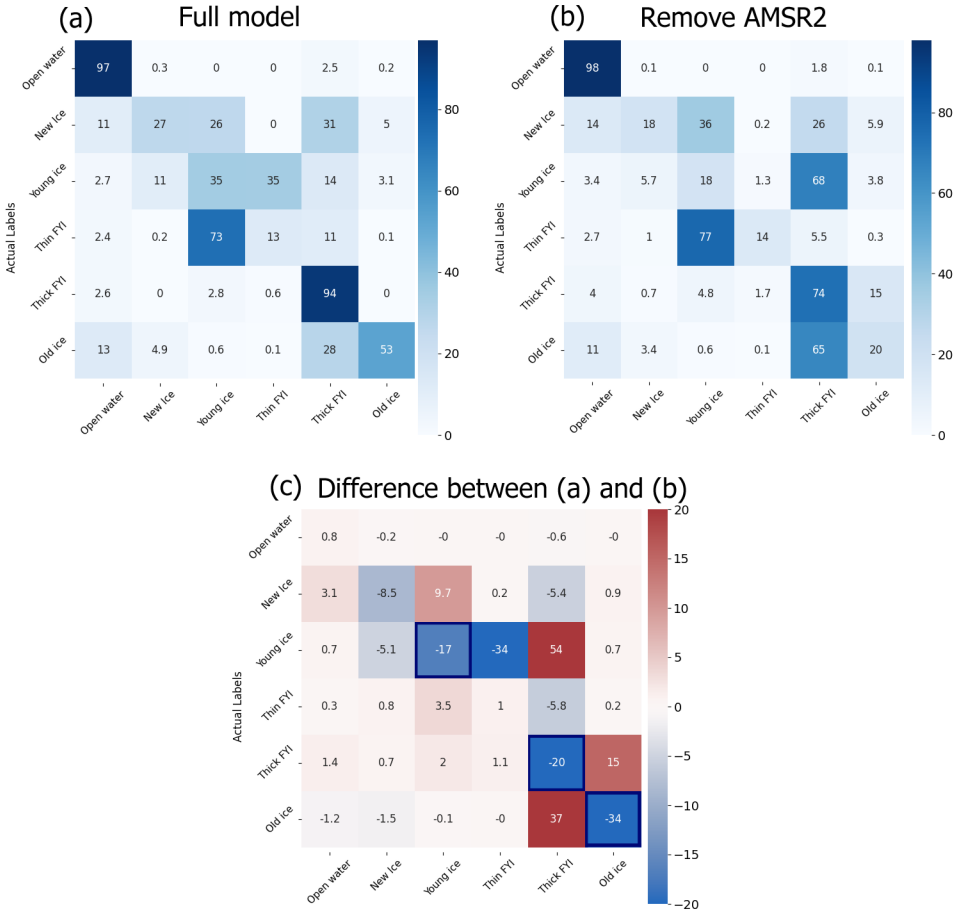


Figure 4.10: (a) Confusion matrix for SOD classification results (testing set) using the 'Full model.' (b) Confusion matrix for SOD classification results (testing set) from the model without AMSR2 features ('Remove AMSR2' in Table 4.6). (c) Difference between (a) and (b).

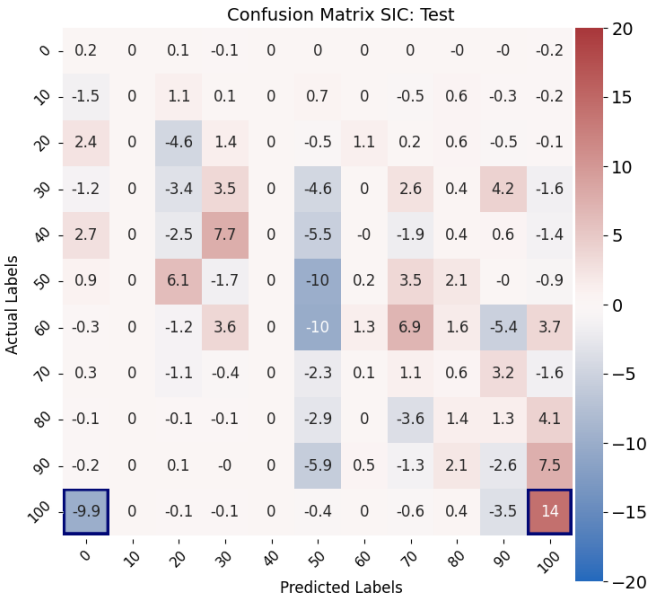


Figure 4.11: The value change in each element of the SIC classification confusion matrix (test dataset after adding AMSR2 features, in comparison with the "HH, HV only" model.

Table 4.9: Comparison of the monthly performance on SOD between "Full Model" and "Remove AMSR2". Statistical significant decreases are shown in bold.

Months	Full Model			Remove AMSR2			Decrease
	SOD	Std	Number of runs	SOD	Std	Number of runs	
January	82.98	18.34	45	80.96	17.32	52	-2.02
February	86.82	17.46	44	85.15	19.74	47	-1.68
March	88.53	11.92	32	**81.94**	21.15	45	**-6.59**
April	91.60	9.33	47	**86.89**	12.72	72	**-4.71**
May	94.65	7.92	44	**87.27**	14.45	53	**-7.38**
June	93.84	6.21	37	**88.82**	9.54	58	**-5.03**
July	86.95	13.36	61	87.12	18.48	67	0.18
August	91.25	11.27	97	90.18	13.39	92	-1.07
September	93.44	8.62	82	91.74	10.47	107	-1.70
October	89.76	14.94	67	92.22	10.12	77	2.46
November	88.30	15.22	45	86.43	19.83	63	-1.87
December	79.11	22.76	53	80.62	23.58	63	1.51

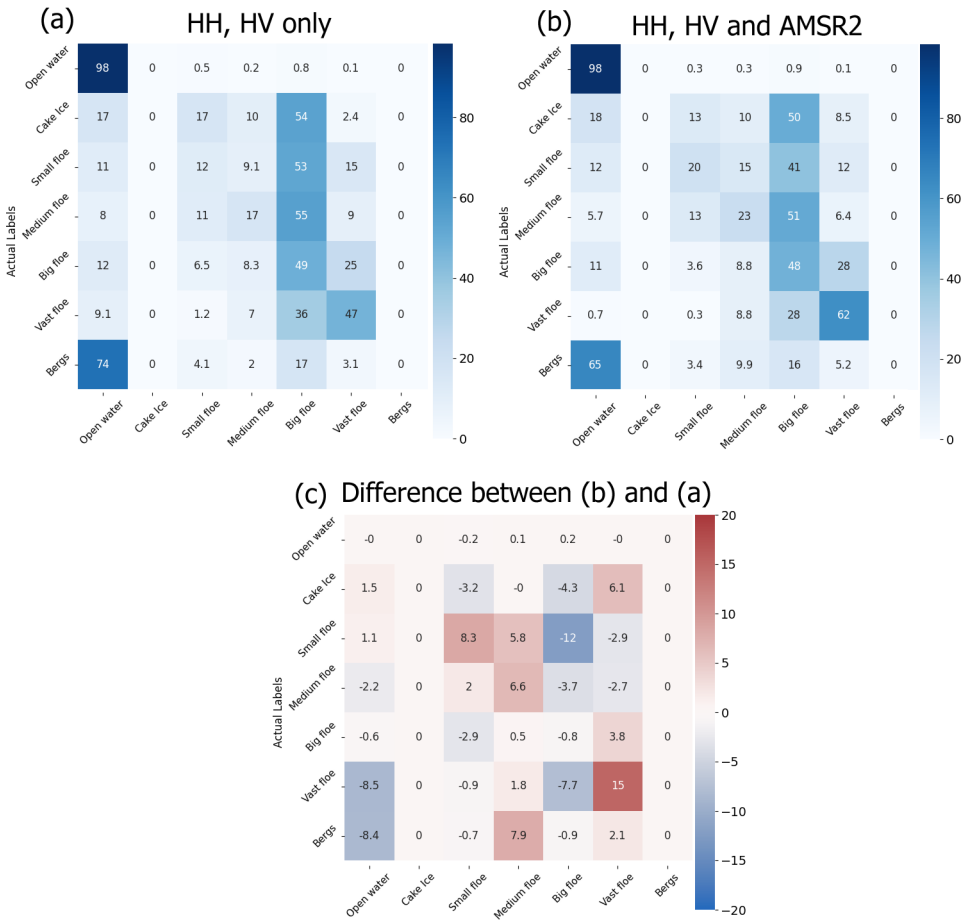


Figure 4.12: (a) The confusion matrix of FLOE classification results (testing set) from the "HH, HV only" model. (b) The confusion matrix of FLOE classification results (testing set) from the model adding AMSR2 features ("HH, HV, and AMSR2" in Table 4.7). (c) Represents (b)-(a).

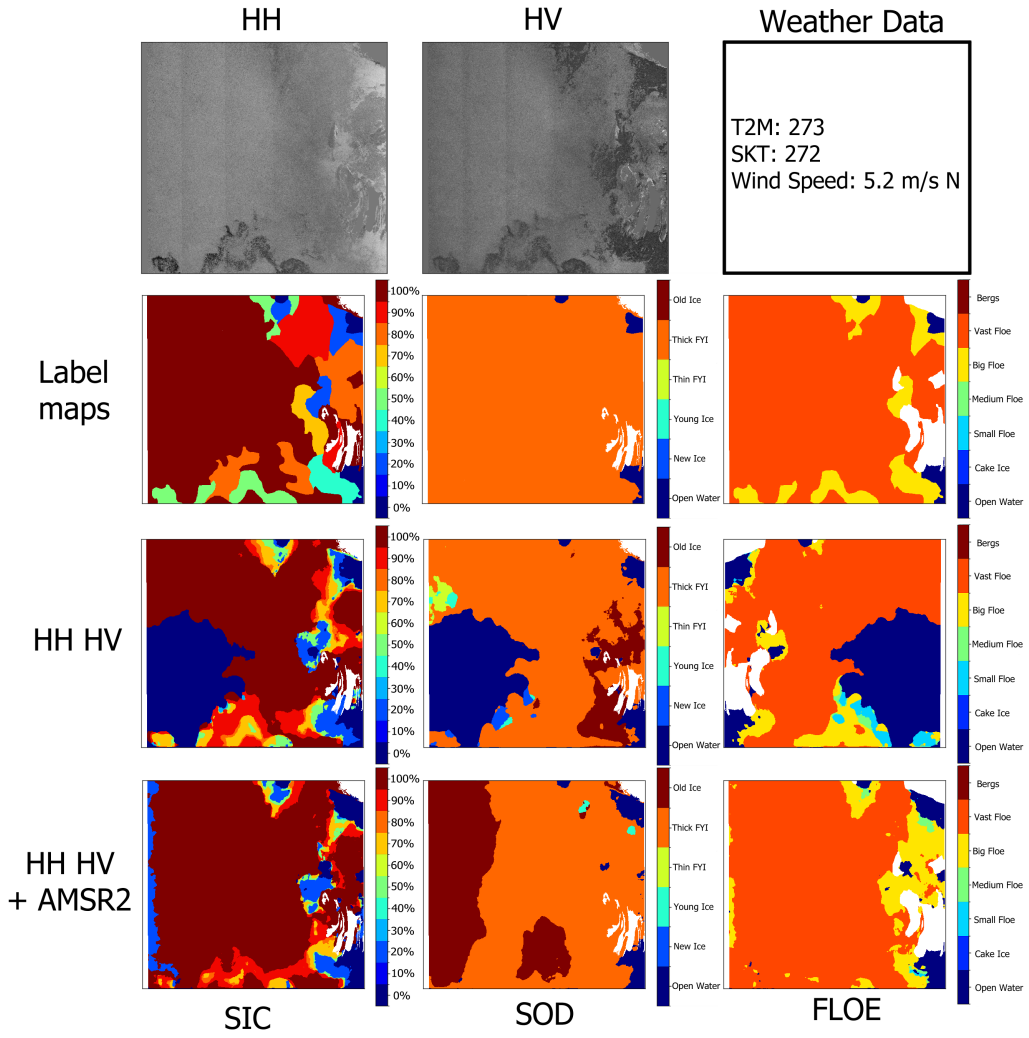


Figure 4.13: An example SAR scene in the testing dataset under melting conditions(file ID name: 20180623T114935_cis, collected on June 23^{rd}, 2018 in Hudson Bay, near the Belcher Islands, center latitude/longitude: $57.32°N, 81.98°W$) with corresponding labels (second row) and mapping results (third and fourth rows). The weather parameters like T2M shown in this scene corresponds to the average value across the entire scene. The third row corresponds to results from the HH HV only model, while the fourth row corresponds to results from HH HV and AMSR2 model (HH, HV, and AMSR2 in Table 4.7).

Table 4.10: Comparison of the monthly performance on SIC and SOD between "HH HV Only" and "HH HV and AMSR2". Statistical significant improvements are shown in bold.

Months	**HH HV Only**		**HH HV and AMSR2**		**Increase**	
	SIC	**SOD**	**SIC**	**SOD**	**SIC**	**SOD**
January	91.13	67.15	93.00	**80.84**	+1.88	**+13.69**
February	97.01	79.59	96.64	83.95	-0.37	+4.37
March	95.03	77.09	94.76	84.04	-0.27	+6.96
April	89.78	87.58	**93.72**	86.40	**+3.95**	-1.18
May	90.54	85.50	**92.56**	**90.26**	**+2.02**	**+4.76**
June	88.18	87.18	91.07	**94.18**	+2.89	**+7.00**
July	84.39	83.40	89.18	**89.43**	+4.78	**+6.02**
August	90.44	86.22	**93.08**	89.27	**+2.64**	+3.05
September	94.60	92.00	94.95	93.74	+0.35	+1.74
October	91.76	88.44	92.95	88.19	+1.19	-0.25
November	92.04	85.54	94.33	86.63	+2.29	1.09
December	93.85	69.15	95.28	**77.27**	+1.44	**+8.12**

4.4.4 Analysis on Spatial-Temporal Features

Examples of comparison between mapping results obtained from the full model and the model removing location and/or time inputs are depicted in Figs. 4.14 and 4.15, which all demonstrate the effectiveness of those features in improving SOD classification accuracy. For instance, in Fig. 4.15, the full model is able to detect the presence of old ice, while the model without location and time information inputs totally misclassified the old ice pixels.

Furthermore, Fig. 4.16 shows the difference in each element of the SOD class-wise confusion matrix between the full model and the model removing time and/or location information. Compared to the full model, removing time inputs reduces the classification accuracy of new ice (-13%), young ice (-18%), and old ice (-13%) significantly (Fig. 4.16(a)). This can be explained by the fact that new and young are mostly present in the freeze-up season and adding time inputs facilitates the model to learn such patterns. Table 4.11 supports this hypothesis by comparing the monthly SOD score between HH HV only and HH HV + Time. The bold text indicates statistical significance. The performance of the model drastically increases in December and January when time is introduced, indicating that much of the performance is gained during the freeze-up season.

In contrast, removing location information inputs mainly reduces the classification accuracy of thick FYI (-13%) and old ice (-17%). The possible reason is the relatively stable spatial distributions of sea ice extent and ice types. For example, old (multi-year) ice mainly occurs in the Canadian archipelago and east coast of Greenland. In consequence, adding location information inputs enables the model to learn about the spatial distributions of different ice types. Removing both time and spatial information further reduces the classification accuracy of most ice types, as demonstrated in Fig. 4.16(c).

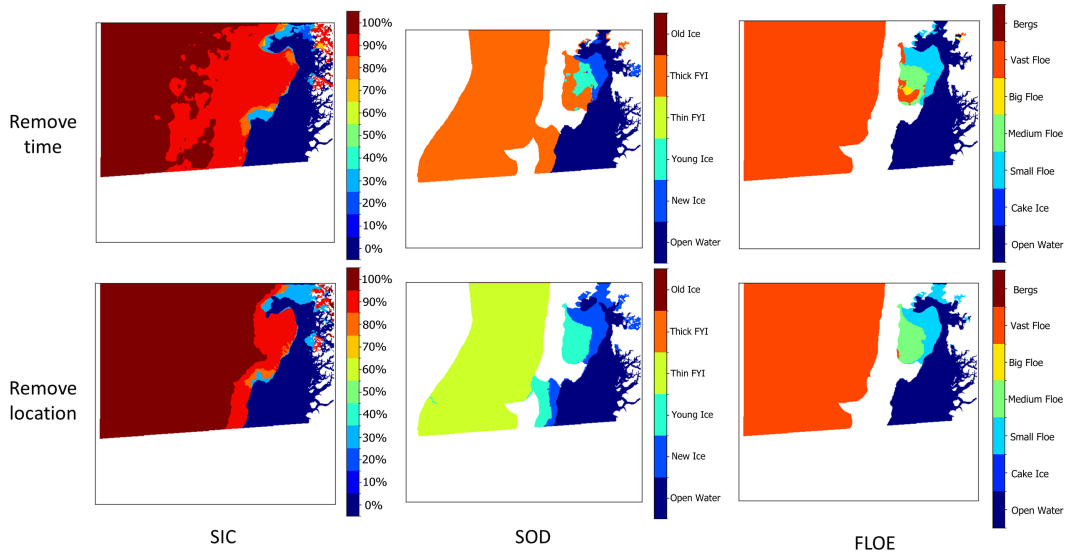


Figure 4.14: Mapping results of the SAR scenes in Fig. 4.6 from more models. First row: the mapping results from the model without time information inputs ("Remove Time" in Table 4.6). Second row: the mapping results from the model without location information inputs ("Remove Location" in Table 4.6).

Table 4.11: Comparison of the monthly performance on SOD between "HH HV Only" and "HH HV + Time". Statistical significant improvements are shown in bold.

Months	HH HV Only			HH HV + Time			Improvement
	SOD	Std	Sample Number of runs	SOD	Std	Number of runs	
January	67.15	28.15	64	**85.12**	14.68	47	**+17.98**
February	79.59	24.83	49	82.12	22.22	43	+2.53
March	77.09	27.24	46	**87.44**	12.46	50	**+10.36**
April	87.59	11.41	61	**90.93**	8.90	40	**+3.35**
May	85.50	16.64	73	**92.66**	8.85	44	**+7.15**
June	87.18	9.30	43	87.67	16.95	35	+0.49
July	83.40	17.65	65	83.24	21.05	57	-0.16
August	86.22	18.72	99	88.73	17.60	93	+2.51
September	92.00	13.66	113	93.64	8.34	85	+1.64
October	88.44	17.30	80	**92.82**	9.99	71	**+4.37**
November	85.54	19.74	66	87.42	18.93	57	+1.88
December	69.15	29.15	61	**83.03**	16.39	38	**+13.88**

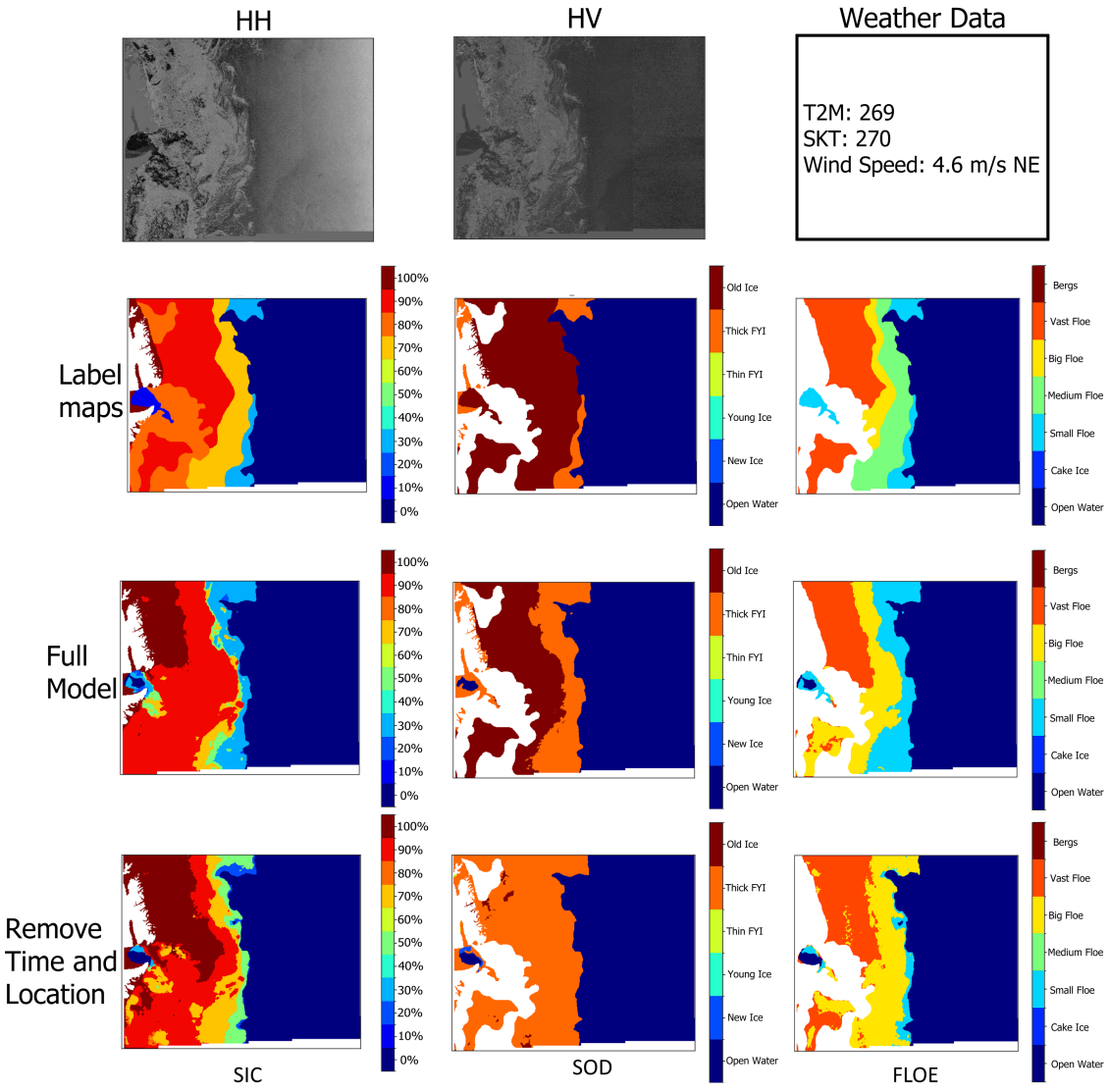


Figure 4.15: An example SAR scene in the testing dataset (file ID name: 20210506T075557_dmi, collected on May 6^{th}, 2021 in Central East Greenland, center latitude/longitude: $68.89°N, 17.49°W$) with corresponding labels (second row) and mapping results (third and fourth rows). The weather parameters like T2M shown in this scene corresponds to the average value across the entire scene. The third row corresponds to results from the full model, while the fourth row corresponds to results from the model without time and location inputs ("Remove time and location" in Table 4.6).

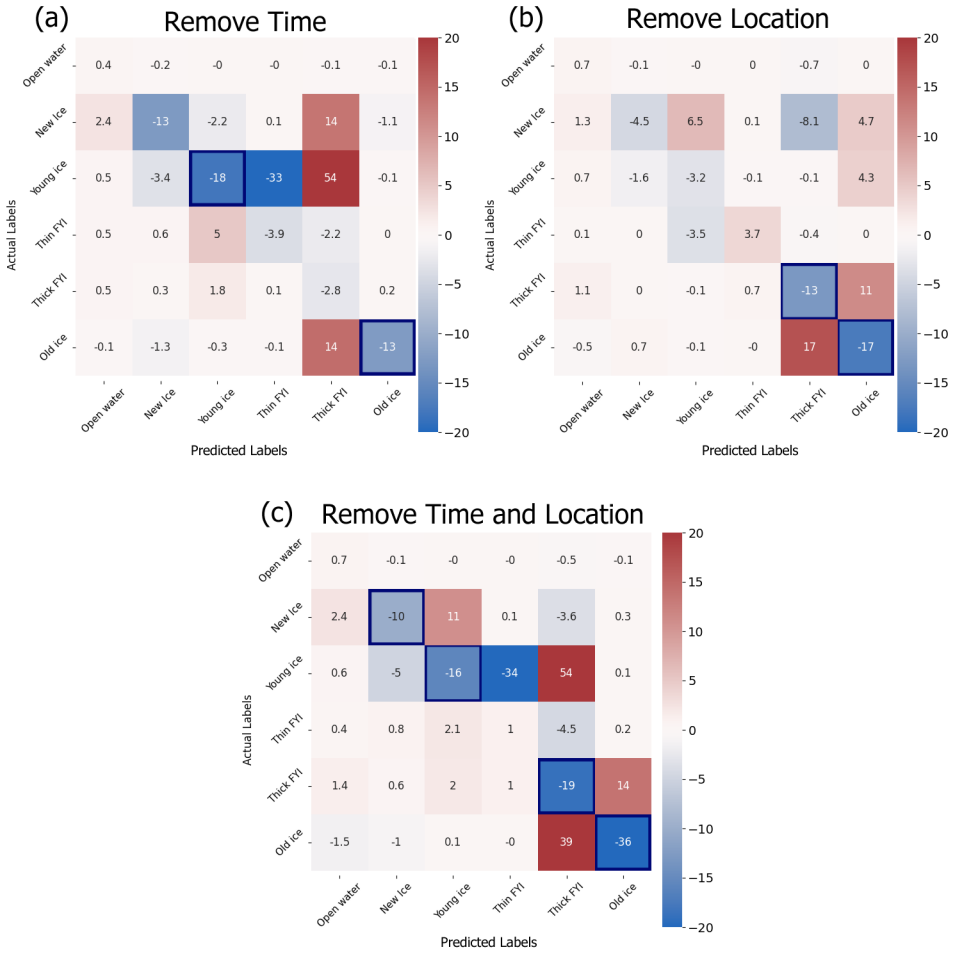


Figure 4.16: (a) The value change in each element of the SOD classification confusion matrix after removing time inputs from the full model. (b) The value change in each element of the SOD classification confusion matrix after removing location inputs from the full model. (c) The value change in each element of the SOD classification confusion matrix after removing time and location inputs from the full model.

Table 4.12: Numerical results obtained from models with SAR HH and HV inputs and individual ERA5 features. The blue and red colors correspond to positive and negative contributions on performance, respectively.

Model description	Cross validation (%)				Testing (%)			
	Combined Score	SIC	SOD	FLOE	Combined Score	SIC	SOD	FLOE
HH, HV Only	88.01	90.74	86.10	86.37	75.63	77.57	76.19	70.60
HH, HV, and wind speed	88.56	90.92	86.88	87.22	**79.94 (+4.31)**	**86.71 (+9.13)**	**77.15 (+0.96)**	**71.97 (+1.38)**
HH, HV, and SKT	**88.90 (+0.89)**	**91.61 (+0.87)**	**87.13 (+1.03)**	87.04	**71.43 (-4.20)**	**67.82 (-9.76)**	76.35	**68.81 (-1.79)**
HH, HV, and T2M	88.65	91.11	**87.38 (+1.27)**	86.26	**71.96 (-3.67)**	**69.13 (-8.44)**	76.16	**69.21 (-1.38)**
HH, HV and TCWV, TCLW	88.45	91.12	86.75	86.49	**71.62 (-4.00)**	**69.70 (-7.87)**	**75.05 (-1.14)**	**68.60 (-2.00)**

4.4.5 ERA5 Feature Analysis

To investigate the effect of different ERA5 features on model performance, each individual feature is added to the "HH, HV only" model for comparison, as presented in Table 4.12. A slight improvement in cross-validation accuracy can be observed in all models with ERA5 feature inputs. However, only the model that incorporates wind speed input improves the accuracy of the test, while others show lower scores, especially in SIC. Therefore, we look into each SAR scene and find the one with low SIC scores, as depicted in Fig. 4.17. The corresponding ERA5 feature maps of this SAR scene is presented in Fig. 4.18. Fig. 4.17(d) shows that a large area of ice regions with 100% SIC are misclassified as open water from "HH, HV only" model, which can be caused by surface ice melt. In consequence, the patterns of ice in SAR imagery become blurry. Adding non-wind ERA5 features (e.g., T2M in Fig. 4.17(e)) leads to poorer results with even larger areas being misclassified. This issue is significantly alleviated after adding wind speed as data inputs, as manifested in Figs. 4.17(f) and 4.19 (classification accuracy of 100% SIC increase by 16%). Thus, it can be concluded that incorporating wind information as data inputs could further improve model performance in SIC estimation under challenging circumstances such as melting on ice surface, while other weather parameters might not be helpful in this case.

Furthermore, as shown in Table 4.7, adding all ERA5 features significantly improves the testing performance. In contrast, removing the ERA5 features from the full model as shown in Table 4.6 does not significantly reduce the scores in either testing or cross validation. Indicating that the ERA5 features are redundant when the AMSR2 features are already present. These findings contradict what is found in the literature. Algorithms that utilize AMSR2 normally benefit from the addition of ERA5 weather data. Han et

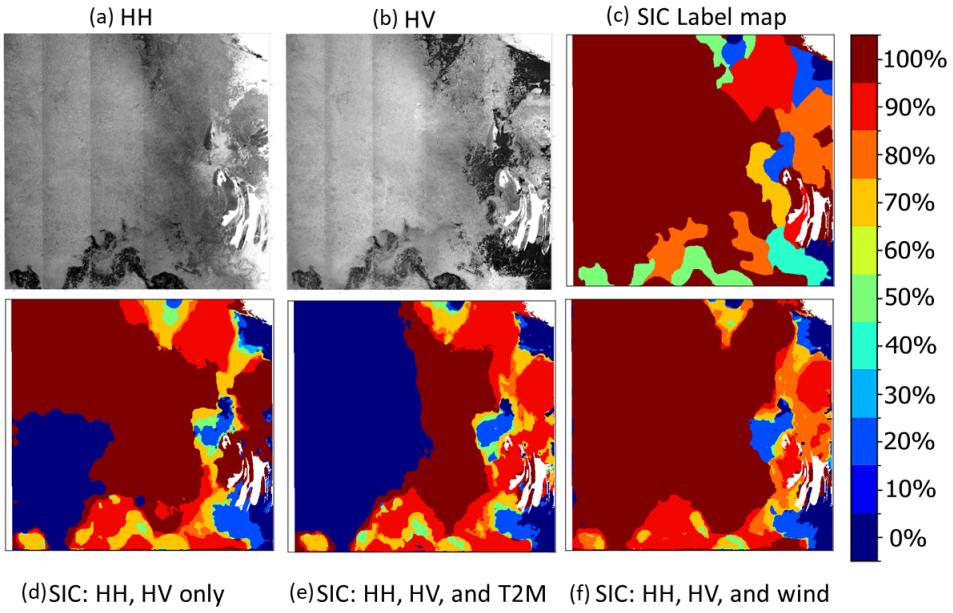


Figure 4.17: An example SAR scene in the testing dataset (file ID name: 20180623T114935_cis, collected on June 23^{rd}, 2018 in Hudson Bay, near the Belcher Islands, center latitude/longitude: $57.32°N, 81.98°W$), with HH and HV channels depicted in (a) and (b), respectively. (c) The corresponding SIC label map. (d) Mapping results from "HH, HV only" model. (e) Mapping results from "HH, HV, and T2M" model in Table 4.12. (f) Mapping results from "HH, HV, and wind speed" model in Table 4.12.

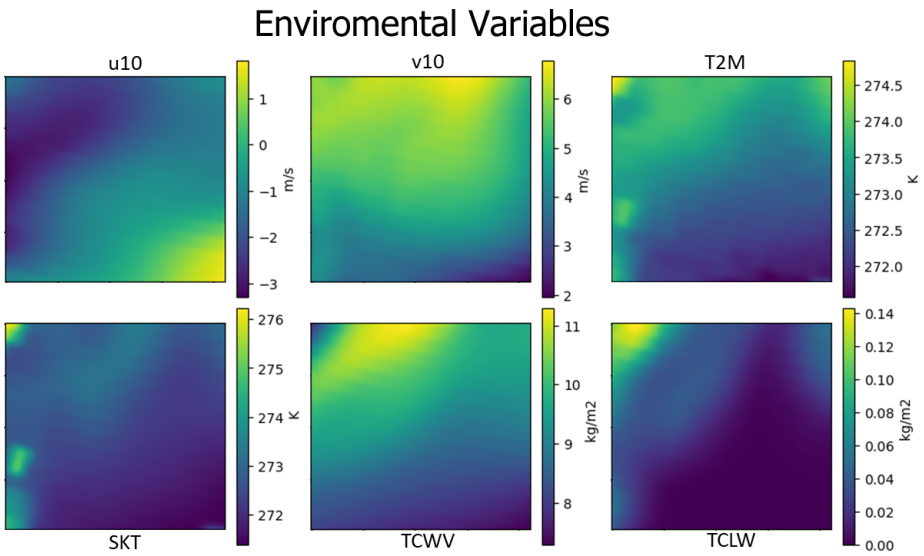


Figure 4.18: The corresponding ERA5 feature maps of the SAR scene in Fig.4.17.

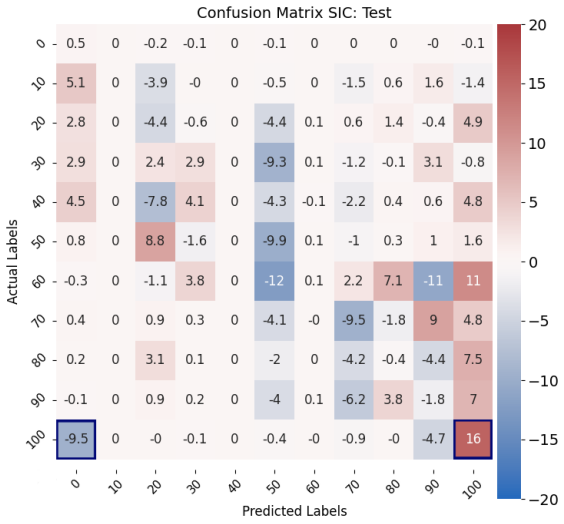


Figure 4.19: The value change in each element of the SIC classification confusion matrix after adding ERA5 features, in comparison with the "HH, HV only" model.

al [25] used a random forest algorithm with AMSR2 data + ERA5 data to improve the performance of SIC estimation. They found that their algorithm using AMSR2 + ERA5 had at least an 8.3% increase in R^2 score over other algorithms that only utilized AMSR2 data. Furthermore, Nihashi *et al.* [47] found that adding ERA5 data in conjunction to AMSR2 improved the thickness estimation of thin sea ice. However, none of these studies used SAR imagery data. It can be hypothesized that when both SAR imagery and AMSR2 data are present, the addition of ERA5 weather data becomes unnecessary.

4.4.6 Feature Importance Ranking

The ranking for the relative importance of different data inputs could provide valuable guidance for operational sea ice mapping. Thus, based on the detailed analysis above, we implement a relative importance score (RIS) to evaluate each feature. For example, the RIS of AMSR2 6.9 GHz in SIC (RIS_{SIC}) can be expressed as

$$\begin{aligned} RIS_{SIC} = {} & SIC^{val}_{HH,HV,6.9GHz} - SIC^{val}_{HH,HV} \\ & + SIC^{test}_{HH,HV,6.9GHz} - SIC^{test}_{HH,HV}, \end{aligned} \tag{4.19}$$

where $SIC^{val}_{HH,HV,6.9GHz}$ and $SIC^{test}_{HH,HV,6.9GHz}$ correspond to the SIC scores of the model with SAR HH, HV, and AMSR2 6.9 GHz inputs in cross validation and testing sets. $SIC^{val}_{HH,HV}$ and $SIC^{test}_{HH,HV}$ refer to the scores of the model with SAR HH and HV inputs only. All the scores used for calculation can be found in Table 4.7. This equation is also implemented to calculate the RIS of the combined scores, SOD, and FLOE. The equation indicates that the larger the improvement compared to the "HH, HV only" model, the more important the feature is.

The rankings calculated from Equation 4.19 are summarized in Table 4.13. Features that improve the scores significantly (here determined as RIS values higher than 3%) are highlighted in blue. In contrast, features that reduce the scores (here determined as negative RIS values) are highlighted in red. Features with positive RIS but less than 3% are not highlighted. These rankings are consistent with our findings above. AMSR2 features (except 89.0 GHz) play a vital role in improving the accuracy of all parameters. Time and location information are the most effective data inputs in boosting SOD classification accuracy. Wind data could effectively improve SIC estimation accuracy. Features including distance-to-land map and incidence angle do not seem to affect model performance, while others such as AMSR2 89.0 GHz and non-wind ERA5 data could negatively affect model accuracy.

Table 4.13: A ranking of importance for different features in the combined score and individual parameters. Features that show significant improvement, no significant effects, and negative effects in mapping accuracy are in blue, black, and red, respectively.

Combined Score Ranking	SIC Score Ranking	SOD Score Ranking	FLOE Score Ranking
AMSR2 6.9 GHz	AMSR2 36.5 GHz	Time information	AMSR2 6.9 GHz
AMSR2 7.3 GHz	AMSR2 18.7 GHz	AMSR2 6.9 GHz	AMSR2 36.5 GHz
AMSR2 10.7 GHz	AMSR2 10.7 GHz	Location information	AMSR2 23.8 GHz
AMSR2 36.5 GHz	AMSR2 7.3 GHz	AMSR2 7.3 GHz	AMSR2 18.7 GHz
AMSR2 18.7 GHz	AMSR2 6.9 GHz	AMSR2 10.7 GHz	AMSR2 10.7 GHz
AMSR2 23.8 GHz	AMSR2 23.8 GHz	AMSR2 18.7 GHz	AMSR2 7.3 GHz
Location information	Wind	AMSR2 36.5 GHz	Wind
Wind	AMSR2 89.0 GHz	AMSR2 23.8 GHz	Distance
Time information	Location information	Wind	Location information
AMSR2 89.0 GHz	IA	Distance	Time information
IA	Distance	T2M	IA
Distance	Time information	IA	Skin Temperature
T2M	T2M	Skin Temperature	AMSR2 89.0 GHz
Skin Temperature	Skin Temperature	TCWV, TCLW	T2M
TCWV, TCLW	TCWV, TCLW	AMSR2 89.0 GHz	TCWV, TCLW

While ASMR2 channels highly improve the model's performance, adding more than one AMSR2 frequency may not improve the performance of the model due to redundant data. As such, it is advised to instead add data from different sources. Here our recommendation would be to utilize one frequency channel from AMSR2 and time and location channels.

4.5 Conclusion

To provide insights and guidance for automated sea ice mapping from SAR imagery and multi-source data, this study conducts a comparative study to evaluate the importance of various data inputs on sea ice parameter retrieval. The AI4Arctic dataset with SAR imagery and corresponding AMSR2, ERA5, ancillary data is utilized to train multitask models based on a U-Net architecture. For feature importance analysis, models with different combinations of data inputs are compared with each other via ablation studies and the alternate adding of individual data input.

The ablation studies demonstrate that compared to the model incorporating all available inputs, removing AMSR2, time and location information channels has a significant impact on model performance, especially the classification accuracy of major ice types (young ice, thick FYI, and old ice) in SOD. Besides, time information further boosts the classification of ice types that are time-sensitive (e.g., new and young ice), while location

information facilitates the model to have a better knowledge of the spatial distributions of FYI and old ice. The removal of other feature types (e.g., ERA5 and IA) does not reduce accuracy scores, which indicates their redundancy among all available features.

We also evaluate the impact of adding individual feature on the model with dual-polarized SAR imagery as inputs. Results validate the effectiveness of most of the AMSR2 channels (except 89.0 GHz) on improving the mapping accuracy of all the three sea ice parameters (SIC, SOD, and FLOE). Nevertheless, adding multiple AMSR2 channels with different frequencies might not further improve model performance due to information redundancy. Among all the ERA5 weather parameters, adding wind speed data inputs improves SIC estimation accuracy by correcting the misclassification of ice regions, whereas other features might negatively affect estimation accuracy. Finally, based on the statistical results, a metric is proposed to rank the relative importance of all auxiliary data inputs in sea ice mapping, which helps us identify features that are useful in enhancing model performance.

Based on the findings of this study, we recommend the following data sources to be included alongside dual-polarized SAR for enhanced performance. First, one frequency from AMSR2, with 6.9 GHz being the most advantageous (which would include both vertical and horizontal polarizations). Additionally, incorporating latitude and longitude channels, as well as a temporal channel indicating the month, is advised to further improve model accuracy.

Chapter 5

Conclusion

This thesis presented two distinct research areas within the domain of automated sea ice mapping. The first area focused on understanding the dependence of incidence angle (IA) on GLCM texture features and its impact on sea ice classification. The second area evaluated the importance of various data inputs for automated sea ice mapping, using a comprehensive dataset to train and test a U-Net based model.

In the study on IA dependence, GLCM features were extracted from SAR images in dB units. Linear regression and class separability metrics were used to analyze their dependence on IA, and a Bayesian classifier was trained to measure classification performance. The R^2 values showed that IA accounted for at most 25% of the variance, indicating that the relationship between GLCM features and IA is quite weak. Class separability, measured using the Jeffries-Matusita distance, improved slightly with the inclusion of IA, particularly for open water and HV polarization, but the changes were not substantial for features with initially high separability. The results indicated that the IA effect had a minor impact on classification performance, with improvements of +0.91% and +0.85% for GLCM features derived from HH and HV, respectively. When combining all features (HH+HV+$GLCM_{HH+HV}$), the performance improvement due to IA was negligible (+0.06%).

The study on data input selection utilized the AI4Arctic dataset, incorporating SAR imagery, AMSR2, ERA5, and ancillary data, to train a U-Net based model. Ablation studies and the addition of individual data inputs were conducted to assess their impact on model performance. The results demonstrated that removing AMSR2, time, and location data significantly decreased model performance, especially for the classification accuracy of major ice types in SOD. Time information improved the classification of time-sensitive ice

types, and location data enhanced the model's understanding of the spatial distributions of FYI and old ice. Adding AMSR2 channels generally enhanced mapping accuracy, while the ERA5 data had mixed effects, with wind speed data improving SIC estimation but other features negatively impacting accuracy.

These findings are crucial for developing more accurate and efficient automated sea ice mapping systems. The minimal impact of IA dependence on GLCM features suggests that accounting for IA may not be necessary, simplifying the feature extraction process. Identifying the most valuable data inputs allows for the optimization of model performance, ensuring better resource allocation and enhanced operational capabilities in sea ice monitoring. This research provides a foundation for future studies and developments in automated sea ice mapping, contributing to more effective climate monitoring and maritime navigation safety.

5.1 Future Work

Although this thesis covers substantial technical ground, many experiments and research areas require deeper investigation. A brief summary of future work for Chapters 3 and 4 is presented here:

- Chapter 3: Incidence Angle Dependence of GLCM Texture Features
  - While Chapter 3 evaluated the dependence between GLCM texture features and IA when the GLCM features were extracted from backscatter in dB, similar experiments should be conducted with GLCM texture features extracted in linear units. Previous research, such as [40, 58], and [22], found that GLCM texture features are dependent on IA when calculated from linear units. Furthermore, [58] highlighted that GLCM features extracted from linear units exhibit higher class separability.
  - A potential hypothesis for the IA dependence with respect to GLCM features is the decreasing signal-to-noise ratio (SNR) with increasing IA, as noise can affect the expected texture. To validate this hypothesis, future research should investigate the relationship between GLCM features and SNR.
  - Additionally, if there is a relationship between GLCM features and SNR, it would be valuable to explore how this dependence varies with different denoising algorithms and multilook processes.

- Chapter 4: Data Input Selection for Automated Sea Ice Mapping
  - An intriguing question for future work is why the inclusion of passive microwave data enhances the performance of FLOE classification. The resolution of passive microwave data varies between 3 km and 65 km, which is too low to delineate actual floes. One hypothesis is that FLOE may be correlated with SOD, so improved SOD classification performance would also enhance FLOE classification. However, this does not fully explain why AMSR2 improves FLOE classification. Otherwise, temporal data should increase FLOE classification performance as it significantly improves SOD performance. Further investigation into these relationships is necessary.
  - The current thesis manually assesses the importance of each feature by observing performance changes, a process that is time-consuming and limits the depth of analysis. Future research should focus on developing an automated program capable of accurately determining the usefulness of each feature without the need to retrain the model, thereby enabling more comprehensive evaluations, such as performance changes per SOD class.
  - Finally, while the U-Net model presented in [7] is the top-performing model for the AI4Arctic Challenge dataset, its operational utility is limited due to the requirement that all data channels (SAR, AMSR2, ERA5, etc.) be present. To overcome this limitation, future work could explore the use of a modified Vision Transformer (ViT), where tokens represent specific data channels. Since transformers allow for a variable number of tokens, this modified ViT could handle any combination of available data types, similar to the approach described in [30].

References

[1] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74, 2021.

[2] Hugo Boulze, Anton Korosov, and Julien Brajard. Classification of sea ice types in Sentinel-1 SAR data using convolutional neural networks. *Remote Sensing*, 12(13), 2020.

[3] Nikhil Buduma and Nicholas Locascio. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, Inc., 1st edition, 2017.

[4] Jørgen Buus-Hinkler, Tore Wulf, Andreas Rønne Stokholm, Anton Korosov, Roberto Saldo, Leif Toudal Pedersen, David Arthurs, Rune Solberg, Nicolas Longépé, and Matilde Brandt Kreiner. Ai4arctic sea ice challenge dataset, Nov 2022.

[5] Canadian Ice Service. *Manual of Standard Procedures for Observing and Reporting Ice Conditions*. Canadian Ice Service, Ottawa, ON, Canada, revised 9th edition, 2005.

[6] Tom Carrieres, Mark Buehner, Jean-Franois Lemieux, and Leif Toudal Pedersen, editors. *Sea Ice Analysis and Forecasting: Towards an Increased Reliance on Automated Prediction Systems*. Cambridge University Press, Cambridge, 2017.

[7] X. Chen, M. Patel, F. J. Pena Cantu, J. Park, J. Noa Turnes, L. Xu, K. A. Scott, and D. A. Clausi. MMSeaIce: A collection of techniques for improving sea ice mapping with a multi-task model. *The Cryosphere*, 18(4):1621–1632, 2024.

[8] Xinwei Chen, Fernando J. Pena Cantu, Muhammed Patel, Linlin Xu, Neil C. Brubacher, K. Andrea Scott, and David A. Clausi. A comparative study of data

input selection for deep learning-based automated sea ice mapping. *International Journal of Applied Earth Observation and Geoinformation*, 131:103920, 2024.

[9] Xinwei Chen, Muhammed Patel, Fernando J Pena Cantu, Jinman Park, Javier Noa Turnes, Mingzhe Jiang, Linlin Xu, K Andrea Scott, David A Clausi, and Weimin Huang. The influence of input variable selection on deep learning-based sea ice parameter inversion from multi-sensor satellite data. In *OCEANS 2023-MTS/IEEE US Gulf Coast*, pages 1–4. IEEE, 2023.

[10] Xinwei Chen, K Andrea Scott, Mingzhe Jiang, Yuan Fang, Linlin Xu, and David A Clausi. Sea ice classification with dual-polarized SAR imagery: A hierarchical pipeline. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 224–232, 2023.

[11] Kohei Cho and Kazuhiro Naoki. Evaluation of AMSR2 thin ice area extraction algorithm applied to the sea ice zones of the Northern Hemisphere. *40th Asian Conference on Remote Sensing, ACRS 2019: Progress of Remote Sensing Technology for Smart Future*, (Acrs):1–10, 2020.

[12] David A. Clausi. Comparison and fusion of co-occurrence, Gabor and MRF texture features for classification of SAR sea-ice imagery. *Atmosphere-Ocean*, 39(3):183–194, sep 2001.

[13] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[14] Josefino C. Comiso and Fumihiko Nishio. Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research: Oceans*, 113(C2), feb 2008.

[15] Anca Cristea, Jeroen van Houtte, and Anthony P. Doulgeris. Integrating incidence angle dependencies into the clustering-based segmentation of SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2925–2939, 2020.

[16] Jonathan J. Day, Sarah Keeley, Gabriele Arduini, Linus Magnusson, Kristian Mogensen, Mark Rodwell, Irina Sandu, and Steffen Tietsche. Benefits and challenges of dynamic sea ice for weather forecasts. *Weather and Climate Dynamics*, 3(3):713–731, July 2022.

[17] Iris de Gélis, Aurélien Colin, and Nicolas Longépé. Prediction of categorized sea ice concentration from Sentinel-1 SAR images based on a fully convolutional network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:5831–5841, 2021.

[18] Wolfgang Dierking. Sea ice monitoring by synthetic aperture radar. *Oceanography*, 26(2):100–111, jun 2013.

[19] Donald L. Forbes, editor. *State of the Arctic Coast 2010 – Scientific Review and Outlook*. Helmholtz-Zentrum Geesthacht, Germany, Geesthacht, Germany, 2011. Available at: http://arcticcoasts.org.

[20] Carolina Gabarró, Nick Hughes, Jeremy Wilkinson, Laurent Bertino, Astrid Bracher, Thomas Diehl, Wolfgang Dierking, Veronica Gonzalez-Gambau, Thomas Lavergne, Teresa Madurell, Eirik Malnes, and Penelope Mae Wagner. Improving satellite-based monitoring of the polar regions: Identification of research and capacity gaps. *Frontiers in Remote Sensing*, 4, 2023.

[21] Government of Canada. Latest ice conditions. https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/latest-conditions.html, 2024. Accessed: 2024-06-18.

[22] Wenkai Guo, Polona Itkin, Suman Singha, Anthony Paul Doulgeris, Malin Johansson, and Gunnar Spreen. Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture. *The Cryosphere*, 17(3):1279–1297, mar 2023.

[23] Bhisham C Gupta and Irwin Guttman. *Statistics and probability with applications for engineers and scientists*. John Wiley & Sons, 2014.

[24] Mryka Hall-Beyer. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38(5):1312–1338, 2017.

[25] Hyangsun Han, Sungjae Lee, Hyun-Cheol Kim, and Miae Kim. Retrieval of summer sea ice concentration in the Pacific Arctic ocean from AMSR2 observations and numerical weather data using random forest regression. *Remote Sensing*, 13(12):2283, jun 2021.

[26] Yanling Han, Yekun Liu, Zhonghua Hong, Yun Zhang, Shuhu Yang, and Jing Wang. Sea ice image classification based on heterogeneous data fusion and deep learning. *Remote Sens.*, 13(4):592, 2021.

[27] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, nov 1973.

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[29] M. M. Holland and C. M. Bitz. Polar amplification of climate change in coupled models. *Climate Dynamics*, 21(3-4):221–232, 2003.

[30] Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, Hao Li, Pedram Ghamisi, Xiuping Jia, et al. SpectralGPT: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[31] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, jan 1989.

[32] Keiji Imaoka, Misako Kachi, Hideyuki Fujii, Hiroshi Murakami, Masahiro Hori, Akiko Ono, Tamotsu Igarashi, Keizo Nakagawa, Taikan Oki, Yoshiaki Honda, and Haruhisa Shimoda. Global change observation mission (GCOM) for monitoring carbon, water cycles, and climate change. *Proceedings of the IEEE*, 98(5):717–734, 2010.

[33] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*, 1:448–456, feb 2015.

[34] John R. Jensen. *Remote sensing of the environment: An earth resource perspective.* Pearson custom library. Pearson, Harlow, England, second edition, pearson new international edition. edition, 2014 - 2014.

[35] Juha Karvonen. Baltic sea ice concentration estimation using Sentinel-1 SAR and AMSR2 microwave radiometer data. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2871–2883, 2017.

[36] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[37] Anton Korosov, Denis Demchev, Nuno Miranda, Niccolò Franceschi, and Jeong-Won Park. Thermal denoising of cross-polarized Sentinel-1 data in interferometric and extra wide swath modes. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2021.

[38] Steven Leigh, Zhijie Wang, and David A. Clausi. Automated ice–water classification using dual polarization SAR satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5529–5539, 2014.

[39] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle. In *Annals of Glaciology*, volume 61, pages 260–270. Cambridge University Press, dec 2020.

[40] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. Incident angle dependence of Sentinel-1 texture features for sea ice classification. *Remote Sensing*, 13(4), 2021.

[41] Dettwiler MacDonald and Associates Ltd. *RADARSAT-2 Product Format Definition*. MacDonald, Dettwiler and Associates Ltd, Richmond, B.C., Canada, 15th edition, October 2016.

[42] Mallik S. Mahmud, Torsten Geldsetzer, Stephen E. L. Howell, John J. Yackel, Vishnu Nandan, and Randall K. Scharien. Incidence angle dependence of HH-polarized C- and L-band wintertime backscatter over Arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11):6686–6698, 2018.

[43] David Malmgren-Hansen, Leif Toudal Pedersen, Allan Aasbjerg Nielsen, Matilde Brandt Kreiner, Roberto Saldo, Henning Skriver, John Lavelle, Jørgen Buus-Hinkler, and Klaus Harnvig Krane. A convolutional neural network architecture for Sentinel-1 and AMSR2 data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):1890–1902, 2021.

[44] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P. Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1):6–43, mar 2013.

[45] Lawrence R. Mudryk, Jackie Dawson, Stephen E. L. Howell, Chris Derksen, Thomas A. Zagon, and Mike Brady. Impact of 1, 2 and 4 °C of global warming on ship navigation in the Canadian Arctic. *Nature Climate Change*, 11(8):673–679, aug 2021.

[46] Anmol Sharan Nagi, Devinder Kumar, Daniel Sola, and K Andrea Scott. RUF: Effective sea ice floe segmentation using end-to-end RES-UNET-CRF with dual loss. *Remote Sens.*, 13(13):2460, 2021.

[47] Sohey Nihashi, Kay I. Ohshima, and Takeshi Tamura. Reconstruct the AMSR-E/2 thin ice thickness algorithm to create a long-term time series of sea-ice production in Antarctic coastal polynyas. *Polar Science*, 39(August 2023):100978, mar 2024.

[48] Shuhratchon Ochilov and David A. Clausi. Operational SAR sea-ice image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4397–4408, 2012.

[49] Jeong-Won Park, Anton Andreevich Korosov, Mohamed Babiker, Joong-Sun Won, Morten Wergeland Hansen, and Hyun-Cheol Kim. Classification of sea ice types in Sentinel-1 synthetic aperture radar images. *The Cryosphere*, 14(8):2629–2645, 2020.

[50] Alan J. Parkinson. *Climate Change and Artic Sustainable Development*, chapter Sustainable Development, Climate Change and Human Health in the Artic, pages 156–163. United Nations Educational, Scientific, and Cultural Organization, Place de Fontenoy, 73752 Paris 07 SP, France, 2009.

[51] E Peacock, AE Derocher, GW Thiemann, and I Stirling. Conservation and management of Canada's polar bears (ursus maritimus) in a changing arctic. *Canadian Journal of Zoology*, 89(5):371–385, 2011.

[52] Fernando J. Pena Cantu, Xinwei Chen, Linlin Xu, Max Ian A. Manning, K. Andrea Scott, and David A. Clausi. Incidence angle dependence of texture features from dual polarization RADARSAT-2 sea ice imagery. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 172–175, 2023.

[53] Yibin Ren, Xiaofeng Li, Xiaofeng Yang, and Huan Xu. Development of a dual-attention U-Net model for sea ice and open water classification on SAR images. *IEEE Geosci. Remote Sens. Lett.*, 19:1–5, 2022.

[54] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[55] Martin S.J. Rogers, Maria Fox, Andrew Fleming, Louisa van Zeeland, Jeremy Wilkinson, and J. Scott Hosking. Sea ice detection using concurrent multispectral and synthetic aperture radar imagery. *Remote Sensing of Environment*, 305:114073, 2024.

[56] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[57] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[58] Randall Kenneth Scharien and Sasha Nasonova. Incidence angle dependence of texture statistics from Sentinel-1 HH-polarization images of winter Arctic sea ice. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[59] K Andrea Scott, Zahra Ashouri, Mark Buehner, Lynn Pogson, and Tom Carrieres. Assimilation of ice and water observations from SAR imagery to improve estimates of sea ice concentration. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):27218, 2015.

[60] K. Andrea Scott, Mark Buehner, and Tom Carrieres. An assessment of sea-ice thickness along the Labrador coast from AMSR-E and MODIS data for operational data assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5):2726–2737, may 2014.

[61] Mark C. Serreze and Roger G. Barry. Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, 77(1):85–96, 2011.

[62] Masanobu Shimada. A new method for correcting ScanSAR scalloping using forests and inter-scan banding employing dynamic filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12):3933–3942, 2009.

[63] Mohammed Shokr and Nirmal Sinha. *Remote Sensing Principles Relevant to Sea Ice*, chapter 7, pages 271–335. American Geophysical Union (AGU), 2015.

[64] David Small. Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(8):3081–3093, 2011.

[65] L.-K. Soh and C. Tsatsoulis. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780–795, 1999.

[66] A. Stokholm, J. Buus-Hinkler, T. Wulf, A. Korosov, R. Saldo, L. T. Pedersen, D. Arthurs, I. Dragan, I. Modica, J. Pedro, A. Debien, X. Chen, M. Patel, F. J. P. Cantu, J. N. Turnes, J. Park, L. Xu, K. A. Scott, D. A. Clausi, Y. Fang, M. Jiang, S. Taleghanidoozdoozan, N. C. Brubacher, A. Soleymani, Z. Gousseau, M. Smaczny, P. Kowalski, J. Komorowski, D. Rijlaarsdam, J. N. van Rijn, J. Jakobsen, M. S. J. Rogers, N. Hughes, T. Zagon, R. Solberg, N. Longépé, and M. B. Kreiner. The autoice challenge. *The Cryosphere*, 18(8):3471–3494, 2024.

[67] Andreas Stokholm, Tore Wulf, Andrzej Kucik, Roberto Saldo, Jorgen Buus-Hinkler, and Sine Munk Hvidegaard. AI4SeaIce: Toward solving ambiguous SAR textures in convolutional neural networks for automatic sea ice concentration charting. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[68] Lei Wang, K. Andrea Scott, Linlin Xu, and David A. Clausi. Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4524–4533, 2016.

[69] Qiyao Yu and David A. Clausi. IRGS: Image segmentation using edge penalties and region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2126–2139, 2008.

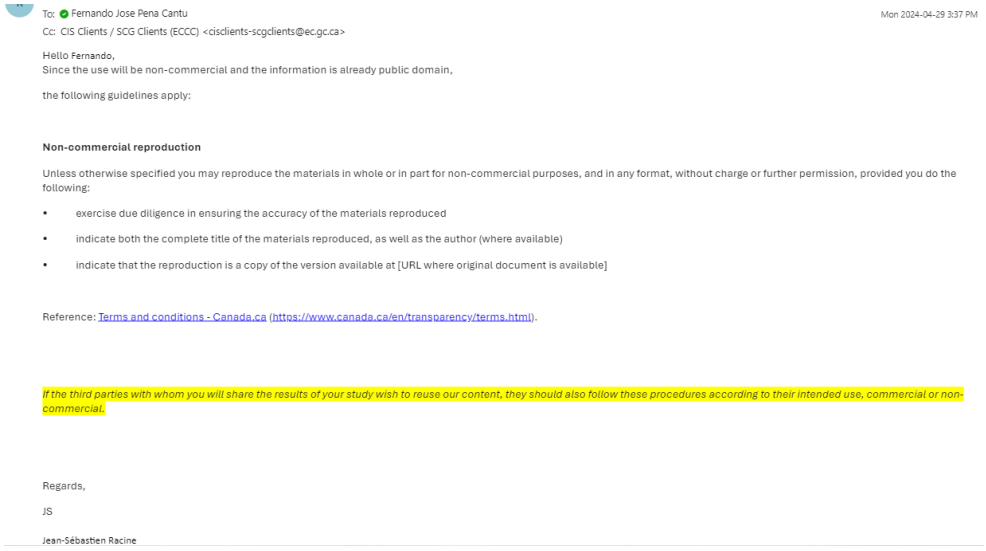
[70] B. Yue and D.A. Clausi. Sea ice segmentation using Markov random fields. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 4, pages 1877–1879 vol.4, 2001.

[71] Natalia Yu Zakhvatkina, Vitaly Yu Alexandrov, Ola M. Johannessen, Stein Sandven, and Ivan Ye Frolov. Classification of sea ice types in ENVISAT synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2587–2600, 2013.

Letters of Copyright Permission

The letters granting permission to use copyrighted materials within this thesis are included in this section.

Canadian Ice Service [5]:



To: Fernando Jose Pena Cantu — Mon 2024-04-29 3:37 PM

Cc: CIS Clients / SCG Clients (ECCC) <cisclients-scgclients@ec.gc.ca>

Hello Fernando,
Since the use will be non-commercial and the information is already public domain,

the following guidelines apply:

Non-commercial reproduction

Unless otherwise specified you may reproduce the materials in whole or in part for non-commercial purposes, and in any format, without charge or further permission, provided you do the following:

- exercise due diligence in ensuring the accuracy of the materials reproduced
- indicate both the complete title of the materials reproduced, as well as the author (where available)
- indicate that the reproduction is a copy of the version available at [URL where original document is available]

Reference: Terms and conditions - Canada.ca (https://www.canada.ca/en/transparency/terms.html).

If the third parties with whom you will share the results of your study wish to reuse our content, they should also follow these procedures according to their intended use, commercial or non-commercial.

Regards,

JS

Jean-Sébastien Racine

Incident angle dependence of Sentinel-1 texture features for sea ice classification [40]:

MDPI Open Access Information and Policy

All articles published by MDPI are made immediately available worldwide under an open access license. This means:

- everyone has free and unlimited access to the full-text of *all* articles published in MDPI journals;
- everyone is free to re-use the published material if proper accreditation/citation of the original publication is given;
- open access publication is supported by the authors' institutes or research funding agencies by payment of a comparatively low **Article Processing Charge (APC)** for accepted articles.

Permissions

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle [39]:



CAMBRIDGE UNIVERSITY PRESS

Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle

Author: Johannes Lohse, Anthony P. Doulgeris, Wolfgang Dierking©right=Copyright © The Author(s), 2020. Published by Cambridge University Press
Publication: Annals of Glaciology
Publisher: Cambridge University Press
Date: Jun 23, 2020

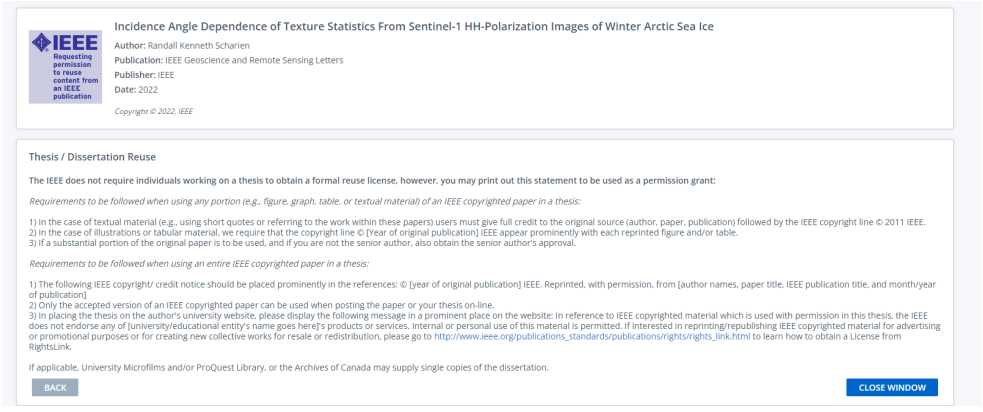
Copyright © 2020, Cambridge University Press

Creative Commons

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

Incidence angle dependence of HH-polarized C- and L-band wintertime backscatter over Arctic sea ice [58]:



IEEE Requesting permission to reuse content from an IEEE publication

Incidence Angle Dependence of Texture Statistics From Sentinel-1 HH-Polarization Images of Winter Arctic Sea Ice

Author: Randall Kenneth Scharien
Publication: IEEE Geoscience and Remote Sensing Letters
Publisher: IEEE
Date: 2022

Copyright © 2022, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

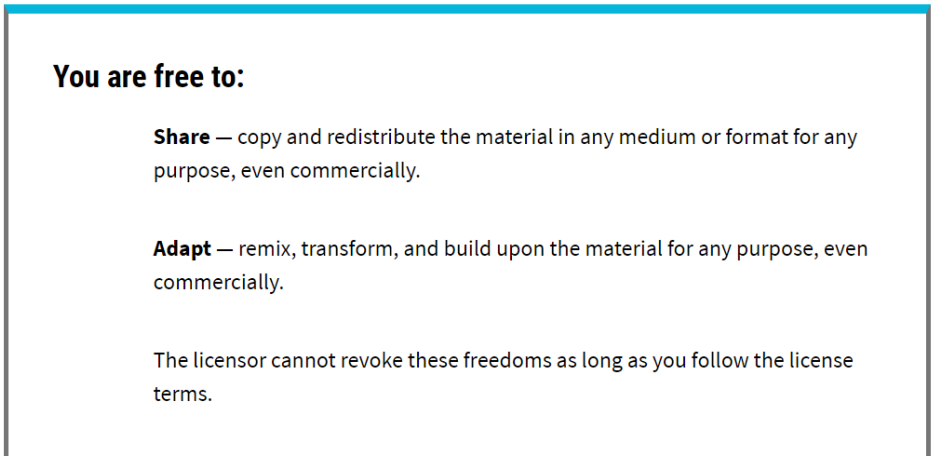
Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture [22]:



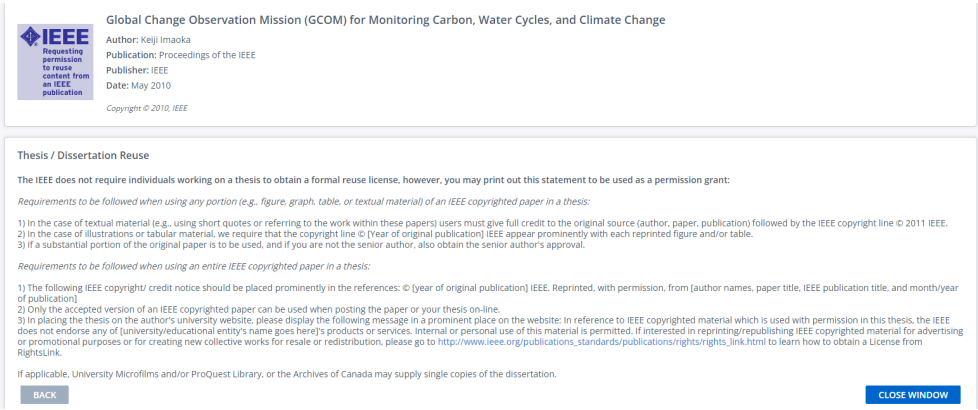
You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Global change observation mission (GCOM) for monitoring carbon, water cycles, and climate change [32]:



IEEE Requesting permission to reuse content from an IEEE publication

Global Change Observation Mission (GCOM) for Monitoring Carbon, Water Cycles, and Climate Change

Author: Keiji Imaoka
Publication: Proceedings of the IEEE
Publisher: IEEE
Date: May 2010

Copyright © 2010, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW